

多言語表示に対する視線情報を用いた第一・第二言語推定

王 琛[†] 浅井 洋樹[†] 山名 早人[§]

[†] 早稲田大学大学院基幹理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1

[§] 早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1

国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-1

E-mail: {magaret1103, asai, yamana}@yama.info.waseda.ac.jp

あらまし 近年、人工知能を例としたインテリジェント技術の進化とウェアラブルデバイスの普及とともに、視線情報を利用した研究が数多く行われるようになってきた。また 2020 年の東京オリンピック開催を控え、空港や街中ではデジタルサイネージといった、様々な言語圏の訪問者が理解可能な表示を要求される場面が増加している。しかし、こうしたデバイスでは、ユーザの明示的な表示言語選択無しに、ユーザ自身の理解できる言語を表示することは、これまで提案されてきた技術では困難である。ユーザが当該画面を見るだけでユーザの理解可能な第一・第二言語を自動検知できれば、非侵襲での自動言語選択を実現することができ、訪問者の満足度や広告効果が向上すると考えられる。視線情報を用いたこれまでの研究では、視線の移動軌跡や注視時間などを用いた英語習熟度の推定や、画像の興味領域の測定などが行われている。しかし、視線情報を用いた第一・第二言語の推定に関する調査は著者の知る限りまだ行われていない。そこで本研究では、ユーザが多言語で表示された画面を見るだけで視聴者の第一・第二言語を自動検知し、非侵襲での自動言語選択を実現する手法を提案する。

キーワード 視線情報, 多言語表示, 言語推定

1. はじめに

近年、人工知能のようなインテリジェント技術の進化と、Google Glassをはじめとしたウェアラブルデバイスの普及とともに、視線情報を利用した研究が数多く行われるようになってきた。2020 年の東京オリンピック開催を控え、日本語を母語としない多くの旅行者が日本を訪問することが想定される。このような背景のもと、空港や街中でデジタルサイネージ(図 1)に代表される多言語表示が増えている。

こうしたデバイスでは、内容を表示する前に、表示言語を選択する画面(図 2)を表示する場合が多い。ユーザが自身の理解できる言語を選択してから、内容が画面上に表示される。しかし、ユーザの明示的な表示言語選択無しに、ユーザ自身の理解できる言語を表示することはできない。ここで、ユーザが当該画面を見るだけで理解可能な言語を自動検知できれば、非侵襲での自動言語選択を実現することができ、デジタルサイネージやウェアラブル設備への応用は格段に広がると思われる。

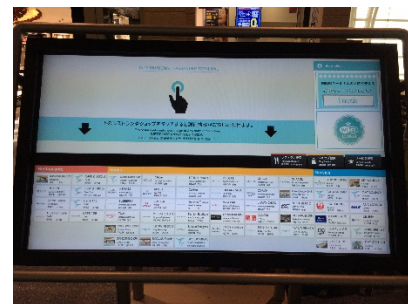


図 1 羽田空港でのデジタルサイネージ¹



図 2 表示言語を選択する画面²

「目は口ほどにものを言う」といわれるように、人間の視線はその人の思考を反映する。上記に述べた視線情報を用いて、目の動きと人間の思考に関する研究が行われている。例として、藤好ら[1]は英文問題解答

¹筆者撮影である

²筆者撮影である

時の視点情報を用いて、英語習熟度を推定している。Dixon ら[2]は女性の視線情報から好みの男性の体タイプを推定する手法を提案した。しかし、目の動きから第一・第二言語の推定に関する研究は著者らの知る限り存在しない。本研究は目の動きから理解できる言語を自動推定する手法を検討する。グローバリゼーションの推進とともに、二言語を理解し、話せる人も一般的になっている。本稿では、母語と第二言語を理解できる言語として定義し、視線情報から母語と第二言語を推定する可能性を検討する。

人間は、複数言語が同時に表示されている画面を見る瞬間に、無意識的に自身が理解できる言語を見つけて読むと考えられるため、自身が理解できる言語を最初に見つけて長く注視する可能性が高い。これを検証するために、複数言語が同時に表示される以下の複数表示方式を検討する：1)国旗・国章の固定表示，2)一単語の固定表示，3)一文全体の固定表示，4)一文スクロールの固定表示，5)ニュースリストの固定表示，6)一文全体の動的表示，7)一文スクロールの動的表示，8)ニュースリストの動的表示。

