

# アイトラッカで取得した視点情報と文書理解度の関係

吉村 和代<sup>†</sup> 川市 仁史<sup>††</sup> Kai Kunze<sup>††</sup> 黄瀬 浩一<sup>††</sup>

<sup>†</sup> 大阪府立大学工学部

<sup>††</sup> 大阪府立大学大学院工学研究科

〒 599-8531 大阪府堺市中区学園町 1-1

E-mail: †{yoshimura,kawaichi,kunze}@m.cs.osakafu-u.ac.jp, ††kise@cs.osakafu-u.ac.jp

**あらまし** 本稿では、アイトラッカで取得した視点情報を用いて、文書に対する理解度を推定する手法を提案する。本手法ではまず、アイトラッカで得られた視点情報に対し、眼球運動の性質を利用した特徴量を抽出し、8次元の特徴ベクトルを生成する。そして、得られた特徴ベクトルを用いて、読み判定をし、ユーザが読んでいるか読んでいないかを判別する。さらに、理解度推定をすることで、文書に対する理解度を推測する。被験者 10 人、10 文書を用いた実験の結果、読み判定については識別率 91%を得、有効性を確認した。一方、文書理解度推定については、5段階で表現される問題の正答率により理解度を表現したところ、平均二乗誤差 1.059 で正答率を推定するに留まり、十分な精度を得るに至らなかった。本稿ではその理由についても考察する。

**キーワード** アイトラッカ, 視点情報, 文書理解度

## 1. はじめに

近年、ライフログに関する多くの研究が行われている。ライフログとは人間の生活や行動、体験を長期間にわたり、画像や音声、位置情報などのデジタルデータとして記録する技術、あるいはその記録自体のことである。ライフログを分析することで、ユーザの趣味嗜好を推測することができ、各ユーザに効果的なサービスの提案、付加価値の向上が望めるとされている。スマートフォン等のデバイスの普及により、ユーザの撮影した画像を記録したり、歩いた距離を記録したりするものなど様々なライフログが提案されている。

このようなライフログの中に、ユーザが見た風景などの情報を記録するものがある。人は日常的にもものを見ることで多くの情報を得ている。したがって、人が見た情報を記録することで、より高度なライフログが実現可能であると考えられる。このようなライフログの実現には、ユーザが見た情報を記録するデバイスが必要となる。これを可能にするものとしてアイトラッカが挙げられる。アイトラッカはユーザの眼球運動を測定する装置である。アイトラッカを使用することでユーザがどこを見ているかを特定することができる。

ところで、人が目から得ている情報の中には、風景やものなどを「見る」ことによって得られる情報と、文字や文章を「読む」ことによって得られる情報がある。この中でも特に、「読む」ことによって得られる情報を記録することを、我々は「リーディングライフログ」と呼んでいる。リーディングライフログを用いてユーザが読んだ情報を記録、分析することができれば、ユーザの興味関心を推測することができ、個人に適したサービスの提供が可能となる。

リーディングライフログの分析目的の一つとして、文書に対

する理解度を推測するものが考えられる。これは読書時の眼球運動から、ユーザの読んでいる文書に対する理解度を推測するものである。眼球運動という測定可能なものから、理解度という測定が容易には可能でないものを判定することができれば、大変便利であると考えられる。

このような分析を実現するためには、(1) 読み判定: ユーザが読んでいるか読んでいないかを判定すること (2) 文書理解度推定: 理解度と視点情報の関係性を把握することの 2 点が必要である。そこで本稿では、これらの機能を実装する手法を提案する。視点情報を分析するために、まずユーザの眼球運動をアイトラッカを用いて記録する。次に、視点情報を一定時間で分割し、それぞれの視点情報を複数の特徴量で表す。得られた特徴量を用いて視点情報を分析し、読み判定、文書理解度推定をする。本手法の有効性を検証するために、読み判定、文書理解度推定の性能評価を行なった。その結果、読み判定については、識別率 91%を得た。文書理解度推定については、5段階で表現される問題の正答率により理解度を表現したところ、平均二乗誤差 1.059 で正答率を推定することができた。

## 2. 関連手法

本研究では、ユーザが読んでいるか読んでいないのかの判別と視点情報から文書に対する理解度の推測を目標とする。前者は行動認識の一種であり、「読む」という行動を認識するものである。

