

マルチモーダルデータによる 携帯映像閲覧者の興味推定

倉野 大二郎^{1,a)} 松村 耕平^{1,b)} 角 康之^{1,c)}

概要: タブレット型 PC やスマートフォンなどのモバイルデバイスの普及により、映像コンテンツを閲覧する形態が変わりつつある。そのため、外で歩きながらや家で寝転びながら端末を持って映像を閲覧することを可能とし、映像を見る際のデバイスと映像閲覧者との身体的関わりが深くなったといえる。本論文では、センサ群を備えたモバイルデバイスを用いて、映像閲覧者の無意識的な行動や発話、しぐさなどからその映像における閲覧者の興味の変化を推定し、映像閲覧者の反応から映像のハイライトとなるシーンを推定する方法を提案する。

Estimating Video Viewer's Interests by Multi-modal Data Captured by Smartphone

KURANO DAIJIRO^{1,a)} MATSUMURA KOHEI^{1,b)} SUMI YASUYUKI^{1,c)}

Abstract: Spreading of tablet PC and smartphone changes our viewing style of video contents and TV programs. We can view videos anywhere and anytime during walking outside and lying in bed, that increases bodily interactions between viewers and video contents. This paper proposes a method to estimate instant changes of video viewer's interests by the viewer's nonverbal responses sensed by smartphone's camera, microphone, and accelerometer.

1. はじめに

近年では、YouTube やニコニコ動画など、映像クリエイターが自由に映像を制作・公開し、それを手軽に視聴することができる投稿型映像閲覧サービスが普及している。さらに、Ustream やニコニコ生放送のように、遠隔からイベントを「横目」で眺めるような状況が増えている。映像コンテンツの閲覧者（以下、閲覧者と呼ぶ）は、映像コンテンツから受けた感銘や興味を持ったシーンをお気に入りへの登録やコメントでの記述といった明示的な行為を通して表している。このような閲覧者の明示的な行為を行わずに、閲覧者が映像コンテンツから受けた感銘や興味を持ったシーンを、閲覧者の反応から推定する研究が盛んに行わ

れている。

これまで、閲覧者の興味を推定するために、閲覧者の表情の変化 [1] や視線の移動 [2]、などに着目して研究が進められてきた。しかし、それらの研究は高度な映像・画像認識技術に強く依存する。また、閲覧者、及びディスプレイの位置が拘束されてしまう問題がある。

しかし、タブレット型 PC やスマートフォンなどのモバイルデバイスの普及によって、投稿型映像閲覧サービスの閲覧方法も従来のテレビを閲覧するような形態の他に、閲覧者が端末を手に持って閲覧する形態が増えてきている。この変化によって、他の作業をしながら（移動しながら、家事をしながら、仕事の合間に）映像閲覧することが可能になり、「ながら見」的に映像を閲覧することが多くなってきた。このような環境では映像・画像認識技術だけで閲覧者の反応を判断するのは限界がある。

一方、スマートフォンやタブレット型 PC は、映像を閲覧する手段としてだけでなく、フロントカメラや、マイク

¹ 公立はこだて未来大学
Future University Hakodate

a) d-kurano@sumilab.org

b) matsumur@acm.org

c) sumi@acm.org

ロフォン、三軸加速度センサ、照度センサ、GPS レシーバなど多くのセンサを内蔵し、閲覧者の状況を常時モニタリングすることができる。

そこで本研究では、モバイルデバイスとして、Apple 社製の iPad2 と iPodTouch を利用し、それらが内蔵するフロントカメラ、マイクロフォン、三軸加速度センサなど複数のセンサを用いて、閲覧者の反応を計測する。それらのセンサから取得したデータを用いて、閲覧者の集中度合いや興味の推定を行い、映像のハイライトとなるシーンを推定する方法を提案する。

2. 関連研究

映像閲覧者の反応を映像・音声的にとらえて興味の变化を推定する試みはこれまでに多くなされてきた。例えば、Joho ら [1] は映像提示端末に設置されたカメラでとらえた映像を画像処理し、映像閲覧者の表情変化を読み取って、映像のハイライトシーンを特定することを試みた。また、Li ら [2] はあらかじめ映像コンテンツの視覚的特徴 (visual saliency) を画像処理によって求めておき、それらの特徴に対する映像閲覧者の視線反応の有無を確認することで、閲覧者の集中度や興味の变化を推定する方法を提案した。しかしこれらの方法は、高度な画像・映像処理技術に頼ったものであり、実験環境では動作するものの、日常的な環境で利用するのは難しい。

