

視線追跡データから算出された注目領域に基づく 視線移動の少ない字幕配置法の提案と評価

赤堀 渉^{1,a)} 平井 辰典^{1,b)} 森島 繁生^{2,c)}

概要: 本稿では、字幕付き映像における視線移動の少ない字幕配置法を提案する。従来の画面端に固定された字幕では、映像で視線が集中する領域と字幕が離れている場合に視線移動が増えるため、映像を十分に鑑賞することができなかった。既存研究では、映像内容の理解度の向上のために動的に字幕を配置したが、字幕なし映像のどこに視線が集まるかを考慮していないため、字幕に視線が誘導され映像を集中して鑑賞することができなかった。そこで本研究では、字幕付き映像における新たな字幕配置方法を提案する。提案手法では、字幕なし映像に対して視線が集中する領域と字幕が重なることを避けるため、複数人の視線情報を測定し映像において視線が集まりやすい領域を算出する。その後、鑑賞者の字幕に対する認知的な負荷を考慮するため、字幕配置のルールとして注目される領域の下部に固定して字幕を配置する。また、評価実験により提案手法の有効性を示した。

キーワード: 視線追跡, 注目領域算出, 動的字幕配置

1. 研究背景

海外の映画やドラマといった、音声から発話内容を理解することができない映像に対し、母国語の字幕を配置することで鑑賞者の内容の理解を補助する鑑賞方法が広く普及している。字幕付き映像は音声情報なしに言語情報を取得できるため、聴覚障害者にとって重要な映像鑑賞方法である。また、総務省は2017年度までに全ての放送番組に対して字幕を配置することを指針として定めている [1]。

映像に字幕を配置する際には、鑑賞者が映像と字幕との双方を理解できるように様々な工夫が凝らされている。例えば、字幕として一度に表示できる文字数には制限があるため、字幕翻訳家によって少ない文字数かつ理解しやすい字幕が配置されている。また、字幕は映像の重要な領域と重ならないように、画面端に固定して配置される（以降、このような字幕を固定字幕と呼ぶ）。しかしながら、字幕の有無により映像作品の見た目が大きく変化することや、映像の言語を含むことから、字幕は視線を引き付けるとい



図 1 固定字幕の映像に対する視線移動の様子

う性質がある。図 1 に、固定字幕の映像に対する視線移動の様子を示す。図中の点列から表された視線の軌跡から分かる通り、映像の登場人物の顔と字幕のいずれか一方に視線が集中している。映像で注目したい領域と字幕の位置との距離が大きい場合、視線のぼらつきが大きくなり、鑑賞者の映像に対する注意を散漫にさせてしまう問題がある。

このような問題に対し、映像内容に基づいて字幕の位置を変えることで、視線のぼらつきを軽減する方法が提案されている（以降、このような字幕を動的字幕と呼ぶ）。この方法では、映像から得られる画像情報や視線情報を用いて、映像における重要な領域を算出し、その領域付近に字幕を

¹ 早稲田大学

Waseda University

² 早稲田大学理工学術院総合研究所 / JST CREST

Waseda Research Institute for Science and Engineering / JST CREST

a) akahori@akane.waseda.jp

b) tatsunori_hirai@asagi.waseda.jp

c) shigeo@waseda.jp

配置している。これにより、映像内容の理解度が向上したことが報告されている [2], [3], [4], [5]。しかし、従来の動的字幕は、映像における重要な領域を話者の顔や個人の視線に基づき算出するため、注目される領域と字幕が重なる場合がある。そこで、本研究では動的字幕が映像に過干渉することを防ぐため、映像の重要な領域を複数人の視線の平均と分散を用いて視線が集まりやすい領域として定義する。本研究の目的は、鑑賞者の視線のばらつきを軽減し、映像に集中して鑑賞できる最適な字幕の配置位置を検討することである。そのため、視線が集まる領域を算出する方法として鑑賞者の視線追跡データを用いる。具体的には、視線測定装置を用いて、字幕なし映像に対する複数人の視線追跡データを測定し、視線が集まりやすい領域を算出する（以降、字幕がついていない映像を元映像と呼ぶ）。その結果を基に字幕付き映像を作成することで、提案手法を適用した映像の鑑賞者は視線測定装置を装着することなく注目される領域を考慮した動的字幕付き映像を鑑賞できる。