様々な表示画面を見ている視線を Eye Tribe³で読み取り、注視中に第一・第二言語を分析するシステムを構築する。評価実験では、異なる言語圏の被験者が構築したシステムを使い、各表示画面を見るよう指示し、Eye Tribe で読み取った視線情報から各自の理解できる言語を推定する。

本稿では以下の構成をとる。まず2節では視線情報を用いた関連研究を紹介し、3節で提案手法を説明する。次に4節で実験と評価を行い、最後に5節でまとめを述べる。

2. 視線情報を用いた関連研究

本研究では、視線情報を用いて視聴者の第一・第二言語を推定することを目標とする。しかし、視線情報を用いた研究は多く行われているが、視線情報を用いて理解できる言語の推定に関する調査は著者の知る限りまだ行われていない。以下では、視線情報を用いて人の思考や好みを推定する関連研究を紹介する。

2.1. 視線情報を用いた英語習熟度の推定

藤好ら[1]は題量が少ない英文問題解答時の視線情報から、ユーザの英語習熟度を推定している。装着式アイトラッカーで英文問題と解答を読んでいる時の視線情報を取得する。取得したアイトラッカーのシーンカメラ上の視点情報を LLAH (Locally Likely Arrangement Hashing) による文書画像検索で文書画像上の視点座標に変換する。そして眼球運動の特性を利

用した35個特徴量を抽出し、線形回帰で読者の英語習熟度を推定して TOEIC スコアとして評価する。推定から得られたスコアと本人の TOEIC スコアと比較した差が ± 35 以内であれば推定成功と定義している。推定結果として、被験者9名中6名の推定に成功した。

この手法は、非リアルタイムでユーザの英語習熟度を推定する。しかし、本研究は応用上から考えると、デジタルサイネージやウェアラブル設備への応用にはリアルタイムの推定が適当であるため、この手法をそのまま適用することは困難である。

2.2. 視線情報からの好みの体系推定

Dixon ら[2]は女性の視線情報から好みの男性の体タイプと体の部位を推定する研究を行った。男性の体タイプを Mesomorphs, Ectomorphs と Endomorphs 三種類、各タイプに7エリアに分けて、それぞれの写真複数枚を女性の被験者に見せる実験を行った。被験者が各写真を見ている5秒間の視線情報を取得し、最初の1秒間と以後の4秒間で注視点の数と注視時間を分析して、好みの男性の体タイプと部位を推定している。結果として、女性全体の視線から見ると、Mesomorphs タイプが好みになることが多いことと、女性は男性の全身を見るタイミングでは背中の上半部と肩に注目することが多いことがわかった。

この研究は一人の好みを推定するのではなく、収集したデータの統計を解明するための研究であるため、非リアルタイムの手法を使用している。

2.3. 視線情報から注目語抽出に基づく検索意図のリアルタイム推定

梅本ら[3]はユーザがウェブページを閲覧する時の視線情報を用いて、注目している対象語に基づく検索意図をリアルタイムで推定する研究を行った。梅本らは「注目度の高い単語ほど、ユーザの検索意図に適合している」ことを仮定し、それを検証するために視線情報をもとに常に注目語を抽出できるプロトタイプを構築した。

ユーザがことプロトタイプで表示したウェブページを閲覧している間に、実際に注目されている語彙の同定を繰り返し行い、tf-idfによりページ中の各語に対する注目度を計算する。そして得られる各閲覧ページにおける注目語集合を1つ集約することで、ユーザの検索意図をリアルタイムで推定する。語彙の注目度を計算するために、梅本らは MGT(Multiply Gaze by Tf), nMGT(normalized MGT), DGT(Divide Gaze by Tf), nDGT(normalized DGT)の4モデルを提案し、ベースラインとしての TF モデルと比較手法 GAZE モデルとともに合計6種類の手法間で推定精度を比較した。結果として、MGT と nMGT モデルは比較的に高い精度での意図推定が可能であるが、DGT と nDGT モデルの