一方で、画像処理だけでなく、システム利用者の様々な非言語的反応を統合的にとらえて、利用者の興味を推定しようとする試みが増えている。Bohus ら [3] は、研究所の受付に設置された情報キオスク端末上にコンピュータグラフィックスによる擬人化エージェントを表示し、カメラ映像、マイク音声から得られる利用者の会話や視線移動、立ち位置の変化に基づいて、受付業務会話を行うシステムを構築した。平山ら [4] は縦長の身体サイズの情報キオスク上に、Bohus らと同様に会話エージェントを搭載し、カメラ映像から利用者の視線方向や認識し、発話に現れない利用者の潜在的な興味の变化を読み取る手法を提案した。河原ら [5] は学会におけるポスター発表のような状況を推定し、発表者 1 人と聞き手 2 人の 3 人会話における聞き手反応、つまり、聞き手の相槌、うなずき、視線変化に注目し、聞き手の内容理解や興味の变化を推定することを試みた。

本研究も、システム利用者の興味反応を、視線変化、発話の有無、身体的動作といったマルチモーダルなデータから読み取ろうという意味で、これらの研究と関連が高い。しかし、既存の研究の多くはシステムやセンサ類が環境に固定されており、システムのサービスエリアに利用者が入ってきたときにだけ動作することを想定している。また、精度の高い画像・音声処理を前提としており、特別なセンサを身に着ける必要がある場合が多かった。

しかし、昨今のスマートフォンやタブレット PC には、

利用者側を向いたカメラ (以下、フロントカメラ)、マイク、加速度センサなど多くのセンサが埋め込まれており、利用者の反応や状況を常時計測することが可能になってきている。また、これらの端末は手で持ちながら利用することが想定されているので、利用者との身体接触性が高く、そのことはつまり、利用者の顔画像や音声反応を適切なエリアで計測することが容易であり、かつ、利用者の身体移動や姿勢の変化などを推定する [6] のにも適している。

したがって本研究では、スマートフォンなどのモバイルデバイスで映像閲覧する際の閲覧者の反応を端末上のセンサで計測し、それによって閲覧者の集中度や興味の变化を推定する手法を提案する。本研究は、それぞれのモダリティの判定精度を追及するよりも、複数のモダリティの組み合わせで傾向を読み取ろうとすることに特徴がある。

3. タブレット PC を使用した映像閲覧者のマルチモーダルデータの収集

3.1 本研究の構成

本研究は、映像閲覧者の反応に関わるマルチモーダルデータを収集し、音声反応と加速度変化の共起性から映像コンテンツのハイライトとなるシーンや、映像コンテンツのシーンの切れ目を推定する手法の妥当性を検討する。そのために本研究では、映像を閲覧しながらマルチモーダルデータを取得できるデータロガーと、取得したマルチモーダルデータを用いて分析を行うためのマルチモーダルデータ分析ツールの 2 つを利用する。

データロガーは多様なセンサを持ち、映像閲覧者の表情や発話、持っている端末の揺れの度合いを取得することができるモバイルデバイスを用いることとする。本研究においては、iPad2 と iPodTouch を用いる。データロガーのシステム概要図を図 1 に示す。

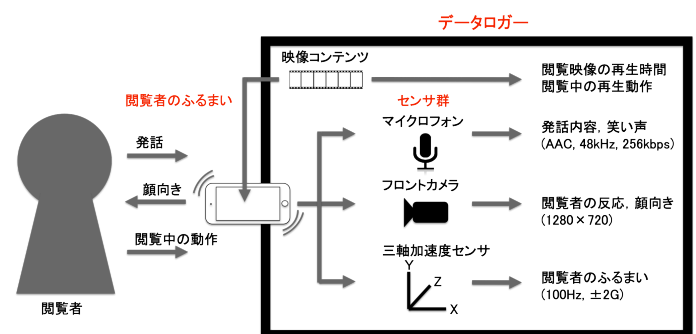


図 1 データロガーのシステム概要図

マルチモーダルデータ分析ツールは様々なデータを同期状態で閲覧し、複数の映像、音声、信号データの閲覧とラベリング作業が行えるマルチモーダルデータ分析環境 iCorpusStudio[7]*1 をベースに、閲覧している映像のタイ

*1 <http://www.ii.ist.i.kyoto-u.ac.jp/iCorpusStudio/>

ムラインに追従した動画プレーヤーおよび、データロガーで取得した三軸加速度のグラフ化をプラグインとして開発し、機能を拡張して用いる。

3.2 マルチモーダルデータとして利用するセンサ群

以下の3つのセンサからマルチモーダルデータを取得する。

マイクロフォン

映像閲覧者の発話を取得するために、iPad2/iPodTouchのモノラルマイクロフォンを利用する(256kbps)。

フロントカメラ

映像閲覧時の状態や、発話状況を分析するためにiPad2/iPodTouchのフロントカメラを利用し、閲覧の開始から終了までの様子を記録する(720p)。

三軸加速度センサ

映像閲覧者の端末の持ち換えや、寝返りなどの閲覧状態の変化、笑った時の端末の揺れを認識するために、iPad2/iPodTouchの三軸加速度を利用する($\pm 2G$)。また、取得する周波数は100Hzとする。



