第二言語の映像視聴時の認知負荷を考慮した字幕提示手法の提案

西 優己[1]　中村 優吾[2]　Billy Dawton[2]　福嶋 政期[2]　荒川 豊[2]

概要：映画やテレビ番組などを見る際，字幕は映像の内容理解を促進する役割を担っている．本研究では，第二言語の映像視聴時において認知負荷を考慮した字幕提示手法を提案する．一般的な字幕は画面下部に提示されるが，字幕と映像の視線の切り替えが多く発生し，認知負荷の要因になっている．提案手法では，認知負荷を減らし，映像集中を妨げないための字幕提示として，字幕の提示位置，提示時間，情報量に着目した．評価実験では，22 歳から 25 歳の 12 名の参加者に英語の映像を視聴してもらい，提案手法が映像視聴時の視線分布，理解度，認知負荷に与える影響を調査した．その結果，一般的な字幕に比べて提案手法による字幕では画面中央付近に視線が分布していることが分かった．映像の理解度は一般的な字幕に比べて劣るものの，同程度の理解が保たれたことが示唆された．しかしながら，映像視聴時の認知負荷は一般的な字幕よりも大きく，参加者は提案手法による字幕で負荷の大きい状態で映像を視聴していたことが確認された．

Proposal of Subtitle Presentation Method Considering Cognitive Load during Second Language Video Viewing

YUKI NISHI[1]　YUGO NAKAMURA[2]　BILLY DAWTON[2]　SHOGO FUKUSHIMA[2]
YUTAKA ARAKAWA[2]

1. はじめに

第二言語の映画やテレビ番組を理解する上で，母国語の字幕は重要な役割を担っている．外国語の映像を視聴する際，母国語の字幕が内容理解に有益であることが報告されている [1] が，字幕と映像の視線の切り替えが眼精疲労の原因となったり，字幕を注視することにより映像内の重要なシーンを見逃したりする可能性がある [2] などのデメリットも存在する．そのため，ユーザが字幕付きの映像を視聴する際の認知負荷を適度に減らしつつ，映像に集中できる状態を作ることが望ましい．

これに対し，映像内の話者付近に字幕を表示させる話者追従型字幕 [3] が提案されている．話者追従型字幕では，画面下部に提示された一般的な字幕に比べて映像視聴時の視線が画面中央付近に集中し，字幕よりも話者に注意が向けられたことが報告された．しかし，字幕の位置が動的に変化することから，一般的な字幕よりもテキストを探すのに苦労したことが報告されている．この問題は，字幕を動的に変化させず，話者に近い位置で固定することで解決できると考える．また，提示する分量や時間に着目した，Rapid Serial Visual Presentation(RSVP)[4] がある．RSVP は高速逐次視覚提示とも呼ばれ，単語などを一定の速度で連続的に提示する手法である．この手法は，スマートグラスを用いた読書 [5] などで用いられており，映像内の字幕にも適用できるのではないかと考えた．さらに，音声の内容が分かっている場合でも，テキストを読む傾向があることが分かっており [6]，視聴する人にとって不要な情報を減らすことが望ましいと考える．

そこで，本研究では，第二言語の映像視聴時の字幕の提示位置，提示時間，情報量の 3 つに着目し，映像集中を妨げないための字幕提示手法を提案した．また，英語の映像を日本語字幕で視聴することを想定している．提案手法では，一般的な字幕に比べて情報量を削減し，提示時間を短くした字幕を画面中央に提示した．字幕の提示時間は，各

[1] 九州大学大学院システム情報科学府
[2] 九州大学大学院システム情報科学研究院

語 300ms としている．字幕の情報量は，文章でなく単語単位で提示し，かつ提示する語を語彙レベルに応じて選定することで削減した．実験では，12 名の参加者に 3 つの字幕タイプで映像を視聴してもらい，提案手法が映像視聴時の視線分布，理解度，認知負荷に与える影響を調査した．1 つ目は，画面下部に文章単位で提示される一般的な字幕である．2 つ目は，画面中央に単語単位で提示される提案手法による字幕である．3 つ目は字幕がない状態での視聴である．視線データはスクリーンベースのアイトラッカーである Tobii Pro Nano[*1]で計測し，視線分布を表すヒートマップを作成した．映像の理解度は，映像視聴後に実施する理解度テストの点数で評価し，認知負荷は NASA-TLX(NASA Task Load Index)[7] による参加者の主観的な評価とした．実験の結果，映像視聴時の視線分布に関して，一般的な字幕は画面下部に集中し，提案手法では画面中央に分布が多かったことから，映像に集中できていることが示唆された．映像の理解度は一般的な字幕と比較して低かったが，字幕無しに比べて有意に高い結果となった．認知負荷は一般的な字幕に比べて大きく，参加者は集中して視聴する必要があることが分かった．

