情報処理学会論文誌 Vol.64 No.2 377–387 (Feb. 2023)

表情に基づく動画再生速度の自動調節による外国語学習支援システムの実装と理解度に与える影響の評価

西田 直人[1,†1,a)] 横山 海青[1,b)] 志築 文太郎[1,c)]

受付日 2022年5月10日, 採録日 2022年11月8日

概要：リスニングやシャドーイングなどの外国語を学習する方法があるなかで，多視聴という方法がある．多視聴とは，多くの動画を見ることにより外国語を学ぶ方法である．しかし，多くの外国語学習者は母語話者と同じように外国語を聞き取る，または読み取ることが難しく，動画の標準再生速度についていけない場面に直面する．そこで我々は，外国語学習者の理解度に応じて動画の再生速度を自動調節することにより，動画の視聴を支援するシステムを開発した．我々のシステムは，外国語学習者がコメディドラマ内の笑いどころにおいて笑っていると動画の内容を理解している，また，逆に笑っていないと動画の内容を理解していないと判断する．実験結果より，本システムは英語初学者から中級者の層（実験環境においては TOEIC Listening&Reading スコア 550 以上 700 未満の層）に対し，動画内容の理解を促進させる働きがあることが判明した．本システムを用いることにより，母語話者向けの動画を多視聴の教材とする際に，外国語学習者のリスニング能力を超えた教材も扱うことができるため，教材選定の選択肢を広げることができる．本論文において我々は，予備実験の詳細について述べるとともに，本実験の実験人数を追加し，統計的に有意な再解析を行った結果を述べる．再解析の結果，先行研究と同様に，本システムは初学者から中級者に対し有効であることを示した．

キーワード：コンピュータを使った語学学習，多視聴，表情認識

Implementation of Language Learning Assistance System by Video Playback Speed Adjustment Based on Facial Expressions and Evaluation of Its Effect on Learners' Comprehension

Naoto Nishida[1,†1,a)] Kaisei Yokoyama[1,b)] Buntarou Shizuki[1,c)]

Received: May 10, 2022, Accepted: November 8, 2022

Abstract: Among the various methods used for learning a second language (L2), such as listening and shadowing, Extensive Viewing involves watching plenty of videos. However, it is difficult for many L2 learners to smoothly and effortlessly comprehend the video content created for native speakers at the original speed of the video. Therefore, we developed a language learning assistance system that automatically adjusts the playback speed according to the learner's comprehension. If the learners laugh at the punchlines of comedy dramas, then our system determines that the learners have understood the content, and vice versa. Experimental results show that our system can aid learners with a relatively low L2 ability (from 550 to 695 in terms of the TOEIC Listening&Reading Score) in effectively understanding the video content. Therefore, our system can widen the learners' possible options with regard to the native speakers' videos as Extensive Viewing material. In this paper, we discuss our preliminary study in detail and the re-analysis of the main experiment in that we included additional participants for statistical significance. The results show that our system is desirable for participants with a relatively low English proficiency, thereby supporting the results of our prior experiment.

Keywords: computer-assisted language learning, extensive viewing, facial expression recognition

1. はじめに

外国語を学ぶ手法の1つとして，多視聴（*Extensive Viewing*）と呼ばれる方法がある．Ivone と Renandya は多視聴を，「容易に理解でき，娯楽性に富む対象言語の教材に長期間にわたって大量に触れること」と定義している [2]．多視聴には，ドラマやアニメといった教材が主に用いられる．

多視聴を行う利点としては，以下の事項があげられる：

- 連結した発音による音声変化に慣れることができる．
- 耳から聴いて把握できる語彙が増える．
- つなぎ言葉，スラング，および話し言葉に慣れることができる．
- 映像や字幕といった視覚情報があるため，音声のみのコンテンツに比べ内容把握がたやすい．
- 娯楽性が高く，学習意欲を維持しやすい．
- 自分の決めた時間に行うことができる．