図2 記録の様子

次に、データロガーの傾きと三軸加速度の関係を図3に示す。

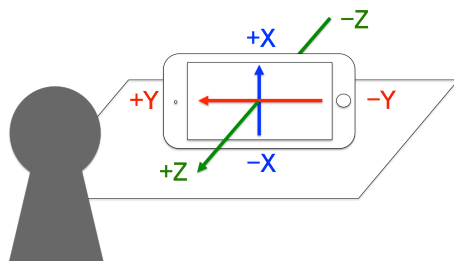


図3 データロガーの傾きと三軸加速度の関係図

3.3 閲覧動作記録

データロガーを用いて取得したデータに加えて、閲覧している映像の再生位置や閲覧中に行う一時停止や巻き戻しなどの映像の再生状態をCSVファイル形式に出力し、マ

ルチモーダルデータ分析ツールを用いて分析する。

また、iPad2とiPodTouchを用いることによって、様々な状況での映像閲覧者の閲覧動作を記録することができる。

図4は、いくつかの状況の例である。図中、a, b, c, dは、それぞれ、端末を手にとって映像を閲覧している様子、2人で映像を閲覧している様子、家で寝転がりながら映像を閲覧している様子、電車に乗りながら映像を閲覧している様子である。



図4 様々な状況での映像閲覧

3.4 閲覧動作の分析

マルチモーダルデータ分析ツールを用いることで、取得したセンサ群のデータについて可視化し、共通の時間軸を用いて分析を行うことができる。

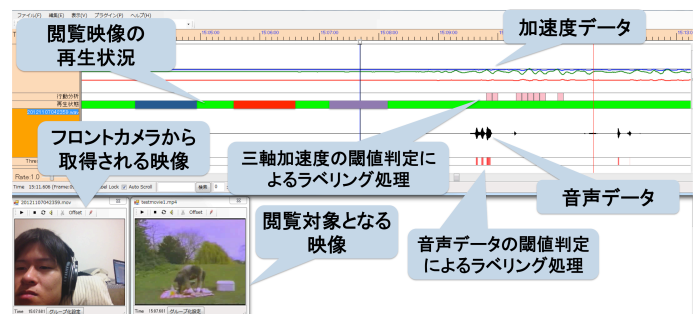


図5 マルチモーダルデータ分析ツールを用いて分析を行う様子

図5はマルチモーダルデータ分析ツールのスクリーンショットである。分析には閲覧している映像コンテンツ、フロントカメラからの閲覧者の状態、マイクロフォンからの音声データ、取得した三軸加速度と映像閲覧時の再生状態のログデータを用いる。閲覧映像の再生状況は、再生中は黄緑色、位置停止は赤色、早送りは青色、巻き戻しは紫色と色分けして表示される。また、三軸加速度データはX軸が青色、Y軸が赤色、Z軸が緑色で表示される。

また、三軸加速度が、ある閾値を超えて変化した際にラベルを追加し、閲覧者の興味変化の推定を行う手がありとすることができる。データを組み合わせ方によって、様々な事象での興味変化の推定を自動で行うことが期待できる。

3.5 顔検出による閲覧者の人数や顔向きの取得

閲覧者の人数を把握するために、データロガーに顔検出プログラムを導入した。また、図6に示すように、閲覧者がフロントカメラに対して横を向いている際は、顔を検出しないことから、閲覧者の顔向きによって映像に対しての集中度合いを推定することが期待できる。

閲覧者の顔認識を行なっている人数をCSVファイルに出力することができるので、状況に応じて適宜、音声データの代わりに利用するなど、興味を推定するためのデータの組み合わせを変更することによって様々な状況に応じた閲覧者の興味を推定することが期待できる。