³ <https://theeyetribe.com/>

推定精度はベースラインや比較手法より低いことがわかったと報告している。

本研究で用いられているような視線情報の利用方法は、様々な分野で利用されているが、被験者の第一・第二言語を推定する研究は、著者が知る限り存在しない。

3. 提案手法

本節では、提案手法である、多言語表示に対する視線情報を用いた第一・第二言語推定手法について述べる。

3.1. 概要

本研究では「自身が理解できる第一・第二言語を最初に見つけて長く注視する可能性が高い」と仮定する。これを検証するために、複数言語が同時に画面に表示させる。ユーザが多言語表示画面を見る視線を Eye Tribe で読み取り、時系列の視点座標が得られる。視線情報としての時系列注視点を一定時間間隔で切り分け、言語別に注視の時間割合を算出する。画面全体をスキャンする視線移動を含め、注視時間割合の第一ピーク、すなわち最初に一番長く停留する言語が母国語、第二ピークが第二言語になると仮定する。

3.2. 多言語表示

提案手法では、多言語で記述された単語や文、あるいは画像を表示し、被験者の第一・第二言語を推定する。また、表 1 に示すように静的な固定表示だけではなく、動的表示方式も考える。固定表示とは、一画面で表示された内容を固定し、変化はしない方式である。動的方式では一画面で表示された内容は一定時間で変更し、変更前と変更後の位置も変化するので、読んでいけば視線が移動することになる。その視線の移動を検出することで第一・第二言語を推定することを狙う。

表 1 多言語表示リスト

固定表示	動的表示
● 国旗・国章(図 3) ● 一単語(図 4) ● 一文全体(図 5) ● 一文スクロール(図 6) ● ニュースリスト(図 7)	● 一文全体 ● 一文スクロール ● ニュースリスト

一単語(図 4)で表示された単語は Google Trends⁴からピックアップした意味が理解しやすい単語を使用する。一文表示では、意味を持ち、かつ、わかりやすい文が求められるため、グローバルなニュースサイトからピックアップしたニュースのタイトル文を使用する。全体表示(図 5)は一文全体をそのまま表示し、スクロ

ール表示(図 6)は一文の一部のみ表示して左から右へスクロールする。ニュースリスト表示(図 7)では、画面を複数ブロックに分けられ、一ブロックでは一種言語の複数ニュースタイトルを箇条書きで表示する。