以上の利点から，近年多視聴は外国語学習者の注目を集めている．

しかし，多くの外国語学習者は，母語話者と同じように対象言語を聴き取ることが難しい．そのため，多視聴を行う際，動画の通常再生速度において，話の流れについていけない場面に直面する外国語学習者は多い [3]．多視聴を行う際には，外国語学習者の外国語能力に合わせた教材を選定する必要がある [2] ため，このように母語話者向けの動画コンテンツを通常の再生速度のまま教材として扱うことは難しい．

その一方で，既存の GUI を用いて動画コンテンツの再生速度を変更する場合，視聴と同時に学習のためのメモをとることが難しく [4]，通常の速度で理解できないシーンのたびに速度変更の操作を行う必要がある．このように，長期的に行う学習方法である多視聴に不便さおよび負荷があると，外国語学習者は学習に対するモチベーションが下がり，学習の継続に支障を及ぼす可能性がある．

そこで我々は，外国語学習者の個人の理解度に合わせ，動画再生速度を自動で最適化する支援システムを開発した（**図 1**）[1]．本システムは，動画中の笑いどころにおける外国語学習者の表情に応じて動画の再生速度を自動調節する．本システムにより，GUI 操作による作業負荷が増えることなく，外国語学習者が動画内容を理解しやすくなる．そのため，多視聴の動画教材を選定する際に，外国語学習者にとって本来難易度が高すぎる教材にも選択肢を広げることが可能となる．

[1] 筑波大学
University of Tsukuba, Tsukuba, Ibaraki 305–8577, Japan
[†1] 現在，東京大学
Presently with The University of Tokyo
[a)] nawta@g.ecc.u-tokyo.ac.jp
[b)] kyokoyama@iplab.cs.tsukuba.ac.jp
[c)] shizuki@cs.tsukuba.ac.jp

本論文では，我々が以前に発表した論文 [1] において記述不足であった，本システム開発のための予備実験について述べた後に，実施した本実験（論文 [1] の Section 4）の実験人数を追加し再解析した結果を述べる．

本研究の貢献を以下に示す：

- 予備実験を通じて，動画中の笑いどころにおいて外国語学習者が笑った場合，外国語学習者は動画内容を理解していることを発見した．
- 外国語学習者の理解度に応じ，動画再生速度を自動調節するシステムを開発した [1]．
- 本システムは初学者から中級者に対し，動画内容の理解を促進させる効果があることを示した [1]．
- 実験人数を増やし再解析した結果，先行研究において行った実験 [1] と同様に，初学者から中級者に対して本システムが有効であることを示した．

2. 関連研究

本研究の立ち位置を示すために，我々はまず，コンピュータを用いた既存の多視聴への支援手法について調査した．次に，動画再生速度の調節について，具体的な支援手法について考察するため，我々は外国語学習者の理解度と再生速度の関係について調べた．我々はさらに，学習への作業負荷を軽減するために用いることができる手法も調べた．最後に，我々は多視聴の目的および学習形態の特徴から，多視聴に用いる機材における制約を調べた．

2.1 コンピュータを用いた既存の多視聴への支援手法

多視聴の文脈において，コンピュータが外国語学習を支援する既存手法としては，外国語学習者の語彙を増やすことに着眼した研究が多い．たとえば，Hu らは外国語学習者が対象言語の字幕上にマウスを置いた場合，その位置の対訳を表示させるシステムを開発した [5]．Fujii らは英語学習者の習熟度合いを判定し，適切な量の対訳を表示させるシステムを開発した [6]．

しかし，多視聴において，再生速度の調節という観点から外国語学習を支援する方法は少ない．さらに，既存のシステムは，ユーザが動画の速度を変更するために，マウスカーソルを用いた一連の動きを行う必要がある [5]．多視聴は外国語学習者が年単位で継続する必要がある学習方法であるため，学習にかかる作業負荷は可能な限り小さいことが望ましい．

そのため，我々は動画再生速度の調節という観点から，多視聴の学習を支援するシステムを開発した．本システムは既存手法よりも外国語学習者の多視聴にかかる学習コストが小さくなるような設計となっている．