本稿の構成は以下の通りである．2 章では字幕提示手法に関する関連研究を示し，3 章で提案手法について示す．4 章で実験について説明し，5 章で実験結果を示す．最後に 6 章でまとめと今後の展望について述べる．

2. 関連研究

字幕提示手法に関する研究は様々なものがある．Kurzhals ら [8] は，字幕を映像中の話者付近に配置する話者追従型字幕 (Speaker-following Subtitles) について研究している．話者追従型字幕では，画面下部に提示された一般的な字幕に比べて映像視聴時の視線が画面中央付近に集中し，字幕よりも話者に注意が向けられたことが報告された．しかし，字幕の位置が動的に変化することから，一般的な字幕よりもテキストを探すのに苦労したことが報告されている．提示する文量や時間に着目した Rapid Serial Visual Presentation(RSVP)[4] に関する研究も行われている．RSVP は高速逐次視覚提示とも呼ばれ，文章を分割したものを同じ位置に提示する手法である．Rzayev ら [5] は，スマートグラスを用いた読書におけるテキスト位置と提示方法について研究している．座った状態で読書をした場合に，文章単位で提示される方法と比べて RSVP 条件でのテキスト理解度が高かったことが報告された．Erin ら [9] は，スマートウォッチを用いた読書における RSVP の影響について調査している．RSVP 条件での読書は，一般的な方法と比較して同程度の理解度が得られたが，満足度に関するアンケートでは一般的な方法が上回った．したがって，認知負荷を考慮したストレスの少ない字幕提示手法が必要とされる．

3. 提案手法

本研究では，映像集中を妨げない字幕提示手法として，字幕の提示位置，提示時間，情報量の 3 つに着目した手法を提案する．提案手法では，一般的な字幕に比べて情報量を削減し，提示時間を短くした字幕を画面中央に提示した．字幕の提示位置に関して，一般的な字幕は画面下部に提示されるが，提案手法では画面中央とした．字幕の提示時間は，各語 300ms に統一し，字幕の情報量は，文章単位でなく単語単位で提示し，かつ提示する語を語彙レベルに応じて選定した．提示する語彙の選定には New Word Level Checker[*2]というツールを用いた．New Word Level Checker は，日本人英語学習者に向けて提供されているツールの一つであり，入力された英文に含まれる語彙のレベルを分類するものである．本ツールの用途として，教師が英語教材の難易度を評価したり，学生が書いた英文の語彙カバー率を調査したりすることなどが挙げられている [10]．本ツールに登載されている語彙リストとして New JACET8000, JET2000，SVL12000 などがある．今回は CEFR（Common European Framework of Reference for Languages）[11] の語彙リストを使用した．CEFR は，外国語の運用能力を測る国際基準であり，6 段階 (A1，A2，B1，B2，C1，C2) に分類されている．A1 が最も基礎的なレベルとされ，C2 はネイティブレベルに相当するとされている．実験用の映像に含まれる単語を New Word Level Checker で分類したところ，約 6~7 割が A1 に分類され，C1，C2 に該当する単語は含まれていなかった．今回は，最も基礎的な A1 レベルの単語を除外し，A2 レベル以上の単語 (A2，B1，B2) を提示対象とし，語彙の選定を行った．A2 レベルでは，簡単な文章を理解したり，日常的な範囲での会話ができるが，外国語のメディア (映画やラジオなど) を視聴するのは難しいとされている．映像の内容理解に必要だと想定される語彙のみ残すことで，情報量を減らしつつ，従来の字幕と同程度の理解度を保てると考えた．このように，理解が容易だと想定される語彙の情報量を削減し，提示時間を短くした字幕を中央に提示することで，字幕の役割である内容理解の促進に加え，映像の視聴をより促すことができると考えた．図 1 に，一般的な字幕と提案手法による字幕レイアウトの比較を示す．著作権の関係で作品そのものの画像は記載しておらず，生成した類似の画像を記載している．図 1a の一般的な字幕では，字幕が文章単位で画面下部に提示されており，図 1b では，字幕が単語単位で画面中央に提示されていることが分かる．また，表 1 に今回作成した字幕の情報量の比較を，表 2 に一