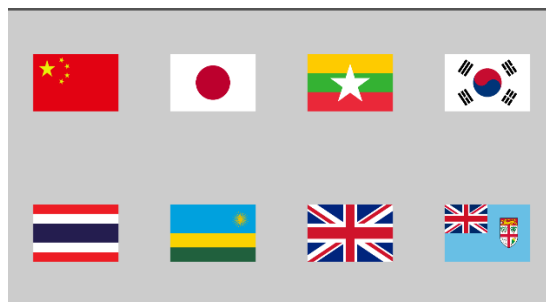


図 3 国旗表示の例

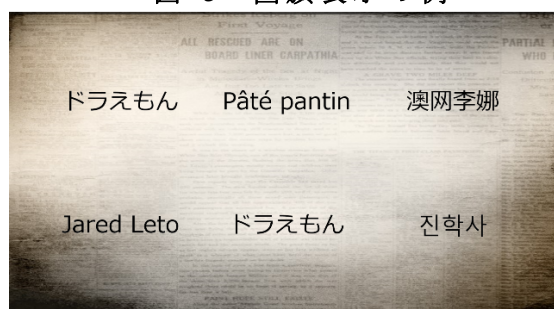


図 4 一単語表示の例

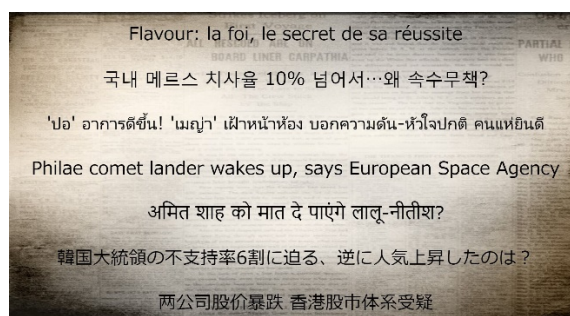


図 5 一文全体表示の例

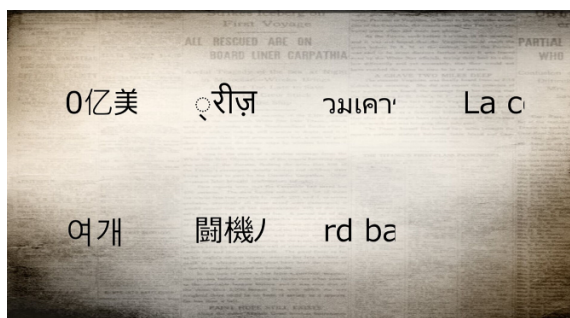


図 6 一文スクロール表示の例のスクリーンショット

⁴ <https://www.google.com/trends>

1. 「女性アイドルだけ総選挙」1位は球団第1. 메리츠에 송 미혁이 가운데 속 타는 한국경 사시. 4키70cm 2. スマホやゲームがもたらす視力影響は 2. NIA, 영국 데이터 혁신 전문기관과 맞손...데 3. 適切な睡眠はスナックのママになるべき!? 이터 사업 협력 논의 4. 人気占い師が助言 3. 국산 중형차보다 저렴한게... "나, 외제차 타 는 사람이야" 5. <青龍連続殺人>結婚相談所で資産家探し 4. C조소영, 전국 당일배송...배송전정 철회 見合いは100人 5. "눈 중 잘 주는 게 중... 최고 일 딱 내려오고 있어요"	1. Lack of diversity' among Scottish school heads 2. US actor John Stamos arrested for drunk driving 3. US tackles crowdfunding man who spent money on himself 4. Jeb Bush: Foreign policy advisers seek to prevent 'unforced errors' 5. Oculus and Xbox create virtual reality tie-up
1. ข่าวสาร: สถานการณ์โควิด-19 ในประเทศไทย 2. ข่าวสาร: สถานการณ์โควิด-19 ในประเทศไทย 3. ข่าวสาร: สถานการณ์โควิด-19 ในประเทศไทย 4. ข่าวสาร: สถานการณ์โควิด-19 ในประเทศไทย 5. ข่าวสาร: สถานการณ์โควิด-19 ในประเทศไทย	1. 遠征中国: APEC首脑“低承诺高兑现”? 2. 远征中国: APEC首脑“低承诺高兑现”? 3. 远征中国: APEC首脑“低承诺高兑现”? 4. 远征中国: APEC首脑“低承诺高兑现”? 5. 远征中国: APEC首脑“低承诺高兑现”? 1. CEDEAO: sécurité et bonne gouvernance 2. L'attribution du Mondial 2026 retardée 3. Usine de chocolat en Côte d'Ivoire 4. Ordonnance d'un tribunal contre el- Béchar 5. Musique: David Jenkins, le zoulou blanc

図 7 ニュースリスト表示の例

3.3. 推定手法

3.3.1. 視線測定

本研究では、被験者に対する負担を無くするため、ウェアラブルな視線測定装置ではなく、設置式の装置 Eye Tribe を使用する。ユーザが上記の多言語表示画面を見る視線を Eye Tribe で読み取り、Eye Tribe の API により視線情報を画面上の座標へと変換し、時系列の視点座標を得る。

3.3.2. 注視点処理

人はどこかに注視する時、しばしば注視点外に視線を移すことがあるが、このような不安定な視線データは、推定精度を落とす原因となるため、前後の視点との距離が 50mm 以上ずれているような視点座標は削除することとする。次に、被験者が文を読んでいるかどうかを判定する。文の途中までしか読まない場合や読む方向が文字方向と一致しない場合、被験者は当該文の記述言語を読んでいるわけではなく、好奇心で見ている可能性が高いと仮定する。

人間の眼球運動は注視 (fixation) とサッカード (saccade) の二種類に分類される [4]。注視とは、視点がある範囲内で一定時間とまることである。サッカードは注視点間を移動する目線の動きである。人が何かを見ている時、目は注視とサッカードを繰り返して動いていく。注視点の領域や注視時間、サッカードの速さと距離などは視線情報となるため、時系列の視点座標から注視とサッカード判定を行う。判定方式は Buscher ら [5] の手法を使用する。注視とサッカード判定を行ったら、各言語に注視時間、注視点数、サッカードの速さと長さが得られる。

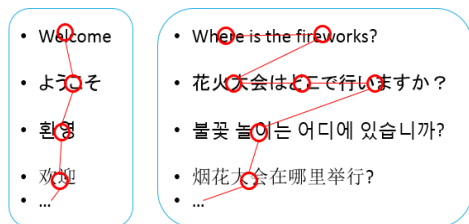


図 8 停留注視点の例 (赤丸：注視点)