本稿では、複数人の視線追跡データを用いた動的字幕の配置方法の提案とその有効性についての検討を行う。複数人の視線追跡データを使用し、映像において視線が集中しやすい領域を算出することにより、元映像において重要な領域と字幕が重なることなく、視線移動の少ない快適な鑑賞を実現する。

2. 関連研究

視線のばらつきを軽減させるため、映像や漫画において鑑賞する際の視線パターンを利用する方法が提案されている。Jain ら [6] は、複数人の視線追跡データを用いて映像における重要な領域を算出し、重要な領域のみが映し出された映像を生成する方法を提案した。その結果、生成した映像に対する視線パターンは元映像を鑑賞する際の視線パターンと類似することが明らかとなった。Cao ら [7] は、読み手の視線の流れを考慮して漫画の吹き出しを配置することにより、読みやすい漫画を自動で生成する方法を提案した。以上から、元映像に対する複数人の視線パターンを利用することは、映像と字幕を鑑賞しやすくする上で重要な要素であることが分かる。

Brown ら [2] は、映像で重要な領域付近に字幕を配置した場合、元映像を鑑賞する際の視線パターンと類似することを実験により明らかにした。しかし、字幕配置方法の提案を行っていない。そこで、定量的に動的字幕を配置する試みとして二つのアプローチで研究がなされている。一つ目は、視線追跡データを用いて動的に字幕を配置する手法である。Katti ら [3] は、鑑賞者から取得した視線追跡データと映像の画像情報から取得した顕著性マップを組み合わせて、個人に最適な字幕配置を実現した。しかし、鑑賞者個人が注目している領域付近に字幕を配置するため、鑑賞者が特に注目しなかったが映像において重要な領

域に注意を払うことができなくなる。また、個人の視線追跡データのみを用いた場合、不完全なキャリブレーションや瞬きにより視線追跡結果にノイズが含まれるため、頑健に字幕を配置することができない。二つ目として、話者の顔認識に基づき字幕を配置する手法が挙げられる [4], [5]。Hong ら [4] は、映像に登場する話者の顔領域付近に字幕を配置することで、聴覚障害者が発話中の話者を認識できない問題を解消し、映像の理解度を向上させた。手順としては、口の動きから話者認識を行い、話者付近の顕著性の低い領域に字幕を配置している。これらの研究は話者付近に字幕を配置することで、鑑賞者の話者認識を容易にさせることを目的としている。しかし、字幕の配置位置を話者の顔領域付近に限定しているため、話者以外の重要な物体に注目しづらくさせてしまう。また、字幕が長い場合、話者付近の重要な領域との重なりを避けるため、字幕の改行やフォントの大きさを小さくするといった工夫が必要となる。

一方、映像の字幕だけでなく、漫画の吹き出しや広告の文字、漫画作品に登場するような効果音を表す文字アニメーションを最適な位置に配置する方法が広く検討されている [8], [9], [10], [11]。漫画や広告画像のような静止画の場合、読み手は自分のペースで画像と文字を鑑賞することができる。また、効果音を表す文字アニメーションは映像全体の内容理解において重要ではないため、鑑賞者は効果音を表す文字を見逃した場合でも映像内容を十分に理解できる。しかし、字幕付き映像の場合、鑑賞者は内容の変化する映像と内容の理解に必要な字幕との双方に注意を払う必要がある。したがって、映像に字幕を配置する際には、鑑賞者が字幕を探す負担を考慮する必要がある。以上を踏まえ、鑑賞者の視線のばらつきを軽減を担保しつつ字幕を配置したことによる重要な映像情報の見逃しを防ぐため、元映像を鑑賞する際の視線パターンを考慮する。鑑賞者の視線を特定の領域に限定させないため、映像の重要な領域全体を俯瞰して鑑賞できるように字幕を配置する必要がある。そこで、視線が集まりやすい領域を一つの矩形で囲んで定義する。加えて、その領域の外部に字幕を配置することにより、重要な物体と字幕との重なりを回避する。また、複数人の視線追跡データを用いるため、不完全なキャリブレーションや瞬きによるノイズの影響を受けにくい。本研究では、様々なジャンルの映像に対し、視線移動を少なくしつつ、映像に対し集中して鑑賞できる字幕配置方法について検討する。