[*1] https://www.tobii.com/products/eye-trackers/screen-based/tobii-pro-nano

[*2] https://nwlc.pythonanywhere.com/

(a) 一般的な字幕

(b) 提案手法による字幕

図 1: 字幕レイアウトの比較

度に提示される字幕の比較を示す．表 1 は，実験で使用する映像に提示される日本語字幕の合計字数を手法ごとに比較したものである．情報量の項目は，提案手法の文字数を通常字幕の文字数で割り，算出している．提案手法では，提示する字幕の字数が通常字幕の 3 割程度になり，情報量が削減されている．表 2 は，各映像において，一度に提示される字幕の字数を比較したものであり，字幕の合計字数 (表 1 参照) を提示回数で割ることにより算出している．提案手法では，通常字幕の半分以下である，約 3 文字の字幕が一度に提示される．

表 1: 作成した字幕の情報量の比較

映像	通常字幕 (字)	提案手法 (字)	情報量 (%)
1	292	107	36.6
2	306	77	25.2
3	288	81	28.1

表 2: 一度に提示される字幕

映像	通常字幕 (字)	提案手法 (字)
1	7.7	3.2
2	8.7	3.1
3	7.6	3.0

4. 実験

提案する字幕提示手法が映像視聴時の視線分布，理解度，認知負荷に与える影響を調査するため，評価実験を行った．

4.1 仮説

字幕の提示位置，提示時間，情報量に着目した提案手法が，映像視聴時の視線分布，理解度，認知負荷に影響を与えるという仮説を立てた．

- **仮説 1: 提案手法による字幕では，一般的な字幕に比べて映像に集中できる．**
  提案手法では字幕を画面中央に固定しており，画面下部に提示される字幕に比べて話者との距離が近く，映像に集中できると考える．
- **仮説 2: 提案手法による字幕では，一般的な字幕と同程度の理解度を保つことができる．**
  提示する字幕を語彙レベルに応じて選定することにより，情報量が従来の 3 割程度に削減されているが，一般的な字幕と同程度の理解度を保つことができると考えられる．
- **仮説 3: 提案手法による字幕では，一般的な字幕に比べて映像視聴時の認知負荷を減らすことができる．**
  字幕の提示時間を短くし，画面中央に提示することで，一般的な字幕に比べて認知負荷の低い状態で映像を視聴することができると考える．

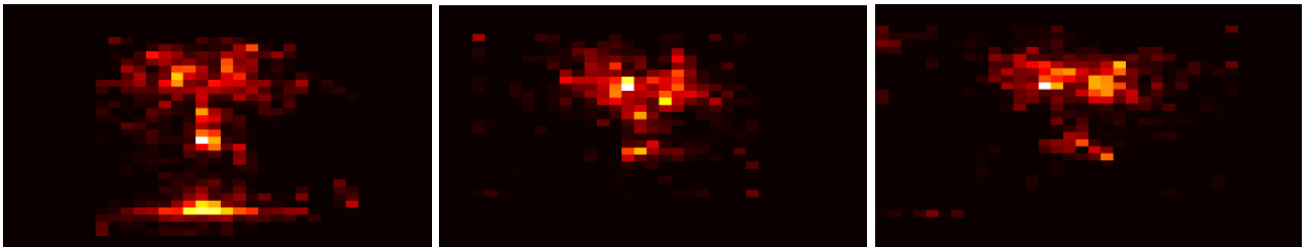
4.2 実験内容

実験は 22 歳から 25 歳の 12 名 (男性 10 名，女性 2 名) に協力していただいた．参加者は英語の映像を視聴し，その後映像の内容に関する設問に回答した．実験条件は，画面下部に提示される一般的な字幕 (通常字幕)，画面中央に提示される提案手法による字幕，字幕なしの 3 つに設定