3.3.3. 注視時間割合の算出

時系列の視点座標を一定時間間隔 (500ms) で切り分け、言語別に時間間隔 (500ms) を占める割合を算出する。図 9 は注視時間のタイムラインの例 (被験者 1 名) を表し、図 10 は一定時間間隔内での注視時間割合の例を示している。画面全体をスキャンする視線を含め、注視時間割合の第一ピーク、すなわち最初に一番長く停留する言語が母国語、第二ピークが第二言語になる可能性が高いと考えられる。

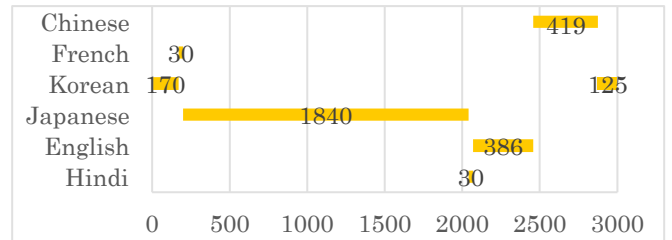


図 9 注視時間タイムラインの例 [ms] (日本人)

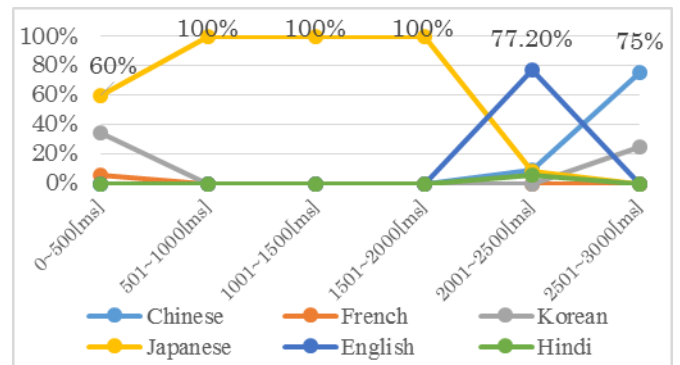


図 10 注視時間割合の例 (日本人)

3.3.4. 視線移動の検出

各画面で固定表示による推定を行うと同時に、画面の切り替えによる動的表示からも推定する。

- ① 被験者が一画面をスキャンし、理解できる言語を読んでいる間に固定表示による推定を行う。
- ② 5 秒後、画面を切り替える。この時、表示方式は変えず、各言語の内容と表示場所のみを変更する。
- ③ ①に戻る。

このような画面切り替え前後で、推定言語が一致すれば、被験者の第一・第二言語の推定の精度が高いと判断することができる。

4. 評価実験

本節では、第 3 節で述べた提案手法を評価する。

4.1. 実験内容

本手法の評価を行うため、8 名の大学生・院生 (日本人 3 名、中国人 2 名、韓国人 1 名、ルワンダ人 1 名、ミャンマー人 1 名) を被験者とし、第一・第二言語推定実験を行った。実験では、置き式視線測定装置 Eye

Tribe をディスプレイの下部に置き、被験者をディスプレイの前に座って、ディスプレイで多言語表示を見る。視点校正のため、実験が始まる前に、各被験者がそれぞれキャリブレーションを行う。実験後被験者の本当の第一・第二言語をヒアリングにより記録する。本手法の各表示方式で推定した第一・第二言語の結果を正解である第一・第二言語と比較し、一致する時の推定正解率をそれぞれ算出する。

実験環境として、今回の評価実験は CPU が Intel(R) Core(TM) i7-4770 CPU @3.4GHz, メモリが 16.0GB, OS が 64bit Windows 7 のコンピュータで実施した。

4.2. 表示内容

一文表示で使用するニュースタイトルの出典サイトを表 2 に表示する。各言語はそれぞれのニュースサイトから 50 個ずつタイトルを記録する。一文表示では毎回各言語ランダムで 1 個表示する。ニュースリストの場合は毎回各言語ランダム 5 個箇条書きで表示する。各表示方式における表示内容と位置を表 3 で示す。一単語、一文表示、ニュースリスト表示はそれぞれ 5 回表示するため、固定表示による推定はそれぞれ 5 回、動的表示で視線の追いかけによる推定はそれぞれ 4 回行うことになる。表示時間の長さに関して、被験者に対象画像を見せてから 5 秒以内の視線が最初の意味を表す可能性が高いので、本研究では被験者が言語を選択して見る意思を推定するために、すべての表示方式に対して各回の表示時間を 5 秒と設定する。