2.2 外国語学習者の理解度と再生速度の関係

外国語学習の文脈において，Blau は発話速度が遅くなる

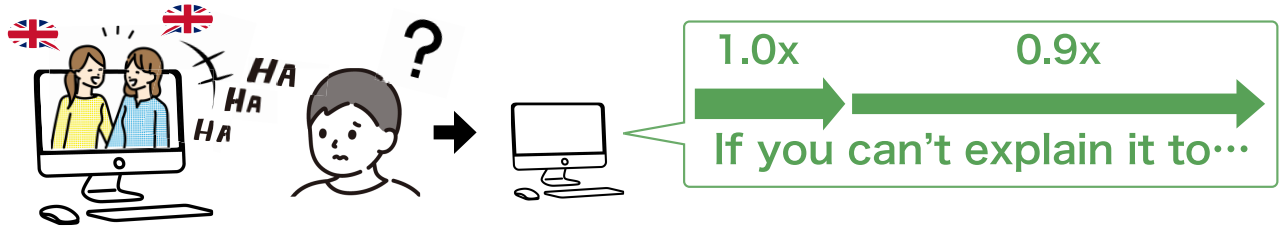


図 1 システム概要（文献 [1] より改変）
Fig. 1 System overview (modified from Ref. [1]).

ほど外国語学習者の聴解精度が向上したことを示した [7]．一方で，動画学習という文脈においては，大規模公開オンライン講座（MOOCs）において，動画速度が上がるほどオンライン動画講座の学習者が内容に対し，深い理解をするようになるということを Kao らは示した [8]．Kao らはさらに，再生速度を調節し動画時間を短縮することにより，オンライン動画講座の学習者が講座を修了する可能性が向上することも示した．

これらの結果から我々は次の 2 点の推測を行った．

- 外国語学習者が動画内容をよく理解できない箇所においては，外国語学習者の聴解の情報処理にかかる時間を増やすため，再生速度を遅くすると外国語学習者にとって有益である．
- 外国語学習者が動画内容を理解できる箇所においては，外国語学習者が集中力を保つことに余計な時間をかけないようにするため，再生速度を速くすると外国語学習者にとって有益である．

2.3 学習への作業負荷の軽減

多視聴の習慣をつけるためには，日々の学習にかかるコストを可能な限り減らす必要がある．Fujii らは無意識な視線の動きや姿勢を用いて外国語学習者の習熟度を予測した [6]．

そのため，我々も外国語学習者の無意識な行動を用い，多視聴の過程において必要となる作業負荷を軽減した．特に，我々は動画教材の笑いどころにおいて外国語学習者が笑っているか否かに着目した．

2.4 多視聴に用いる機材

多視聴の目的の 1 つとして，課外における外国語学習の時間を増やすという目的がある [9]．さらに，多視聴は日々の学習を年単位で続ける必要がある [9]．そのため，いつでもどこでも学習を行えるように，多視聴の準備にかかる手間は最小限にする必要がある．

そのため，我々は市販のノート PC の内蔵カメラを通じて取得できる特徴を支援に用いた．

3. 予備実験

表情および視線といった外国語学習者の理解度を測る適切な指標を調べるため，予備実験を実施した．

3.1 実験参加者

予備実験には情報科学専攻の学生 5 名（平均 23.60 歳，標準偏差 1.520 歳，全員男性）が参加した．2 名が研究室外，3 名が研究室内からの参加だった．5 名全員が英語を第二外国語として学んでいたため，英語を学習対象言語として選んだ．また，実験参加に際して常用する視力矯正器具を装着するよう指示した（裸眼 1 名，眼鏡 2 名，コンタクトレンズ 2 名）．実験中に顔の位置を大きく動かした者はいなかった．