表 3: 実験で使用した映像

映像	タイトル	時間 (分: 秒)
1	Rachel Falls off the Balcony	2:25
2	Phoebe Gives Monica a Haircut	2:03
3	Monica Won't Be Chandler's Girlfriend	2:17

図 2: 実験の様子



(a) 通常字幕　(b) 提案手法　(c) 字幕無し

図 3: 映像視聴時の参加者の視線分布

した．実験の様子を図 2 に示す．主観的な認知負荷の評価には NASA-TLX(NASA Task Load Index) を用いており，参加者は，英語の映像を各条件で視聴し，理解度テストに回答するというタスクについて NASA-TLX の各項目を評価した．実験にはアメリカのコメディドラマ作品であるフレンズの映像を使用し，2 人または複数人の人物が会話しているシーンを切り取ったものを使用した (表 3)．映像は，時間が 2 分から 2 分半程度の長さで WPM(Words Per Minute) が比較的同じものを選定している．実験手順について示す．始めに，参加者は椅子に座り，視聴する映像と提示される字幕の説明を受けた．次に，アイトラッカーのキャリブレーションを行い，視線データが計測できていることを確認した．その後，イヤホンを装着した状態で映像を視聴し，映像に関する理解度テスト (4 択 10 問) と NASA-TLX(7 段階リッカート尺度) に回答した．視線データは Tobii Pro Nano で収集し，視線分布を表すヒートマップを作成した．映像の視聴と回答を 1 セットとしこれを 3 回行った．各参加者が視聴する映像と提示する字幕の組み合わせはランダムにしている．参加者は最後にアンケートに回答した．実験にかかった時間は 1 人あたり 30 分程度であった．

5. 実験結果

評価実験の結果を視線分布，映像の理解度，認知負荷ごとに示す．本実験において視線分布は被験者間比較とし，映像の理解度，認知負荷は被験者内比較としている．

5.1 視線分布

Tobii Pro Nano で計測した視線データから視線の分布を示すヒートマップを作成した．映像視聴時の参加者の視線の分布を図 3 に示す．色が明るい部分に視線がより集中していることを示す. 通常字幕では画面中央付近に加え，画面下部に視線が集中していることが分かる．一方，提案手法による字幕では，字幕無しの場合と分布が比較的似ており，画面の中央付近に視線が集中している．したがって通常字幕に比べてより映像に集中できることが示唆された．