表 2 各言語の出典ニュースサイト

言語	出典ニュースサイト
Chinese	BBC Chinese (http://www.bbc.com/zhongwen/simp)
English	BBC (http://www.bbc.com/)
French	BBC French (http://www.bbc.com/afrique)
Hindi	BBC Hindi (http://www.bbc.com/hindi)
Japanese	Yahoo Japan (http://news.yahoo.co.jp/)
Korean	Naver (http://www.naver.com/)
Thai	Thairath (http://www.thairath.co.th/)

表 3 各表示方式における表示内容と位置

表示方式	表示言語数	固定表示回数	各回表示位置	各回表示内容
国旗	8	2	random	--
国章	8	2	random	--
一単語	6 ⁵	5	random	50 件中 random で 1 件
一文全体	7 ⁶	5	random	50 件中 random で 1 件
一文スクロール	7 ⁶	5	random	50 件中 random で 1 件
ニュースリスト	6 ⁵	5	random	50 件中 random で 5 件

4.3. 実験結果

実験結果としては、各表示方式における固定表示による第一・第二言語の推定正解率は図 11 と図 12 に示す。図 13 と図 14 ではそれぞれ第一と第二言語における判定までの時間と平均正解率の関係を表す。

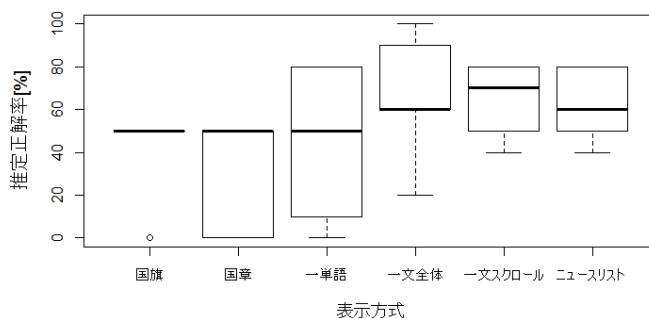


図 11 固定表示による第一言語推定正解率

⁵ Chinese, English, French, Japanese, Korean, Thai
⁶ Chinese, English, French, Hindi, Japanese, Korean, Thai

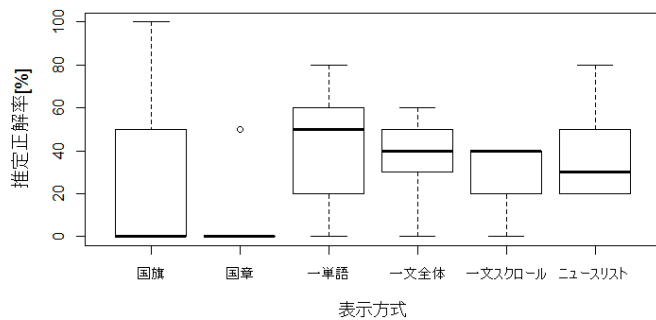


図 12 固定表示による第二言語推定正解率

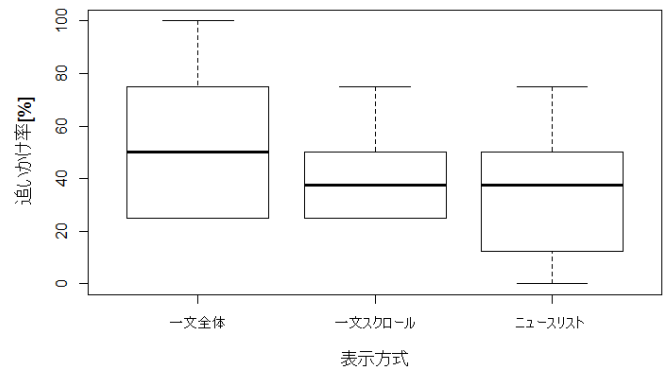


図 15 視線の追いかかけ率

4.4. 考察

4.3 からわかるように、各表示方式において第一言語の推定正解率が第二言語より高くなっている。母語の推定では、一文全体表示の推定正解率が一番高く 67.5%となる。図 13 から画面を表示してから 1500ms では推定正解率が低いことがわかる。原因としては、画面を表示してから 1500ms までは人の視線が画面全体をスキャンし、理解できる言語を探しながら読んでいる視線であることがわかる。また、4500ms 以上を見ると、推定正解率が下がることから、理解可能な言語を読み終えて、好奇心で他の言語を長く注視してしまうことが考えられる。