Bulling らは眼球運動から、ユーザが室内で椅子に座って行う行動を認識する手法を提案した [1]。この手法では、頻出する眼球運動のパターンを符号化することで、ユーザの行動を認識する。この手法は、眼球運動を利用してユーザの行動を認識するという点では我々の研究と共通する。しかし、本研究が目的

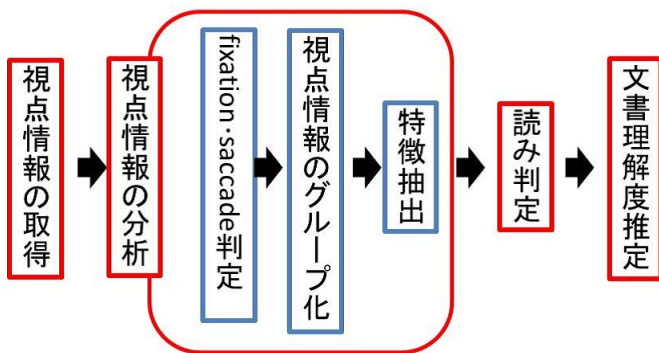


図 1 提案手法の流れ

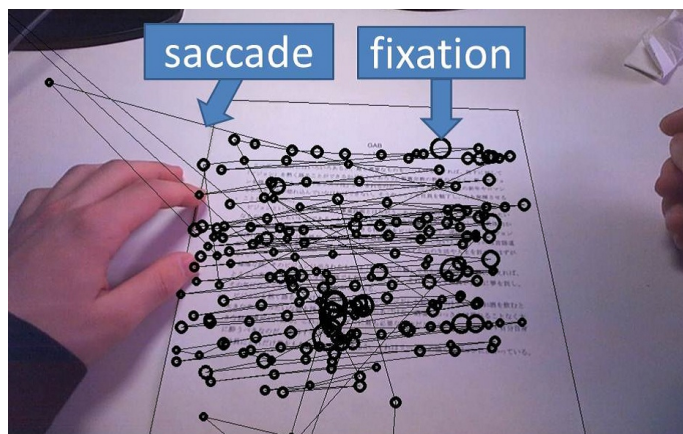


図 2 視点情報に対し、fixation-saccade 判定を施したものの

とする「読み始め」「読み終わり」といったような細かい動作を判別する場合には適用できない。

視点情報から文書理解度を推定する手法として、東中らが、ユーザの文書に対する理解度を自動的に測定する手法を提案する [2]。この手法は、文書閲覧中のユーザの視線から、ユーザが理解していないと思われる単語を自動推定する。この手法は電子文書中の単語領域に対する視線情報と、ユーザが辞書引きを必要とするかどうかを関連付ける関係式を機械学習を用いて作成する。東中らはこの手法を、電子文書の閲覧における意味理解支援に利用する。視点情報からユーザの理解を推定するという点は我々の研究と共通する。しかし、東中らの研究は単語に対する理解度を判定するものであり、本研究が目指す文書に対する理解度とは異なる。

文書の読み方の違いに関する研究に Campbell らの手法がある [3]。この手法は文書を読む際に見られる眼球運動を入力とし、文書を読んでいるか判定をする。特徴量としては方向と距離を利用しており、目の動きごとに点数が設定されている。点数の設定は経験則に基づいて決められている。

Campbell らの手法を改良したものに Buscher らの手法がある [4]。Buscher らは読んでいるか、流し読みかの識別手法を提案する。具体的には Campbell らの手法に、新たに流し読みの点数を設定することにより、読んでいるか流し読みかを識別する。

また、Biedert らはドイツ語で書かれた文書に対して、リアルタイム処理が可能で頑健な reading と skimming の識別器を提案する [5]。まずアイトラッカから得られた視点情報から、virtual median filter を用いてノイズを除去する。virtual median filter は対象となる点とその近傍の点から中央値を算出し、それを対象となる点に置き換えるものである。そして、視点情報を一定数でグループ化し、グループごとに特徴ベクトルを抽出する。これらの研究は読書時における眼球運動を利用する点においては我々の研究と類似する。しかし、読み方と文書に対する理解度の関連は明確になっておらず、読んでいるか流し読みをするかの判定は本研究の目的には適さない。