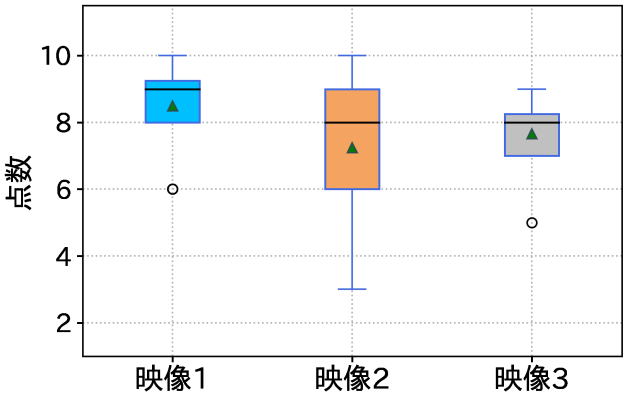


図 4: 映像ごとの理解度

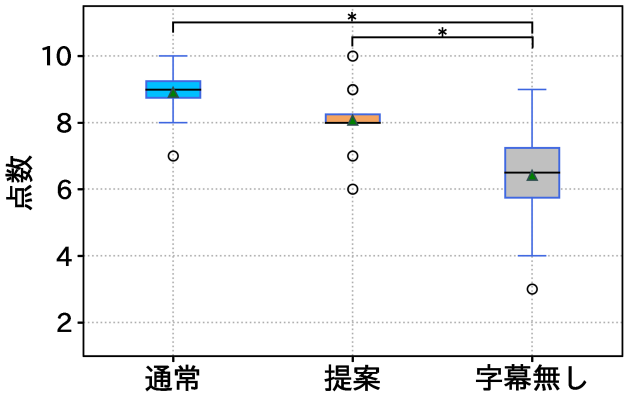


図 5: 手法ごとの理解度

5.2 映像の理解度

理解度テストの結果を図 4，図 5 に示す．図 4 は映像ごとの比較，図 5 は手法ごとの比較である．図中の緑の三角形は平均値，黒線は中央値を表しており，$*$ は有意水準 5%で理解度に差があることを示している．映像ごとの理解度テストの点数は映像 1(M=8.5)，映像 3(M=7.7)，映像 2(M=7.3) の順に高い結果となった．一元配置分散分析 (ANOVA) を実施したところ，映像間の理解度に有意差は生じなかった ($F(2, 33)$=1.65, p >.05)．したがって，映像による難易度調整は問題ないと考えられる．

手法ごとの理解度テストの点数は通常字幕 (M=8.9)，提案手法 (M=8.1)，字幕無し (M=6.4) の順に高い結果となっ

た．一元配置分散分析 (ANOVA) を実施したところ，手法間の理解度に有意差が見られた ($F(2,33)$=11.96, p <.01)．Bonferroni 法により，調整化された有意水準 α=0.0167 を求めた (補正前 p=0.05)．Bonferroni の多重比較検定を実施した結果，通常字幕と字幕無し，提案手法と字幕なしの条件間で有意差が確認された．提案手法による字幕では，字幕無しに比べて理解度が有意に高かった．また，通常字幕と提案手法の条件間での有意差は生じなかったことから，提案手法は通常字幕に劣るが，同程度の理解度が保たれたと考える．

5.3 認知負荷

タスク時の主観的な認知負荷の結果を図 6 から図 8 に示す．図中の緑の三角形は平均値，黒線は中央値を示しており，$*$ は有意水準 5%で差があることを示している．図 6 は知覚的負荷を示したものであり，数値が高いほど，見る，考える，記憶するなどの負荷を要したといえる．参加者の主観的な知覚的負荷は字幕無し (M=5.9)，提案手法 (M=5.0)，通常字幕 (M=3.3) の順になった．一元配置分散分析 (ANOVA) を実施した結果，手法間の知覚的負荷に有意差が生じた ($F(2,33)$=9.15, p <.01)．Bonferroni 法により，調整化された有意水準 α=0.0167 を求め (補正前 p=0.05)，Bonferroni の多重比較検定を実施した結果，通常字幕と提案手法の条件間，通常字幕と字幕無しの 2 条件間で有意差が生じた．提案手法による字幕では，通常字幕に比べてタスク時にかかる知覚的負荷が大きく，集中して映像を視聴する必要があったと考える．

図 7 は，参加者がタスクの達成にどの程度努力を要したかを示したものであり，数値が高いほど，英語の映像を視聴し，理解度テストに回答するというタスクの達成に苦労したといえる．参加者の主観的な努力は提案手法 (M=4.7)，字幕無し (M=4.5)，通常字幕 (M=2.8) の順になった．一元配置分散分析 (ANOVA) を実施した結果，手法間のタスクに要した努力に有意差が生じた ($F(2,33)$=6.56, p <.01)．Bonferroni の多重比較検定を実施した結果，通常字幕と提案手法，通常字幕と字幕無しの 2 条件間で有意差が生じた．提案手法による字幕では，通常字幕に比べてタスクの達成に相当な努力を要したことが分かる．

図 8 は，参加者がタスク時に感じたストレスを示したものであり，数値が高いほど，ストレスが大きかったといえる．参加者による主観的なストレスは字幕無し (M=3.9)，提案手法 (M=3.8)，通常字幕 (M=2.3) の順になった．一元配置分散分析 (ANOVA) を実施した結果，手法間のタスク時のストレスに有意差が生じた ($F(2,33)$=5.59, p <.05)．Bonferroni の多重比較検定を実施した結果，通常字幕と提案手法，通常字幕と字幕無しの 2 条件間で有意差が生じた．提案手法による字幕では，通常字幕に比べてタスク時にかなりのストレスを感じたことが確認された．