動的表示では、一文スクロール表示とニュースリスト表示は画面のスキャンの時間が長くなり、表示内容が変更された後はすぐに追いかかけしにくいいため、追いかかけ率が一文全体表示より低くなる。

5. まとめ

本稿では、ユーザが複数種類の多言語表示画面を見る視線情報を用いて、各言語に注視時間や視線の追いかかけを検出することによって、非侵襲的にユーザの第一・第二言語をリアルタイムで推定する手法を提案した。

また今回の実験はデジタルサイネージではなく、パソコンのディスプレイで行った。実際のデジタルサイネージでは、ユーザが立ったまま見る場合が多いため、その時の視線から推定が難しくなる。また、機械学習やパターン認識などの手法を用いて推定正解率を向上することも今後の課題になる。

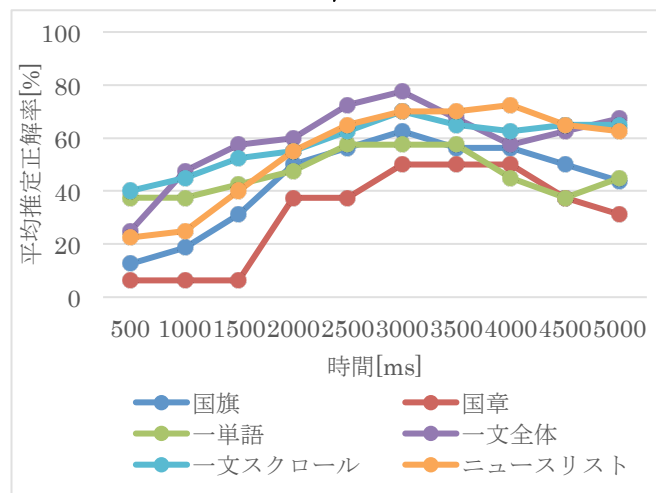


図 13 第一言語における判定までの時間と平均正解率の関係

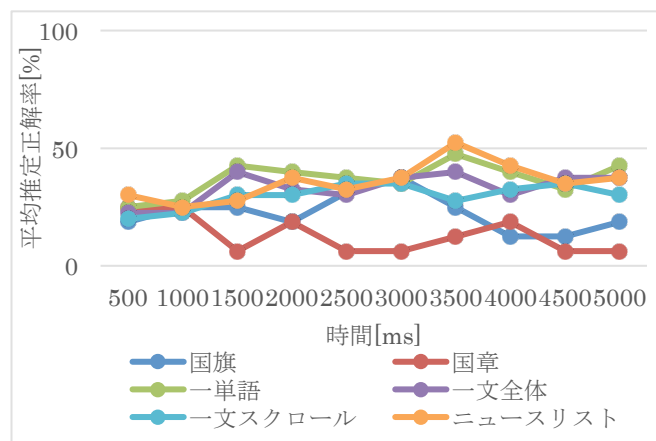


図 14 第二言語における判定までの時間と平均正解率の関係

3.3.4 で述べたように、本研究では各画面で固定表示による推定を行う同時に、画面の切り替えによる動的表示から母語を推定する。各表示方式において 4 回ずつ表示内容変更による 4 回動的表示の中に、視線追いかかけを行った率は追いかかけ率として、図 15 に示す。

参考文献

- [1]. 藤好宏樹, 吉村和代: "英文問題解答時の視点情報を用いた英語能力推定法", 電子情報通信学会技術研究報告: 信学技報 (パターン認識・メディア理解), Vol.115, No.24, pp.49-54 (2015)
- [2]. Dixon, Barnaby J., et al.: "Eye-tracking women's preferences for men's somatotypes," Evolution and

Human Behavior, Vol.35, No.2, pp.73-79 (2014)

- [3]. 梅本和俊, 山本岳洋, 中村聡史, 田中克己: "視線情報からの注目語抽出に基づく検索意図のリアルタイム推定", 情報処理学会論文誌(データベース), Vol. 6, No.3, pp. 120-131 (2013)
- [4]. Keith Rayner: "Eye movements in reading and information processing: 20 years of research," Psychological bulletin, Vol. 124, No.3, pp.372-422 (1998)
- [5]. Georg Buscher and Andreas Dengel: "Gaze-based filtering of relevant document segments," Proc. of International World Wide Web Conference (WWW), 2009.