3. 提案手法

初めに元映像とそれに対する複数人の視線情報を取得する。次に、字幕を配置したい時間における視線情報の平均と分散を計算し、元映像の注目領域を定義する。最後に、注目領域に基づいて元映像に字幕を配置する。以下、注目領域の算出方法と字幕の配置方法について検討する。

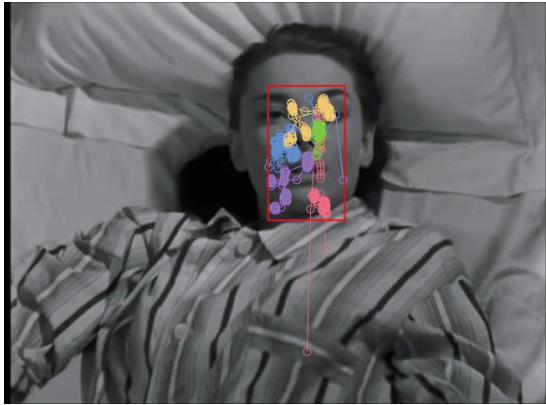


図 2 複数人の視線追跡データと算出された注目領域（赤矩形）

3.1 注目領域の算出

元映像における注目領域は複数人の視線追跡データを用いて定義する。映像を鑑賞する際に注目する領域には個人差があるため、個人が注目する領域を必ずしも他の人が注目するとは限らない。そこで、複数人に対する視線測定によって得られる視線位置の平均と分散を用いることで大多数の人が注目する領域を矩形領域として算出し、その領域は字幕を配置すべきでないとする。これにより、元映像において鑑賞者が重要と感じる領域に字幕が入り込むことを防ぎ、元映像の重要な領域を俯瞰して鑑賞することが可能となる。

注目領域を算出するため、字幕が配置される区間における元映像に対する複数人の視線位置の平均と分散を算出する。時間 t における人物 i の視線位置を $\mathbf{p}_i(t) = (x_i(t), y_i(t))$ 、字幕の提示が開始される時間を t_s 、終了する時間を t_e とすると、 N 人の視線位置の平均 μ と分散 σ は以下の式によって算出される。

$$\mu = (\mu_x, \mu_y) = \frac{1}{N} \sum_{i=1}^N \sum_{t=t_s}^{t_e} \mathbf{p}_i(t) \quad (1)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N \sum_{t=t_s}^{t_e} (p_i(t) - \mu)^2} \quad (2)$$

このようにして得られた視線位置の平均 μ と分散 σ を用いて、平均 μ から $\pm 2\sigma$ の範囲の矩形領域を注目領域として定義する。（視線が正規分布に従うと仮定した場合、 $\mu \pm 2\sigma$ の範囲に約 95% の視線が収まることに相当する。）図 2 に 5 人分の視線追跡データ（2 秒間）を用いて注目領域（図中の矩形）を算出した例を示す。算出した注目領域の外部に字幕を配置することにより、元映像の重要な領域に字幕が入り込むことを防ぐことができる。