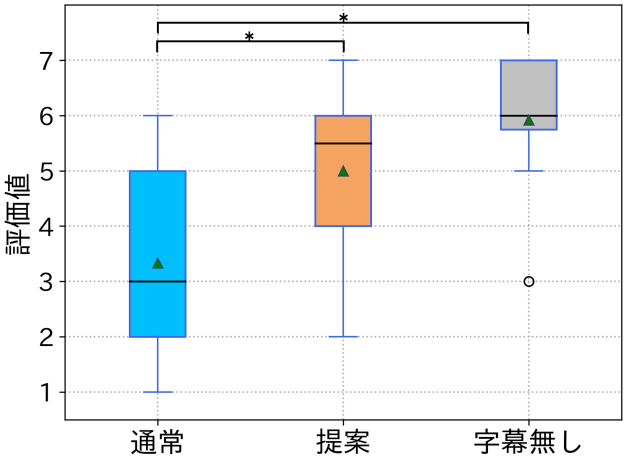


図 6: 知覚的負荷

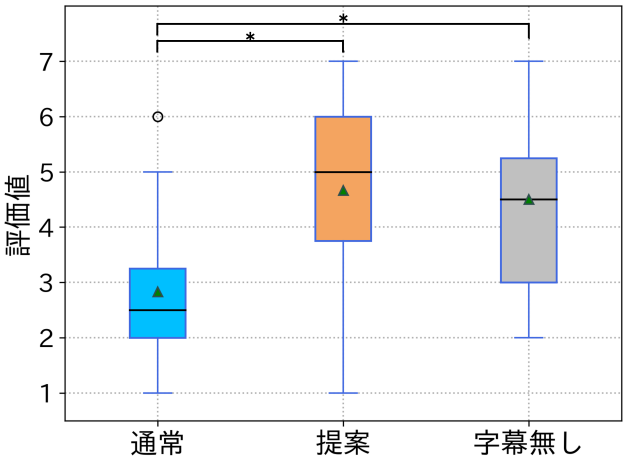


図 7: タスク達成に要した努力

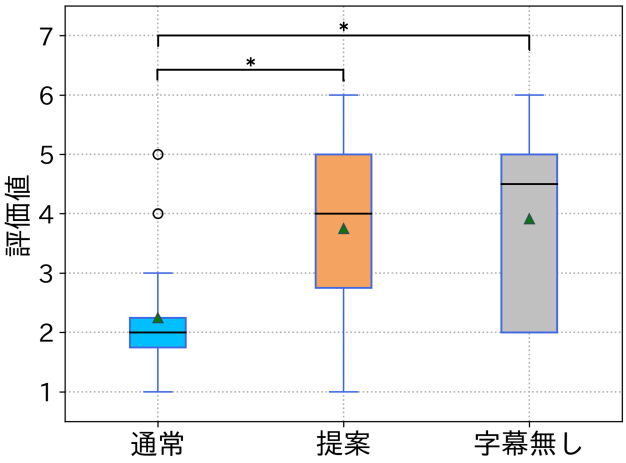


図 8: タスク中に感じたストレス

5.4 アンケート

実験後に実施したアンケートの結果を示す．今後洋画などを視聴する際，実験で用いた手法 (通常の日本語字幕，提案手法による日本語字幕，字幕無し) のどれを用いたいか質問したところ，通常字幕が 9 名，提案手法が 2 名，字幕

無しが 1 名と回答した．参加者から得られた意見や感想として，「普段の字幕に慣れているので，新しい字幕は受け入れにくかった」，「映像を理解するのが難しかったので，もっと情報量が多いと理解の助けになると感じた」，「提案手法ではいつ文字が出るのか分からないのでとても頭を使った」などがあった．一方で，提案手法を選んだ参加者からは，「字幕があるとつい見てしまうため，難しい単語のみ補助してくれるのは良い」という意見が得られた．

6. おわりに

本研究では，第二言語の映像視聴において，字幕の提示位置，提示時間，情報量に着目し，映像集中を妨げない字幕提示手法を提案した．評価実験では，提案手法による字幕提示が映像視聴時の視線分布，理解度，認知負荷に与える影響を調査した．以下に，本研究で設定した仮説に対する結論を示す．