### 3. 提案手法

図 1 に提案手法の流れを示す。まず、ユーザの視点情報を取

得するために、アイトラッカを用いてユーザの眼球運動を記録する。次に、視点情報を一定時間で分割し、それぞれの視点情報を複数の特徴量で表す。このようにして得られた特徴量を用いて、読み判定、文書理解度推定をする。

#### 3.1 視点情報の取得

視点情報はアイトラッカを用いて取得する。アイトラッカとは、眼球運動を測定する装置のことである。アイトラッカは、非装着型のもので装着型のものに分けられる。

非装着型のアイトラッカは、内蔵されたカメラで顔全体を撮影することにより、ユーザの顔の向きと眼球運動などを測定するものである。そしてこれらの情報から、ユーザの視点位置を特定するものである。非装着型はディスプレイ一体型のもが多く、ユーザに負担を与えることなく眼球運動の測定が可能となっている。しかし、ユーザの行動を制限するなどの短所もある。

装着型のアイトラッカは、メガネ型のもが多く、内蔵されたカメラで眼球を撮影することにより、眼球運動を測定する。装着型は、頭部に装着するので場所を選ばずに眼球運動を測定することができる。また、装着型アイトラッカにはカメラを用いず、目の周りに電極を取り付けることにより、眼球運動の測定をするものもある。しかし、装着型は非装着型に比べ、精度が劣るという欠点がある。

本研究では、装着型のアイトラッカを使用することにより、ユーザの眼球運動を測定し、視点情報を取得する。その理由は、非装着型のアイトラッカはユーザの行動範囲を制限するため、ユーザの全ての視点を記録することはできないことである。しかし、リーディングライフログの性質上、ユーザの視点情報を全て把握する必要がある。そこで、本研究ではウェアラブルに記録可能な装着型アイトラッカを用いる。

#### 3.2 視点情報の分析

視点情報の分析は 3 つのプロセスに分類される。まず、視点情報に対し、fixation-saccade 判定をする。そして得られた fixation と saccade に対して、一定時間でグループにする。このようなグループごとに、特徴を抽出する。以下に各プロセスについて詳しく述べる。

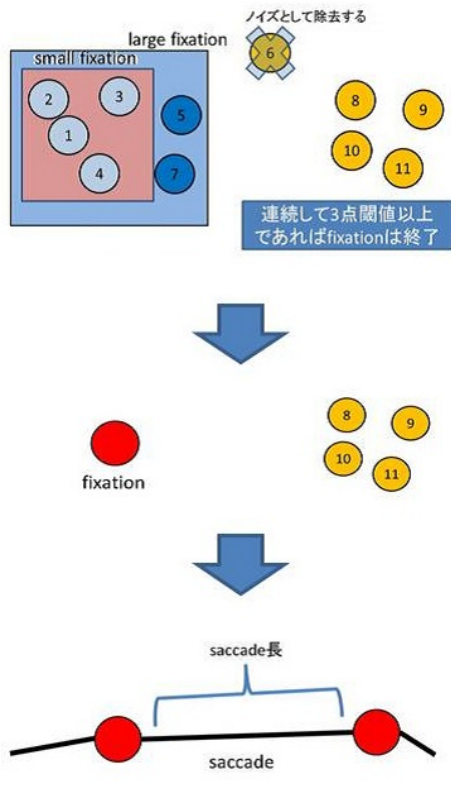


図3 fixation 判定

### 3.2.1 fixation-saccade 判定

図2に視点情報に対し、fixation-saccade 判定を施したものを示す。丸の大きさはfixationの持続時間に比例する。得られた視点情報に対し、fixation-saccadeを判定する。アイトラッカで得られる眼球運動は、blink, fixation, saccadeの3つに分類される。blinkとは、瞬きのことである。また、fixationは視点がある範囲内に一定時間停留することであり、saccadeはfixation間の素早い目の動きのことである。人は、上記にあげた3つの眼球運動を繰り返すことで、ものを見たり文書を読んだりする。本稿では、fixationとsaccadeに注目する。本手法では、fixationの判定手法にBuscherらの手法[6]を用いる。

図3にその手順を示す。始めに視点が全て互いに距離 $T_1$ 以内に連続して4点存在した場合、その4点をminimum fixationとする。この時、 $T_1$ は視点の中心間の距離とする。そして、その次の視点とminimum fixationの全ての視点との距離が閾値 $T_2$ 以内であれば、その点をlarge fixationとする。 $T_1$ と、 $T_2$ は、 $T_1 < T_2$ の関係にある。閾値 $T_2$ 以上である視点はノイズとして除去する。large fixationに含まれない点が連続で3点現れるとfixationは終了し、その最初の視点から新たにminimum fixation判定を始める。最終的には、fixationと判断された視点の重心をfixationの中心位置とする。そしてfixation間をsaccadeとする。