3.2 字幕配置

算出した注目領域を基に見やすい位置に字幕を配置する。提案手法では、3.1 節で算出した注目領域外の下部に字幕を配置する。これは注目領域が横長となった場合には

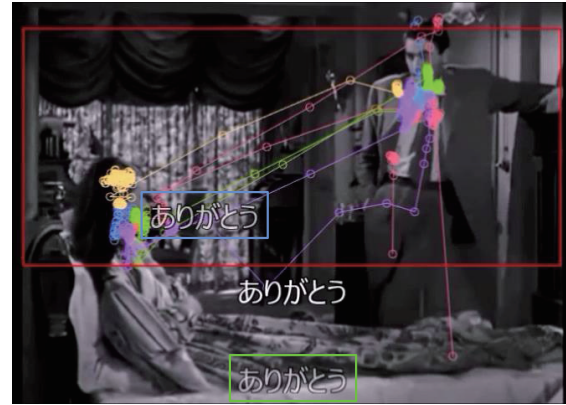


図 3 字幕位置の比較（白文字：提案手法、薄文字/緑矩形：固定字幕、薄文字/青矩形：既存手法）

注目領域の左右に字幕を配置する空間がないこと、注目領域の下部に字幕を常に配置するというルール設定を行うことにより鑑賞者に字幕の出現位置を予期させ、認知的負荷を軽減させることとの二つの理由に基づいている。例えば、注目領域が画面全体となった場合には固定字幕と同じ位置に字幕が表示されることになる。

図 3 に、提案手法によって配置された字幕と固定字幕及び、Hong ら [4] の方法に倣った字幕との比較を示す。ここで、Hong らの方法では話者付近の重要度を高く設定する重要度マップと顕著性マップを組み合わせることで、話者付近の顕著性が低い領域に字幕を配置している。図 3 において提案手法による字幕は赤い矩形で示される重要な領域の下部中央に、固定字幕は画面下部中央の緑矩形の中に、既存手法は画面左の人物横の青矩形の中に薄くそれぞれ表示している。話者付近に字幕を配置する手法を用いた場合には、画面左にいる話者の顔領域付近に字幕が配置されるため、注目したい領域に字幕が入り込んでしまう。一方、提案手法では注目領域の外部に字幕が表示されるため、鑑賞者は注目領域全体を俯瞰して見ることができる。

4. 評価実験

複数人の視線追跡データを用いて字幕を配置する方法の有効性を検証するため、主観評価実験を行った。具体的には、「様々なジャンルにおいて有効であるか」、「長時間の鑑賞において有効であるか」、「話者の顔領域付近に字幕を配置する方法と比較して有効であるか」の三つを検証した。被験者は 20 代男性 17 名、女性 3 名の日本語を母国語とする 20 名であった。

4.1 準備

実験では、アニメ、映画、スピーチ、音楽ライブ、ニュースの 5 つのジャンルから 15 秒間を切り出した映像を 2 つずつ、ドラマのジャンルから 5 分間を切り出した映像を 1 つ作成し、計 11 通りの映像を用いた。また、映像の音声

は全て英語であった。字幕内容や字幕を提示するタイミング、字幕を提示する時間の長さは、映像内容の理解において重要な役割を果たす。本実験では、予め字幕情報があるものは、これらの字幕情報をそのまま使用した。予め字幕情報がないものは、元映像における発話内容、発話開始時間、発話終了時間を基に作成した字幕情報を用いた。

元映像における注目領域を算出する際に必要な視線追跡データは、被験者5名に視線測定装置(EMR-9)を装着させた状態で元映像を鑑賞してもらい、60Hzで記録した視線測定結果を用いた。視線測定を行う際には、ディスプレイの大きさが字幕の見やすさに与える影響を考慮しなければならない[12]。そのため、本研究では大画面テレビや映画館の大きなスクリーンでの鑑賞を想定する。本実験では、41-inchの液晶ディスプレイを使用し、ディスプレイと被験者の間の距離は1.5mとして視線測定を行った。

これらの情報を用いて映像の注目領域を算出し、本手法による字幕付き映像を作成した。また、本手法が「様々なジャンルにおいて有効であるか」を検証するための比較映像としては、15秒間の映像に固定字幕を配置したものを作成した。同様に、「長時間の鑑賞において有効であるか」と「話者の顔領域付近に字幕を配置する方法と比較して有効であるか」とを検証するため、比較映像として5分間の映像に固定字幕を配置したもの、話者付近に字幕を配置したものをそれぞれ作成した。ここで、話者付近に字幕を配置する方法では、映像における話者の顔の位置を目視により決定し、Hongらの手法[4]に倣って字幕を配置した。字幕が長い場合には字幕を配置する十分な領域がないため、適切な位置に改行を施し可能な限り重要な領域と重ならないように話者の顔領域付近に字幕を配置した。また、話者が不在のシーン、話者の位置が大きく移動するシーンでは、固定字幕と同じ位置に字幕を配置するものとした。