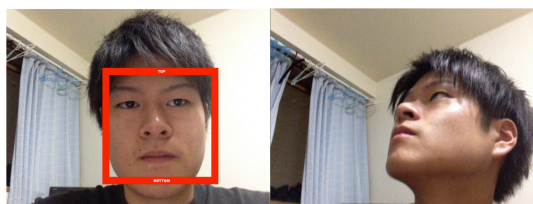


図6 顔向きによる顔検出の様子

4. マルチモーダルデータによる映像閲覧者の興味変化の推定

4.1 映像閲覧中に期待できる反応

閲覧者が共通して起こす反応として期待できる、以下のシーンを設定し、これらのシーンに注目して分析を行った。

閲覧者の笑い声や、端末の揺れによる興味変化の推定

人は笑う時に声を発し、かつ身体が小刻みに振動する。そのため、音声データと三軸加速度データが大きく変化すると考えられる(図7)。

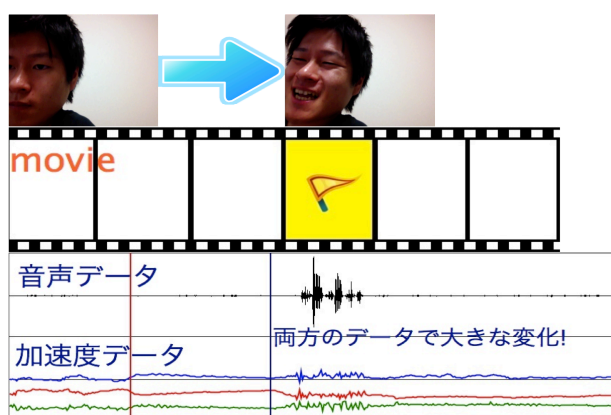


図7 三軸加速度と音声の変化によって興味推定できるケース

閲覧者の寝返り動作での興味変化の推定

人は物事に対して集中していたり注目しているときは、通常時より瞬きの回数が少なくなる[8]。このことから、閲覧者が映像に対して集中している時は、閲覧者の身動きが少なくなると推測できる。すなわち、閲覧者が寝返り動作

やデータロガーの持ち替えなどの身動きを頻繁に行う時は、閲覧者が映像に対して集中を切らしたポイントであると考えられる(図8)。このことから、多くの閲覧者が身動きを行うようなシーンは、CMや場面の切り替わりなどの、シーンの切れ目となるポイントであると推定できる。

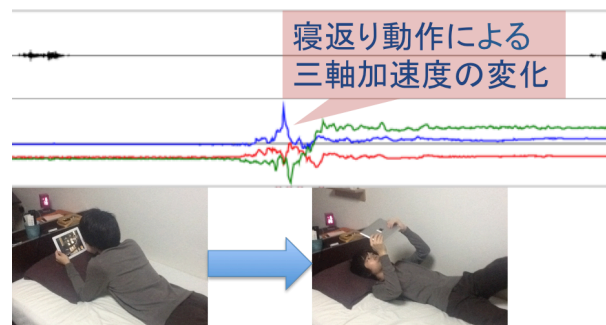


図8 閲覧者の寝返り動作によって興味変化を推定できるケース

4.2 複数人の発話、再生時間軸の変化による興味推定

上記の他に、常に閲覧者の状況を監視することができる本システムだからこそ、次の様な閲覧中の動作からも、閲覧者の興味を推定することが期待できる。

閲覧者が興味を持ったポイントを第三者と共有しようとするとき、閲覧していた映像を巻き戻し、再び興味を持ったポイントを再生すると考えられる。この時、閲覧者が興味を持ったポイントを再生時間軸の変化から推定することが期待できる(図9)。

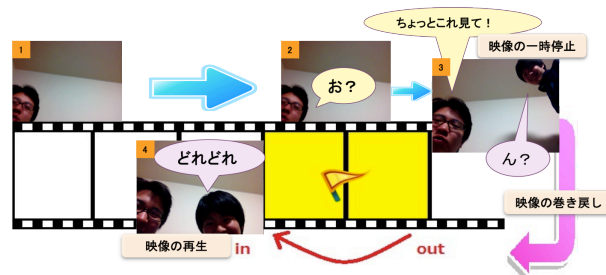


図9 閲覧者の発話、行動から興味推定できるケース

5. 予備検討

閲覧者の反応を検証する前に、予備検討として様々な状況で、映像閲覧を行なってもらった。予備検討では、閲覧者の笑いポイントに注目し、音声反応及び三軸加速度の変化を調査した。

5.1 第三者が多数存在する場での映像閲覧

公共の場など、第三者が多数存在する場での映像閲覧では、閲覧者が周りに人がいることから、笑い声を我慢するなど、閲覧者自ら映像に対する反応を抑えるという事象が見られ、音声反応がほとんど生じなかった。また、外での

閲覧は、風の音や騒音が大きいことから、音声データから閲覧者の反応を検出できなかった。

5.2 電車内での映像閲覧

電車内での映像閲覧では、電車の騒音をデータロガーが拾ってしまい、常に大きな音になっている状態であることがわかった。また、車内の揺れから常に三軸加速度が変化していることがわかった。このことから、閲覧者の反応が電車の揺れや騒音で埋もれてしまうことがわかった。