このようにして視点情報に対し、fixationを判定する。

### 3.2.2 視点情報のグループ化

図4に視点情報のグループ化を示す。連続するfixationと

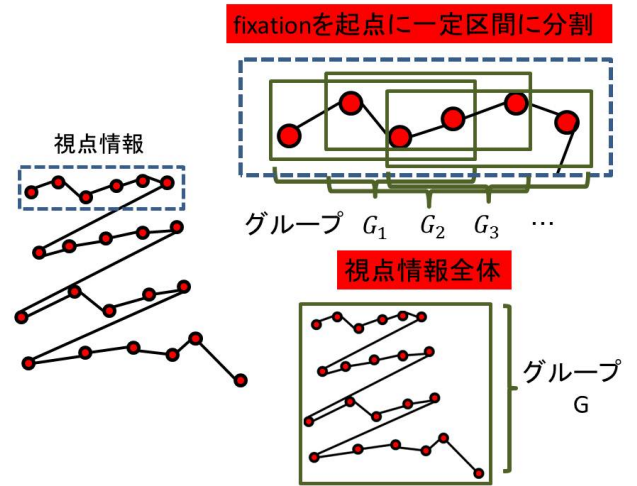


図4 視点情報のグループ化

表1 特徴量

関連するもの	詳細
fixationに関するもの	fixationの個数
	fixation持続時間の合計
	fixation時間時間の平均
saccadeに関するもの	saccade長の平均
	saccade長の最小値
	saccadeをベクトルとみなした際の合成ベクトルのx成分, y成分
ウェーブレット変換に関するもの	ウェーブレット変換後の振幅の大きさの平均

saccadeに対して、グループ化する。グループ化には2種類の方法を用いる。得られたfixationの列を $F = (f_1, f_2, f_3, \dots, f_N)$ とする。第一は、 $F$ を一定時間で分割するものである。まず、最初の $f_1$ を起点として、120フレームで分割する。そして、次に $f_2$ を起点として、120フレームで分割する。この処理を繰り返すことで、複数のグループを得る。第二は、 $F$ 全体を1つをグループ $G$ とするものである。上記にあげたグループから、それぞれ特徴を抽出する。

### 3.2.3 特徴抽出

眼球運動の特性を利用して8つの特徴量を考案する。人はfixationとsaccadeを繰り返すことでものを見たり読んだりする。そこで、fixationとsaccadeの位置や方向、出現割合などを求め特徴量とする。表1に使用する特徴量を示す。fixationに関するものには、fixationの個数、fixation持続時間の合計、fixation持続時間の平均を用いる。saccadeに関するものは、saccade長の平均長、saccade長の最小値がある。ここでsaccade長とは図3に示すfixation間のユークリッド距離である。また、saccadeをベクトルとみなした際の合成ベクトルの $x$ 成分、 $y$ 成分を用いる。これらに加えて、 $x$ 軸方向の視点情報に対し、一次元連続ウェーブレット変換を施して得た、振幅の大きさの平均を用いる。

上記8つの特徴量を用いて、グループごとに、8次元の特徴ベクトルを抽出する。

### 3.3 読み判定

ユーザが読んでいるか読んでいないかを判定する方法について説明する。上記の8次元の特徴ベクトルに対し、SVM(Support vector machine)によって学習、識別をする。この際、読んで



図5 ノイズ除去の様子

いる、読んでいないという2値識別をする。SVMではガウシアンカーネルを用いる。また、特徴ベクトル抽出のグループは分割したものを用いる。それぞれの特徴ベクトルから得られた識別結果をまとめ、多数決で全体の識別結果を決定する。