4.2 評価方法

本手法と固定字幕との比較実験は20名に、本手法と固定字幕との比較及び、本手法と話者付近に字幕を配置する方法との比較実験は被験者それぞれ10名に対して行った。被験者には映像は同じであるが字幕の配置方法が異なる映像を続けて二本鑑賞してもらった後に、次の3つの質問項目に対して5段階評価してもらった。

問1. 映像に集中して鑑賞できたか

問2. 字幕は読みやすかったか

問3. 快適に鑑賞できたか

比較する二つの映像の順序はランダムに提示した。また、全ての映像に対する評価が終わった後に、どのような時に見やすいと感じたか、どのような時に見づらく感じたかという感想を自由記述で回答してもらった。

4.3 結果と考察

4.3.1 様々なジャンルに対する有効性

15秒の映像を用いた本手法と固定字幕との比較実験における問1、問2、問3それぞれに対する評価結果を図4、図5、図6に示す。各図において、質問に対するスコアは3を中心として、本手法が良いほどスコアは5に、比較手法が良いほどスコアは1に近づく。図4、図5、図6から分かるように、本実験で用いた多くの映像において、固定字幕よりも本手法による字幕の方が有効であったことが確認できる。特に、字幕の読みやすさ、鑑賞の快適さに関しては本手法に対する高い評価が得られた。また、被験者による自由記述では、「見ているものと字幕が近い位置にあると見やすかった」という感想が多く得られた。これは、本手法によって算出された領域が被験者の注目する領域を網羅していたためと考えられる。次に、全体を通して評価が良かったアニメ映像、評価が悪かった音楽ライブの映像について考察する。

アニメ映像に対する本手法の評価は全項目を通して高かった。これは実験で用いた映像が動きの少ないシーンであったために、字幕を読みつつ映像に集中できたためと考えられる。今後、動きの激しい映像に対しても本手法の有効性を検証する必要がある。音楽ライブの映像に対する本手法の評価は全項目を通して低かった。加えて、自由記述において「音楽のビデオでは曲が聞きたいため、字幕が動くことは鑑賞の邪魔になる。」という感想が多く挙がった。つまり、音楽ライブのように発話内容よりも映像内容を理解する方が重要な場合、字幕の移動が映像を鑑賞する上で不快さを与えてしまう。一方で、ミュージックビデオ等では文字にアニメーションを付与することにより歌詞の意味を効果的に提示する方法が広く使用されている。これは、文字を映像内容の一部として鑑賞されるためと考えられる。今後、文字をアニメーション化させた場合の本手法の有効性の検証が必要である。

その他に特筆すべき点として、スピーチ映像はアニメ映像と同様にシーン変化が少なかったにも関わらず評価がそれほど高くなかった。これは、スピーチにおいて発話内容が重要であることと、常に発話し続けるために字幕の文字数が多くなりやすいことから、字幕が動くことによって字幕を読みづらく感じたためと考えられる。今後、字幕の長さが長い場合には字幕位置の変化を小さくするといった工夫が必要である。

4.3.2 長時間の鑑賞における有効性

5分の映像を用いた本手法と固定字幕、本手法と話者付近に字幕を配置する方法との比較実験における問1、問2、問3それぞれに対する評価結果を図7に示す。本手法と固定字幕との「映像に対する集中のしやすさ」に対する比較では、5分映像を用いた平均スコアが15秒映像を用いた平均スコアより低かった。また、被験者の自由記述では、「固