電車内で映像を閲覧した際の主な三軸加速度と音声データを図 10 に示す。

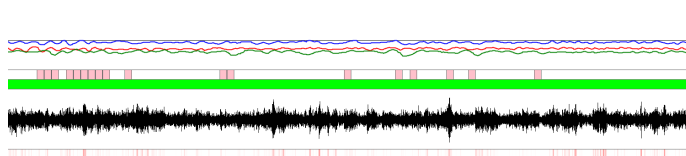


図 10 電車内で映像を閲覧した際の三軸加速度と音声データ

以上、5.1 節、本節の結果から、外部の状況によって閲覧者の反応に影響を与えにくい環境下での映像閲覧に焦点を当て、実験を行う。

6. 実験

6.1 実験内容

4.1 節で述べた予想の妥当性を検証するために、10 名の被験者に本システムを用いて映像を閲覧してもらう実験を行った。また、5 節の結果を踏まえて、実験は室内で被験者が 1 名の状態でいった。本実験では主に閲覧者の笑い動作を検証するために、8 分程度のコメディ短編映画を 2 つと、45 秒の CM を短編映画の合間に挟んだ合計約 18 分の映像を閲覧してもらった。この際、被験者には約 18 分間、映像を継続して閲覧することを依頼した。

6.2 各モーダルデータの閾値計算

本実験によって、映像閲覧実験で取得したデータをもとに三軸加速度、音声データの閾値を設定する。三軸加速度は 30ms の間に三軸 (X 軸, Y 軸, Z 軸) の平均値の差が 0.3G 以上開いた際にラベリング処理を行う。音声データは、データロガーを起動した時点から終了した時点までの音量の平均値を、100 倍した値を閾値とし、それを超えた際にラベリング処理を行う。これらの閾値は、徐々に閾値を変更し、本実験で行った笑い動作を認識する際に、一番笑いの動作の認識率が良かった閾値の組み合わせとなる。

以下に、これらの閾値を用いて、閲覧者の反応を分析した結果を示す。

6.3 閲覧者が起こした動作による三軸加速度の変化の差異

実験で被験者に映像を閲覧してもらった際に、被験者か

らは様々な動作がみられ、三軸加速度データにもそれが表れている。

特徴的な例として、笑った時の動作による変化 (図 11) と、寝返り動作による変化 (図 12 上部)、顔を掻く動作による変化 (図 12 中央)、データロガーの持ち替え動作による変化 (図 12 下部) があつた。

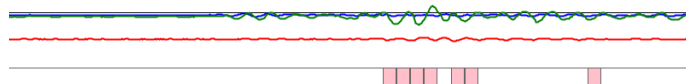


図 11 閲覧者が笑った動作をとった際の三軸加速度

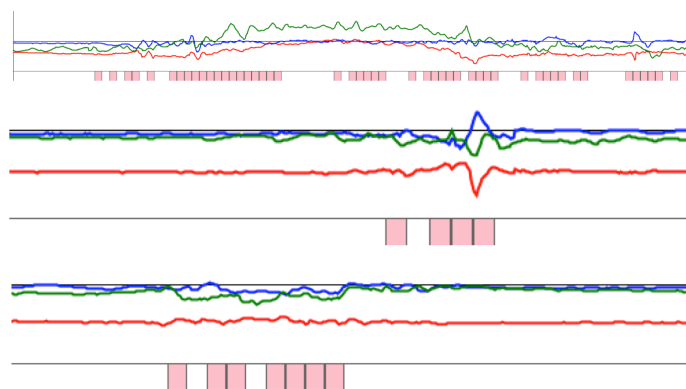


図 12 閲覧者が寝返り動作 (上部)、顔を掻く動作 (中央)、データロガーの持ち替え動作 (下部) をとった際の三軸加速度

笑った際の動作と寝返り動作における三軸加速度の差異

図 11 から笑った際の動作は Z 軸 (緑色) に細かい波が発生することがわかった。次に、図 12 上部から、寝返り動作をとった際の三軸加速度は笑った際の動作 (図 11) に比べて周期のゆるい大きな波になり、波の形状が横に長くなることがわかった。

また、寝返り動作では三軸加速度に大きな変化が長時間生じることから、閾値判定によって貼られるラベル (図 12 下の四角形) の数が笑った際の動作と比較して、多くなることがわかった。

このことから、笑った際の動作と寝返り動作による三軸加速度の変化には、特徴的な差異があり、自動的に識別することが可能である。

笑った際の動作と顔を掻く動作における三軸加速度の差異

閲覧者の顔を掻く動作による三軸加速度の変化 (図 12 中央) は、笑った際の動作 (図 11) と比較して、規則的な周期の細かい波はどの軸にも発生しなかったが、X 軸 (青色) と Y 軸 (赤色) に波の山が半周期分発生していることがわかった。しかし、現在は、閾値判定を閲覧中の微小時間における三軸加速度の変化量で行なっているため、微小時間における笑った際の動作と閲覧者が顔を掻く動作を自動的に識別することは困難である。

笑った際の動作とデータロガーの持ち替え動作における三軸加速度の差異

閲覧者のデータロガーの持ち替え動作による三軸加速度の変化（図 12 下部）は、笑った際の動作（図 11）と比較して、山と谷が認識できる波が発生していないことがわかった。節 6.3 と同様に、微小時間における三軸加速度の変化量では笑った際の動作とデータロガーの持ち替え動作を自動的に識別することは困難である。