### 3.4 文書理解度推定

ユーザの文書に対する理解度を推定する方法について説明する。視点情報には、頭部の動きやキャリブレーションの失敗などによりノイズが見られる。そこでまず、ノイズを除去する。図5にノイズ除去の様子を示す。以下にノイズ除去の手順について説明する。人は文字を順を追って読んでいくため、眼球は水平方向への動きを繰り返す。そこで、saccadeの水平方向との角度がある閾値以上のものは、ノイズである可能性が高い。そこでそのようなsaccade S1に対し、saccade長を求め、閾値以上であるものをノイズとして除去する。またこの際、行を移動する眼球運動と区別するために、次のsaccade S2の長さを求める。そして、S1とS2のsaccade長の差が閾値以内であり、S1のsaccade長が大きければS1を除去する。これはノイズであれば、視点は元の読んでいた部分に戻ることを利用する。以上のようにしてノイズを除去する。また、ノイズとして除去し

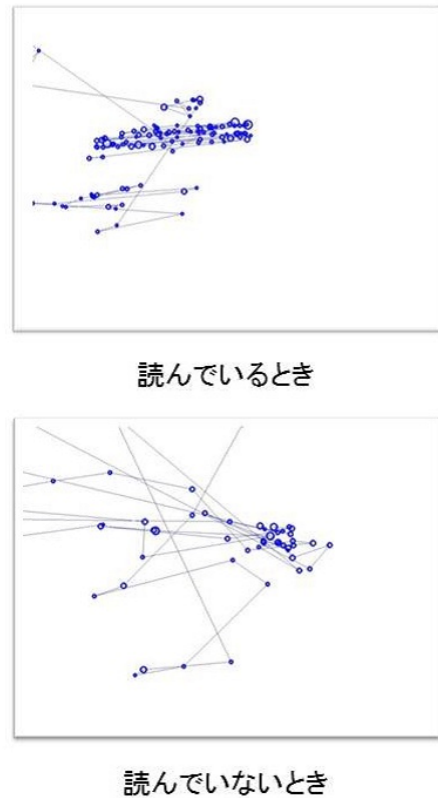


図6 読んでいるときと読んでいないときの視点の動き

た部分の前後のfixationを結び新たなsaccadeとする。

このようにして得られたfixationとsaccadeに対し、グループ化、特徴抽出をする。グループは、分割するものと全体のものの両方を用いる。

そして、特徴ベクトルに対し、SVR(Support Vector Regression)を用いて、理解度を推定する。この際、理解度は問題の正答率を用いて5段階で表す。SVRでもガウシアンカーネルを用いる。また、全体から特徴抽出したものに対しては、学習データ、テストデータ共に正規化をした。ここで正規化とは学習データの分布を用いて特徴量の各々を、-1から1の範囲にスケール変換することである。

学習によってSVRで生成された理解度推定のモデルに、テストデータの特徴ベクトルを入力して、推定された理解度を入力する。この際、1つの入力に対し、出力は1つである。分割して特徴抽出をしたものに関しては、特徴ベクトルが複数得られるため、1つの文書に対して複数の推定された理解度が得られる。そこで、推定された理解度を平均し文書に対する理解度とする。

## 4. 実験

本章では、読み判定の性能評価と、理解度推定の性能評価について述べる。10名の被験者を用いて、被験者ごとにデータセットを作成し、10-fold cross validationによって評価した。被験者は21歳から24歳までの大学生10名(男性9名、女性1名)である。その内、裸眼が6名、コンタクトレンズを装着者が3名、片目のみコンタクト装着者が1名であった。本実験で



表 2 読み判定の識別結果

被験者	1	2	3	4	5	6	7	8	9	10	平均
識別率 (%)	100	100	100	80	100	80	75	100	100	75	91



図 7 実験 2 の様子

は、SMI 社の装着型アイトラッカ iViewX ETG (Eye Tracking Glasses) を用いた [7].

#### 4.1 実験 1: 読み判定の実験

実験 1 では読み判定の性能を評価する。まず、椅子に座った状態でアイトラッカと装着し、キャリブレーションをすることで個人パラメータを設定した。そして、その状態で日本語の文書の黙読、休憩、黙読を繰り返し、視点の記録をした。その際、被験者には自然な状態で読むこと、合図をしたら黙読、休憩をするようにと指示を与えた。この際、文書の読み方には制約は与えていない。被験者 1 名に対し、黙読は 2 回、休憩は 3 回である。また、休憩ではものを取る、周りを見渡す、携帯電話を閲覧するなどの行動がみられた。得られた視点情報に対し、読み判定をした。