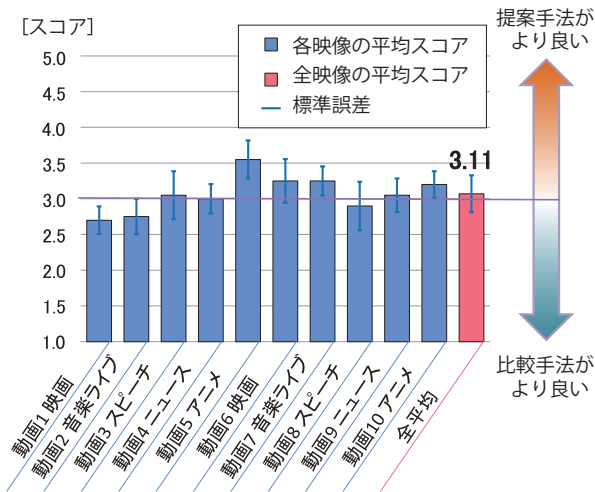


図 4 「映像に集中できたか」という質問に対する評価結果 (15 秒映像)

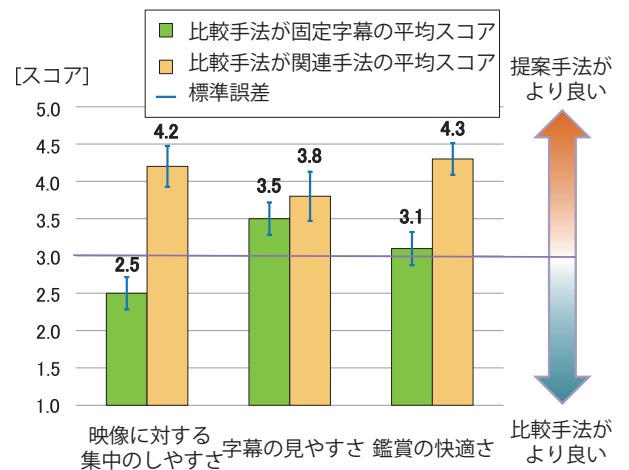


図 7 それぞれの比較対象に対する評価結果 (5 分映像)

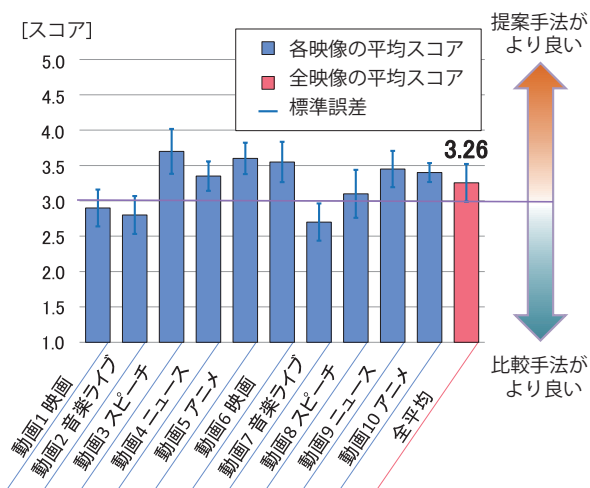


図 5 「字幕は読みやすかったか」という質問に対する評価結果 (15 秒映像)

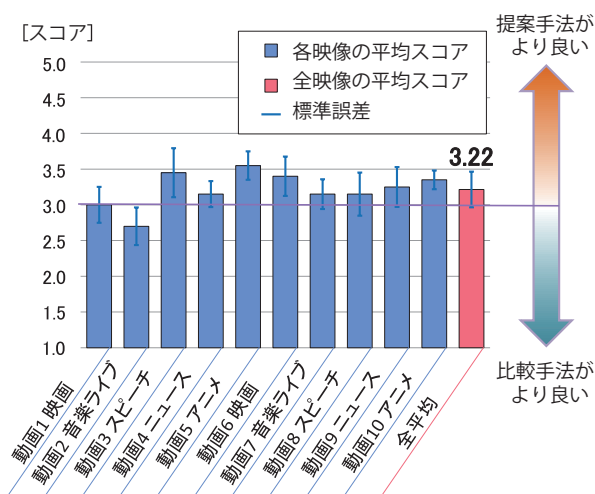


図 6 「快適に鑑賞できたか」という質問に対する評価結果 (15 秒映像)