- **仮説 1: 提案手法による字幕では，一般的な字幕に比べて映像に集中できる．**
  一般的な字幕の場合，画面下部に最も視線が集中していた．一方で，提案手法では画面中央付近に視線が集中し，字幕なしと分布が類似していたことから映像の内容に集中できていたことが示唆される．
- **仮説 2: 提案手法による字幕では，一般的な字幕と同程度の理解度を保つことができる．**
  映像の理解度は高い順に一般的な字幕，提案手法，字幕無しであった．検定の結果，提案手法では，字幕無しの場合より理解度が有意に高く，一般的な字幕と提案手法の条件間では理解度に有意な差が生じなかった．したがって，提案手法では一般的な字幕と同程度の理解度を保つことができたと考える．
- **仮説 3: 提案手法による字幕では，一般的な字幕に比べて映像視聴時の認知負荷を減らすことができる．**
  知覚的負荷，タスク達成に要した努力，タスク中に感じたストレスの全ての項目において，提案手法が一般的な字幕より高い結果となった．字幕を話者に近い位置で固定し，ユーザの負担を減らそうと試みたが，提案手法では映像視聴時の認知負荷を減らすことはできなかった．

課題と今後の展望について述べる．本研究では，認知負荷を考慮した字幕提示手法を提案したが，多くの参加者は認知負荷が高い状態で映像を視聴していた．その要因の一つとして字幕の提示時間が短かったことが考えられる．今回，字幕の提示時間を各 300ms に統一したが，一部の参加者が字幕を見逃すことがあったため，今後の実験では字幕の提示時間を長くする予定である．また，提示する語彙の選定に CEFR を用いたが，実験に用いたフレンズの映像には話し言葉やスラングなどが含まれていた．そのため，単語自体は比較的簡単であっても，連語など理解が難しい語が提示対象にならず，一部の参加者が難しく感じた要因であると考えられる．したがって，視聴する人が欲しい情報を提示できる指標を模索し，語彙を選定する必要がある．また，話し言葉の多い映像だけでなく，堅い雰囲気のスピーチなど，映像の特徴によって手法の有効性が異なると想定される．今後は別の映像や語彙選定の指標を用い，字幕提示手法について引き続き調査を進めていく予定である．

謝辞 本稿で示した研究の一部は，Society 5.0 実現化研究拠点支援事業 (グラント番号: JPMXP0518071489) の支援のもと実施されている．

参考文献

[1] G. Dizon and B. Thanyawatpokin, “Language learning with netflix: Exploring the effects of dual subtitles on vocabulary learning and listening comprehension,” *Computer Assisted Language Learning*, vol. 22, no. 3, pp. 52–65, 2021.

[2] L. Bergen, T. Grimes, and D. Potter, “How attention partitions itself during simultaneous message presentations,” *Human Communication Research*, vol. 31, no. 3, pp. 311–336, 2005.

[3] Y. Hu, J. Kautz, Y. Yu, and W. Wang, “Speaker-following video subtitles,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 11, no. 2, pp. 1–17, 2015.

[4] M. C. Potter, “Rapid serial visual presentation (rsvp): A method for studying language processing,” in *New methods in reading comprehension research*. Routledge, 2018, pp. 91–118.

[5] R. Rzayev, P. W. Woźniak, T. Dingler, and N. Henze, “Reading on smart glasses: The effect of text position, presentation type and walking,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–9.

[6] G. d’Ydewalle, J. Van Rensbergen, and J. Pollet, “Reading a message when the same message is available auditorily in another language: The case of subtitling,” in *Eye movements from physiology to cognition*. Elsevier, 1987, pp. 313–321.

[7] S. G. Hart, “Nasa-task load index (nasa-tlx); 20 years later,” in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.

[8] K. Kurzhals, E. Cetinkaya, Y. Hu, W. Wang, and D. Weiskopf, “Close to the action: Eye-tracking evaluation of speaker-following subtitles,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 6559–6568.

[9] E. Gannon, J. He, X. Gao, and B. Chaparro, “Rsvp reading on a smart watch,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2016, pp. 1130–1134.

[10] B. Milliner, “Evaluating the lexical difficulty of teaching materials with nwlc,” *ELF 2*, p. 49, 2022.

[11] C. of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division, *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press, 2001.