表 2 に被験者ごとの読み判定の識別結果を示す。識別率の最高値は 100%、最低値が 75%、平均値は 91% となった。また、読んでいるときと読んでいないときの視点情報を図 6 に示す。読んでいるときの視点情報は、狭い範囲内で水平方向の動きを繰り返すものが多く見られる。これは、文字を目で追っている動きを表すと考えられる。一方で、読んでいないときの視点情報は、saccade 長が長いものが複数見られる。また、広い範囲で視点の動きが見られた。これは、周りを見渡すときの視点の動きを表すと考えられる。上記の結果より、視点情報は読んでるか読んでいないかの判定に有効であることが示された。よって、読み判定の有効性が確認された。

#### 4.2 実験 2: 文書理解度に関する実験

実験 2 では文書理解度推定の性能を評価する。図 7 に実験風景を示す。実験 2 では、被験者に複数のタスクを与え、タスク実行中の被験者の眼球運動を記録した。タスクは日本語の文書を読み、文書に関する問題を解くというものである。また、被験者は、アイトラッカを装着し、椅子に座った状態で頭部をできるだけ動かさないようにという指示が与えられた。次に、タスクの内容について以下に示す。まず、被験者は制限時間内に

表 3 実験に用いた文書の行数と概略

文書番号	行数	概略
1	16	外国語の概念性
2	15	民族問題
3	16	古典
4	14	自意識
5	21	インターネットの普及による影響
6	12	病名の告知
7	13	北極の氷の融解
8	15	アメリカ人の国民性
9	14	明るい空間と暗い空間
10	19	リーダーの条件

表 4 制限時間

制限時間	長さ
1	通常
2	通常の半分
3	通常の 1/4

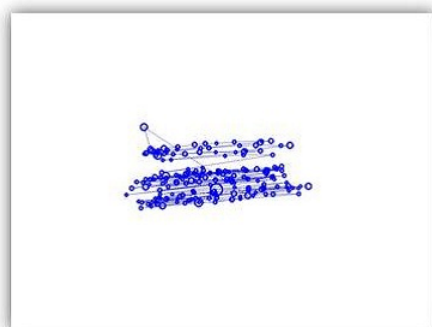
文書を読むように指示される。そして、制限時間が終了すると読んでいた文書は回収される。同時に問題文が実験者によって提示され、被験者は提示された問題に解答をする。被験者が問題に対し解答を終えたとき、被験者は実験者に合図をし、次の文書が提示される。問題数は 1 文書につき 4 問であり、それぞれ 3 つの選択肢が与えられている。その選択肢は全て、設問は本文から考えて (1) 正しい、(2) 正しくない、(3) どちらとも言えないの三者択一であった。用いた文書数は、10 文書である。実験に用いた文書内の概略と行数を表 3 に示す。また、文書に対する理解度に幅を持たせるために、3 種類の制限時間を設けた。表 4 に制限時間を示す。制限時間は文書ごとに設定した。設定の際、文書の行数を考慮し、難易度は考慮しなかった。被験者に実験前に練習問題を数問解いてもらい、制限時間に慣れてもらった。各問題解答後に、制限時間に対して被験者に評価してもらった (長い、ちょうどよい、短い)。得られた視点情報を、文書を読んでいる部分とそうでない部分に分け、文書を読んでいる部分に対して、理解度推定をした。理解度は設問に対する正解数とし、5 段階で表現される。

表 5 に被験者ごとの実験結果を示す。これは、SVR によって推定された理解度と正解との誤差を示す。アイトラッカのキャリブレーションが失敗していた被験者 5 の文書 8 は除いた。全体から特徴抽出したものでは、ノイズ除去していないもので最小となった二乗誤差は 1.059 であった。また分割したものでは、ノイズ除去したもので最小となった二乗誤差は 1.083 であった。また、ノイズ除去前とノイズ除去後の値を比較してみると、分割したものに関しては誤差は小さくなっていった。しかし、全体のものに関しては誤差が大きくなっている。この結果より、分割したものに対してはノイズ除去は有効だが、全体のものに対してはノイズ除去は有効でないということがわかった。

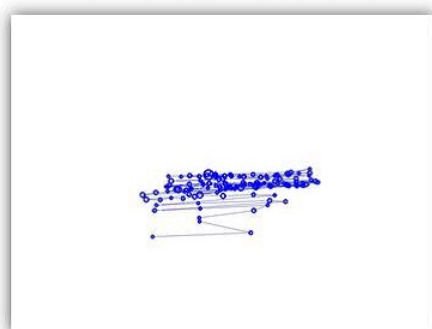
図 8 に同一被験者における理解度が高いときと低いときの視点の動きを示す。理解度が高いときには、saccade 長が短いものが多く見られる。また、理解度が高いときには fixation の密