6.4 閲覧者の笑い声や、三軸加速度の揺れによる閲覧者の興味推定

笑った際に三軸加速度と音声データに大きな変化が見られるかを検証するために、まず閲覧者が笑った際の反応に着目して分析を行った。この時、マルチモーダルデータ分析ツールの閾値判定によって、貼られたラベルを手がかりに分析を行った。三軸加速度データと音声データの閾値判定が行われているポイントの閲覧者の状態を見比べ、その際に閲覧者が笑った状態であるか、それ以外の状態であるかを分析した。

各モーダルデータの閾値判定による分析結果

三軸加速度、音声データの各モーダルデータの閾値判定、また両データの閾値判定が同時に行われていた場面の総数を表 1 に示す。表 1 から、多くの被験者が音声データより

表 1 閾値判定が行われた総数

被験者	三軸加速度	音声データ	両方
1	10	15	3
2	155	10	3
3	118	16	10
4	54	23	18
5	26	10	5
6	43	5	4
7	30	48	8
8	79	33	8
9	6	5	1
10	9	15	1

も三軸加速度の方が閾値判定が行われる回数が多いことがわかった。

次に、三軸加速度と音声データの閾値判定が同時に行われているポイントに着目し、閲覧者が笑っている際の三軸加速度と音声データの変化について分析した。図 13 でみられる、閲覧者が笑った際に三軸加速度と音声データが笑う直前と比較して、大きく変化する事例が多く見られた。

しかし、図 14 のように閲覧者が笑っていない状況でも、三軸加速度と音声データの閾値判定が行われている事例も少数見られた。この事例について調べたところ、寝返り動作中にマイクが閲覧者に触れることによって生じることがわかった。

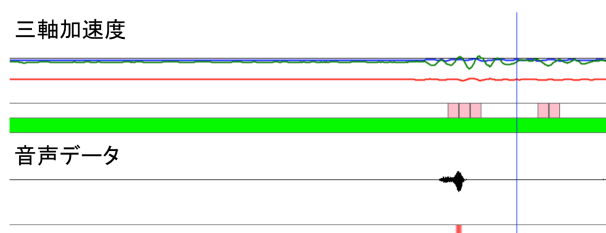


図 13 閲覧者が笑っている際に、同時に三軸加速度と音声データの閾値判定が行われている事例

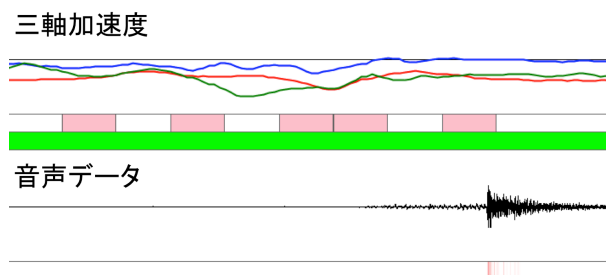


図 14 閲覧者が笑っていないが三軸加速度と音声データの閾値判定が同時に行われている事例

各モーダルデータとデータの組み合わせによる笑い動作の判定認識率

閲覧者の笑い動作を、1) 三軸加速度の閾値判定のみ、2) 音声データの閾値判定のみ、3) 三軸加速度と音声データ 2 つの閾値判定、の 3 通りの方法で自動判定することを試みた。各方法の笑い動作における判定認識率を表 2 に示す。

表 2 笑った際の認識率

被験者	認識率 (%)		
	三軸加速度	音声データ	両方
1	10	100	75
2	18.06	30	75
3	49.15	93.75	100
4	74.07	100	100
5	30.77	100	100
6	88.37	80	100
7	33.33	54.17	75
8	24.1	21.21	75
9	0	40	0
10	22.22	15.38	0

表 2 の通り、多くの閲覧者の笑い動作の判定認識率が 75%以上となった。また、三軸加速度や音声データの各モーダルのみで閾値判定を行うと、多くの誤差を含んでしまいが、両データの閾値判定を組み合わせることによって、笑い動作の判定認識率が向上することが示された。

6.5 閲覧者の寝返りや姿勢を正した際の動作による興味変化の推定

映像のシーンの切り替わりなど、閲覧者の集中度が変わる場面に、寝返り動作や姿勢を正す動作が多く見られるかを検証するために、閲覧者がこれらの動作をとったポイン

トに着目して分析した。分析から、被験者が映像を閲覧している際に多くの寝返りや、映像を閲覧する際の姿勢を正す動作が確認できた。閲覧者が寝返り動作をとった際の三軸加速度の変化を図 15 に示す。