定字幕に慣れているため、字幕が移動することに違和感があった」という感想が多かった。今後の課題として、長時間の鑑賞においても映像に集中できる動的字幕の提示方法の検討が必要である。

4.3.3 話者付近に字幕を配置する方法との比較

話者付近に字幕を配置する方法と比べ、本手法は全ての項目で高い評価が得られた。これは、以下の二つの理由により映像が見やすくなったためと考えられる。一つは、話者付近に字幕を配置する方法では、字幕位置の上下の変動が大きくなりやすかったのに対し、本手法では、常に注目領域の下部に字幕を配置した結果、字幕位置の上下の変動が小さくなりやすかったためである。被験者の自由記述でも、「字幕の高さが変わるときに見づらいと感じる」という感想があった。もう一つは、話者付近に字幕を配置する方法では字幕と話者付近の重要な領域が重なる場合があったのに対し、本手法では、注目される領域の外部に字幕を配置した結果、字幕が重要な領域に入り込むことを防ぐことができたためと考えられる。

4.3.4 本手法の課題

本手法の課題について字幕が見えづらいと感じたときの自由記述より考察する。

「字幕が頻繁に動くとき」や「字幕の高さが変化するとき」、「シーンの切り替わりが多いとき」に見づらいという意見が得られた。これらは、鑑賞者が注目する位置が頻繁に変化するため、不快に感じたと考えられる。したがって、注目されやすい領域の変化が大きい場合に字幕を一定時間固定して配置するといった工夫が必要である。

「シーンが変化していない場合でも多少字幕位置が変動するとき」に見づらいという意見があった。これは、本実験で注目領域を算出するための視線追跡データを5人分しか用いなかったためと考えられる。このため、不完全なキャリブレーションや瞬きによるノイズの影響を受けやすく注目領域が安定して得られなかったと考えられる。今

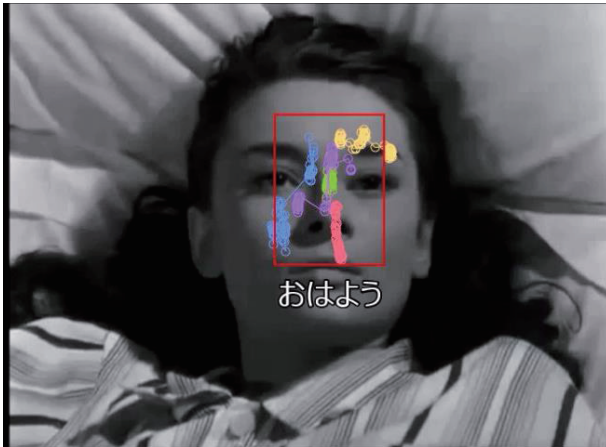


図 8 顔の上に字幕が配置される例

後、注目領域を安定して得るために必要な視線追跡データの人数についても検討が必要である。

「注目する人の顔の上に字幕が配置されたとき」に見づらいつい意見があつた。図 8 に、顔の上に字幕が配置されたときの例を 5 人分の視線追跡データと算出された注目領域と共に示す。本手法では、視線の分散が極端に小さかつた場合、算出される注目領域が小さくなるため重要な領域と字幕が重なってしまうことがある。今後、視線分散が小さい場合の字幕配置方法の検討が必要である。

他にも、「15 秒の映像より 5 分の映像の方が、ストーリーを追って理解できたため、自然と評価できた」という意見があつた。このため、様々なジャンルにおける有効性の検証においても、長い時間の映像を用いることで自然な評価が得られると考えられる。

5. まとめと今後の課題

本研究は、複数人の視線追跡データを用いて映像の中で視線が集まりやすい領域を算出し、その領域の下部に字幕を配置することで、視線移動の少ない字幕付き映像の鑑賞方法を提案した。字幕は高い視線誘導性を持っていることから、本手法は視線が集まりやすい領域へ鑑賞者の視線を誘導するという効果を持つ。本実験では、字幕の位置が字幕付き映像の鑑賞体験に与える影響について調査を行った。その結果、本実験で用いた多くの映像において、固定字幕と比較し本手法は映像と字幕との双方を快適に鑑賞できることが明らかになった。