表 5 実験 2 の被験者ごとの識別率 (%)

被験者	グループ	1	2	3	4	5	6	7	8	9	10	平均
ノイズ除去していないもの	120	0.576	0.901	0.810	1.309	2.071	0.611	1.587	0.575	2.211	0.769	1.142
	全体	0.485	0.989	0.541	1.645	1.990	0.749	1.080	0.570	1.786	0.758	1.059
ノイズ除去したもの	120	0.578	0.901	0.811	1.314	2.069	0.611	1.643	0.577	2.246	0.084	1.083
	全体	0.457	0.871	0.741	1.581	1.940	0.875	1.435	0.734	2.025	0.776	1.144
全てが定数	3	0.775	1.049	0.949	1.22	51.517	0.837	1.378	0.775	1.612	0.894	1.101



理解度が高いときの視点の動き



理解度が低いときの視点の動き

図 8 理解度が高いときと低い時の視点の動き

度が高い部分が複数見られた。

表 5 を見ると、被験者 9 は例外的に二乗誤差が大きいことがわかる。これは、被験者 9 の視線の動きがイレギュラーであり、理解度も他の被験者と比べて低かったことによる。

すべての理解度推定結果を 3 として平均二乗誤差を計算すると 1.101 となった。本実験で、最も平均二乗誤差が小さかったものは 1.059 であり、全ての理解度推定が 3 であったものと比較するとあまり差は見られないという結果になった。よって本手法では十分な推定精度を得られるにいたっていないということがわかった。その理由は以下の通りである。被験者の正答率の平均と標準偏差を求めると、それぞれ 2.50, 0.32 となった。これから、被験者の理解度にそれほど大きなばらつきがないことがわかる。このため、理解度を 3 と推定する結果と比べて十分な優位性を得ることができなかつたと考えられる。提案手法の有効性を十分検証するには、様々な理解度を持つ被験者を加えた実験を行う必要があると言える。

## 5. ま と め

本稿では、アイトラッカから取得した視点情報に対し、理解

度を推定する手法を提案した。本手法では視点情報に対し、読み判定をした後に文書理解度推定をする。性能評価によって、読み判定については高い識別率を得ることができた。しかし、文書理解度推定は十分な推定精度を得るにはいたっていないということがわかった。

今後の課題としては、長時間でのデータに対し、読み判定をすることが挙げられる。そして理解度推定では、視点情報以外の特徴を使用することが考えられる。また、被験者により個人差があることから、多くの被験者のデータを収集し、分析することも課題の 1 つである。

**謝辞** 本研究の一部は、JST CREST および日本学術振興会科学研究費補助金基盤研究 (B) (22300062) の補助による。

## 文 献

- [1] A. Bulling, J.A. Ward, H. Gellersen, and G. Troster, "Eye movement analysis for activity recognition using electrooculography," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.33, no.4, pp.741-753, 2011.
- [2] 東中竜一郎, 大野健彦, "視線に基づく文書理解度測定法とその応用," *情報処理学会研究報告. HI, ヒューマンインタフェース研究会報告*, vol.2003, no.9, pp.31-38, 2003.
- [3] C.S. Campbell and P.P. Maglio, "A robust algorithm for reading detection," *Proceedings of the 2001 workshop on Perceptive user interfacesACM*, pp.1-7 2001.
- [4] G. Buscher and A. Dengel, "Gaze-based filtering of relevant document segments," *Workshop on Web Search Result Summarization and Presentation. Workshop on Web Search Result Summarization and Presentation (WSSP-2009)*, located at in conjunction with WWW, vol.9, pp.20-24, 2009.
- [5] R. Biedert, J. Hees, A. Dengel, and G. Buscher, "A robust realtime reading-skimming classifier," *Proceedings of the Symposium on Eye Tracking Research and ApplicationsACM*, pp.123-130 2012.
- [6] G. Buscher, A. Dengel, and L. vanElst, "Eye movements as implicit relevance feedback," *CHI'08 extended abstracts on Human factors in computing systemsACM*, pp.2991-2996 2008.
- [7] <http://www.smivision.com/> 2013.