図 15 閲覧者が姿勢を正した際の三軸加速度

表 3 は、映像閲覧中の寝返り動作や、姿勢を正す動作をとった回数の総数と、実験映像においてシーンの切れ目となる CM の前後で、このような動作をとった回数を示している。

表 3 寝返り、姿勢を正す動作の回数

被験者	総数	CM 前後
1	7	0
2	34	8
3	22	6
4	9	4
5	14	4
6	4	1
7	16	4
8	18	5
9	3	2
10	5	1

6.6 映像閲覧者が共通して反応したシーン

本実験で閲覧してもらった映像に対して、閲覧者が共通して、笑い動作や寝返り、姿勢を正す動作をとるシーンを分析した。映像時間に対し共通して反応を起こした閲覧者の人数を図 16 にまとめた。

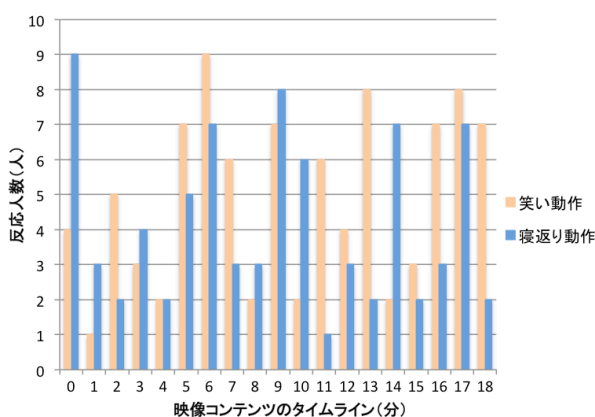


図 16 閲覧中に笑い動作を検出した人数

図 16 より、最も多くの被験者が共通して笑い動作をとったシーンは、映像の再生時間が 6 分から 7 分の時となった。また、図 16 より、被験者の 70% 以上が寝返り、姿勢を正す動作をとったシーンは、映像のタイムラインが 0 分から

1 分、6 分から 7 分、9 分から 10 分、14 分から 15 分、17 分から 18 分の時となった。

本実験で用いた映像は、9 分から 10 分の間に CM があり、14 分から 15 分の間に、短編映画の中でシーンの切り替わりが起こる。

6.7 閾値判定以外での閲覧者の笑い動作、寝返り、姿勢正しの動作

本研究で設定した閾値によって、閲覧者の笑い動作、寝返り、姿勢正しの動作を分析したが、閲覧者は設定した閾値を超えない場合にもこれらの動作をとっていた。閾値判定がどれだけ閲覧者の笑い動作、寝返り、姿勢正しの動作に対応していたかを表 4 に示す。

表 4 閾値判定による各動作の対応率

被験者	笑い動作 (%)	寝返り動作 (%)
1	5.56	100
2	58.06	100
3	87.72	100
4	84.09	100
5	30.43	77.78
6	74.51	80
7	25	94.12
8	86.36	94.74
9	0	100
10	7.41	100

7. 考察

7.1 三軸加速度と音声データの変化による笑い動作の検証

表 2 から、映像閲覧中には、三軸加速度や音声データには多くの誤差が生じることがわかった。このことから、1 つのモダルデータから閲覧者の笑い動作を認識することは難しいことがわかる。しかし、三軸加速度、音声データの閾値判定を and 演算的に組み合わせることによって、笑い動作の平均認識率が、70%を超えた。

一方で、5.2 節で述べたように、電車内などで映像を閲覧する際には、音声データと三軸加速度共に多くのノイズを含んでしまうので、三軸加速度と音声データの組み合わせのみによって、閲覧者の笑い動作を認識するには、場所や状況を限定してしまうことがわかる。

7.2 姿勢正し、寝返り動作による興味変化の検証

表 3 から、閲覧者は映像閲覧中に多くの寝返りや姿勢を正す動作をとっていたことがわかった。実験で用いた短編映画の切れ目である CM 前後で、このような動作をとる被験者は 10 名中 9 名に見られた。

また、図 16 より、本実験において、全体を通して 70% 以上の閲覧者が寝返りや姿勢を正す動作をとったシーンは、映像の開始直後である 0 分台、共通して最も閲覧者が笑い動作をとった 6 分台、短編映画の切れ目となる CM が含ま

れる9分台、短編映画の中でシーンの切り替えが起こる14分台、短編映画の結末となる17分台のシーンに見られた。よって、閲覧者の70%以上が、場面の切り替えとなるシーンや、笑いポイントとなるシーンの前後で寝返りや、姿勢を正す動作をとっていることがわかる。しかし、閲覧者はこのようなシーン以外にも多数の寝返り動作や、姿勢を正す動作をとっていた。また、姿勢を正す動作をとった直後に、映像に対して笑い動作をとった被験者も見られた。