本実験の目的は、字幕の配置の変化による字幕付き映像の鑑賞体験に与える影響を調査することである。ただし、字幕の見やすさという観点では、字幕の色やフォント、アニメーションによる影響を強く受ける。また、これらの影響力は、鑑賞者が映像と字幕のどちらをより重要視するかによって変化することが考えられる。字幕の文字としての可読性の評価は、今後の課題である。

本手法は事前に視線情報を用意する必要がある。しかし、

視線情報を一度用意すれば多くの人々にとって見やすい字幕付き映像ができることと視線追跡装置は普及しつつあることから、本手法の実用性は高まることが期待できる。視線情報を基に、最適な字幕提示方法の検討を行いより映像鑑賞に集中可能な動的に字幕を配置するシステムの構築を目指す。

謝辞 本研究の一部は、JST CREST「OngaCREST プロジェクト」の支援を受けた。また、河村俊哉氏、福里司氏、藤賢大氏、加藤卓哉氏（早稲田大学）には有益な指摘を頂いた。記して感謝します。

参考文献

- [1] 視聴覚障害者向け放送普及行政の指針: http://www.soumu.go.jp/menu_news/s-news/01ryutsu09_02000044.html
- [2] Brown, A., Jones, R., Crabb, M., Sandford, J., Brooks, M., Arm-strong, M., and Jay, C.: Dynamic Subtitles: The User Experience, *Proc. of ACM International Conference on Interactive Experiences for TV and Online Video*, pp.103–112 (2015).
- [3] Katti, H., Rajagopal, A. K., Kankanhalli, M., and Kalpathi, R.: Online estimation of evolving human visual interest, *ACM Trans. on Multimedia Computing, Communications, and Applications (TOMM)*, Vol. 11, Issue 1, No. 8(2014).
- [4] Hong, R., Wang, M., Xu, M., Yan, S., and Chua, T. S.: Dynamic captioning: video accessibility enhancement for hearing impairment, *Proc. of the IEEE International Conference on Multimedia*, pp.421-430(2010).
- [5] Hu, Y., Kaultz, J., Yu, Y., and Wang, W.: Speaker-Following Video Subtitles, *ACM Trans. on Multimedia Computing, Communications, and Applications (TOMM)*, (2015).
- [6] Jain, E., Sheikh, Y., Shamir, A., and Hodgins, J.: Gaze-Driven Video Re-Editing, *ACM Trans. on Graphics (TOG)*, Vol. 34, Issue 2, No. 21(2015).
- [7] Cao, Y., Lau, R. W., and Chan, A. B.: Look over here: Attention-directing composition of manga elements, *ACM Trans. on Graphics (TOG)*, Vol. 33, Issue 4, pp. 94 (2014).
- [8] Chun, B. K., Ryu, D. S., Hwang, W. I., and Cho, H. G.: An auto-mated procedure for word balloon placement in cinema comics, *In Advances in Visual Computing*, pp. 576–585(2006).
- [9] 小川直記, 岡部誠, 尾内理紀夫, 益子宗, 平野廣美: 配置と時間による可読性を維持した画像内文字提示手法, 第 22 回インタラクティブシステムとソフトウェアに関するワークショップ(WISS2014) 予稿集(ロングペーパー), pp.55–60(2014).
- [10] Wang, F., Nagano, H., Kashino, K., and Igarashi, T.: Visualizing video sounds with sound word animation, *Proc. of IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6(2015).
- [11] Fukusato, T. and Morishima, S.: Automatic depiction of onomatopoeia in animation considering physical phenomena, *Proc. of the Seventh International Conference on Motion in Games*, pp. 161–169(2014).
- [12] 中嶋崇大, 中林紗也, 井ノ上寛人, 川澄正史, 鉄谷信二: 画面サイズに応じた字幕表示位置の最適化に関する研究, 情報科学技術フォーラム講演論文集, Vol. 12, No. 3, pp. 219–220(2013).