このことから、寝返りや姿勢を正す動作のみで、閲覧者の集中度合いの変化や映像のシーンの切り替えを、閲覧者個人のデータから推定することは難しいが、閲覧者全員の集合知から、映像のシーンの切り替えを推定することは可能であるといえる。

7.3 実験データを分析する際に設定したシーンの妥当性

閲覧者が笑った際に三軸加速度と音声データに大きな変化が生じるという予想に対して、本実験の閲覧環境では予想通り閲覧者が笑った際に、三軸加速度と音声データに大きな変化が見られた。しかし、外や公共の場での閲覧では、笑いを我慢したり、閲覧者の反応が埋もれてしまった。このことから、限られた状況でのみ妥当であるといえる。

次に、閲覧者の集中度に変化が起こる、映像のシーンの切り替えに閲覧者が寝返り動作や姿勢を正す動作を多くとるという予想に対して、7.2節より、このような動作から、閲覧者の集中度合いの変化を閲覧者個人のデータによって推定することは難しいことがわかった。このことから閲覧者個人の反応から、設定したシーンが妥当であるとはいえない。

しかし、6.6節より、多数の閲覧者のデータを集約することによって、寝返りや姿勢を正す動作から、映像のシーンの切り替えを推定することが可能であるといえる。よって、多数の閲覧者のデータを集合的に考えると、設定したシーンが妥当であるといえる。

8. おわりに

スマートフォンやタブレット PCなどのモバイルデバイスを利用して映像を閲覧する状況を想定し、端末に搭載されたカメラ、マイク、加速度センサで得られるデータから閲覧者の反応を読み取り、集中度や興味の変化を推定する方法を提案した。コメディビデオ閲覧時の閲覧者の笑いに注目したところ、音声反応と端末の振動を組み合わせることで、反応解釈の精度を上げることができた。また、シーンの切れ目や集中の切れ目に姿勢を大きく変えることが多く、複数人の閲覧者の身体的反応を集約することで、シーンの切れ目や盛り上がりシーンの特定が可能になる示唆を得た。

電車の中や周りに多くの人がいる状況での映像閲覧などについても予備検討を行ったが、そのような状況では笑い

を押し殺すことが多く、音声反応や端末の振動による閲覧者反応を読み取ることは難しいことが確認された。

一方で、そういう状況では端末への顔向けが安定しない中で、特に注目するときにだけ映像に注視する傾向が高く、利用者の顔画像検出や画面上の操作を重視すべきであろう。つまり、状況に応じて解釈ルールの適用は大きく変わるものであり、今後は、そういった様々な利用状況の違いに応じた解釈戦略の切り替えに取り組みたい。

そして、そういった状況変化はGPSによる位置情報やスケジュール情報といった情報から推定することが可能であろうから、さらに携帯情報端末を利用する価値が高まる。本研究では映像閲覧という限られた情報サービスに特化して取り組んできたが、proactiveな情報提供を行うパーソナルで携帯型の情報利用機器における利用者とのインタラクションのデザインという大きな枠組みの中で取り組んでいきたい。

参考文献

- [1] Hideo Joho, Joemon M. Jose, Roberto Valenti, Nicu Sebe: Exploring Facial Expressions for Affective Video Summarisation, *Proceeding of the ACM International Conference on Image and Video Retrieval*, 2009.
- [2] Jia Li, Yonghong Tian, Tiejun Huang, Wen Goo: A dataset and evaluation methodology for visual saliency in video, *IEEE international on multimedia and expo*, pp.442-445, 2009.
- [3] Dan Bohus, and Eric Horvitz: Multiparty Turn Taking in Situated Dialog: Study, Lessons, and Directions, *Proceedings of the SIGDIAL 2011 Conference*, pp.98-109, 2011.
- [4] Takatsugu Hirayama, Yasuyuki Sumi, Tatsuya Kawahara, and Takashi Matsuyama: Info-concierge: Proactive multi-modal interaction through mind probing, *The 2011 APSIPA Annual Summit and Conference*, 2011.
- [5] 河原達也: スマートポスターボード: ポスター会話のマルチモーダルなセンシングと認識, *電子情報通信学会誌*, Vol.112, No.141, pp.7-12, 2012.
- [6] Nobuo Kawaguchi, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Tsutomu Terada, Kazuya Murao, Sozo Inoue, Yoshihiro Kawahara, Yasuyuki Sumi, and Nobuhiko Nishio: HASC Challenge: Gathering large scale human activity corpus for the real-world activity understandings, *Augmented Human 2011*, Tokyo, Japan, 2011.
- [7] 角 康之: マルチモーダルデータを用いた会話的インタラクションの構造理解, *人工知能学会誌*, Vol.27, No.4, pp.405-410, 2012.
- [8] 佐藤直樹, 山田昌和, 坪田一男: VDT 作業とドライアイの関係, *あたらしい眼科*, No.9, pp.2103-2106, 1992.