3.2 実験に用いた機器

予備実験には，カメラを内蔵したノート PC（MacBook Pro 13 inch，macOS 11.6，2.8 GHz クアッドコア Intel Core i7，メモリ 16 GB）を用いた．実験参加者の顔が鮮明に見えるように，実験に用いた部屋の照明と内蔵カメラの位置関係について，逆光になるのを避けた．PC ディスプレイに対し参加者の顔の距離は PC 操作時の標準的な距離である 50 cm から 60 cm であった．PC ディスプレイと参加者の顔が正対するように PC を配置した．

3.3 実験手順

参加者はまず，内蔵カメラによって顔のフレーム画像を 30 fps で撮影されながら，5 本の母語話者向け英語動画コンテンツを英語音声および英語字幕の条件（多視聴において想定される学習環境）で視聴した．動画の長さは合計 9 分 12 秒であった．英語の動画コンテンツは，英語学習書 Keynote に掲載されている TED Talks，TED Education，および洋画からおよそ 60 秒から 90 秒ずつ切り取ったクリップをつなげて作成された [10]．

参加者の動画内容に対する理解度を大まかに測るため，および実験参加者の意識を動画に集中させるために，動画コンテンツ中に動画の内容に関する 4 択クイズをクリップとクリップの間に提示した．実用英語検定，TOEIC Listening&Reading，および TOEFL iBT のリスニング解答時間を参考にし，4 択クイズは 1 問あたり 15 秒提示された．実験参加者は紙に解答を書いた．その後，実験参加者は 30 fps で取得された動画コンテンツのフレーム画像に対

し，理解ができていた箇所および理解できていなかった箇所のラベリングを行った．教材動画のフレーム画像に理解できたとラベリングされた時間に対応する顔のフレーム画像を理解できたと定義し，ラベリングした．理解ができていた基準としては Webb [3] の基準に従い，瞬時に大まかに文意がとらえられることを理解の基準とした．実験の所要時間は 1 人あたり約 30 分であった．

3.4 実験結果

3.4.1 理解度の違いが外見的特徴に現れるか

参加者の顔のフレーム画像を解析した．フレーム画像は 5 名全員分で 62734 枚あり，動画の内容を理解できたとラベリングされたものが 52,385 枚，理解していないとラベリングされたものが 10,349 枚であった．

顔の特徴量を取得するために，OpenFace の視線追跡ユニットおよび Action Unit（AU）検出ユニットを使用した [11]．なお，本論文では，視線の特徴量および顔の筋肉部位の変化量である Action Unit 特徴量を顔の特徴と定義する．AU とは，人間の顔の筋肉部位に対応した 66 個の特徴点により表情をモデル化したものである [12]．AU は対応する筋肉部位が動くほど大きい数値を示す．視線追跡ユニットを用いて取得したデータは，G_x（視線のベクトルの x 成分）および G_y（視線のベクトルの y 成分）である．Action Unit 検出ユニットを用いて取得したデータは，AU1，AU2，AU4–6，AU7，AU9，AU10，AU12，AU14，AU15，AU17，AU20，AU23，AU25，AU26，および AU45 である．これらは OpenFace により取得できるすべての Action Unit である [11]．

解析には，正規性の確認のために Kolmogorov–Smirnov 検定を用いた．その結果，検定を行う従属変数（G_x，G_y，AU1，AU2，AU4–6，AU7，AU9，AU10，AU12，AU14，AU15，AU17，AU20，AU23，AU25，AU26，および AU45）について，すべて正規性が確認できなかった．よって，Wilcoxon の順位和検定を用い，動画内容の理解ができている際と理解ができていない際に従属変数に差があるかを調べた．その結果，AU45 以外において 2 群間に差が見られた．Hedges'g を用いて効果量を測定した結果，ほとんどの従属変数において無視可能であった．しかし，AU14 において 0.222 と小さな効果量が検出された [13]．AU14 は頬筋に対応する AU であり，笑顔に関連する．

3.4.2 アンケート結果

予備実験後の自由記述アンケートにおいて，内容を理解できない際に無表情を保ってしまうという意見を得られた．その理由としては，集中しようとして顔がこわばるため，理解ができないから話の笑いどころが分からないため，および，分からない箇所があまりに多いと表情を変えている余裕がないためと述べられていた．

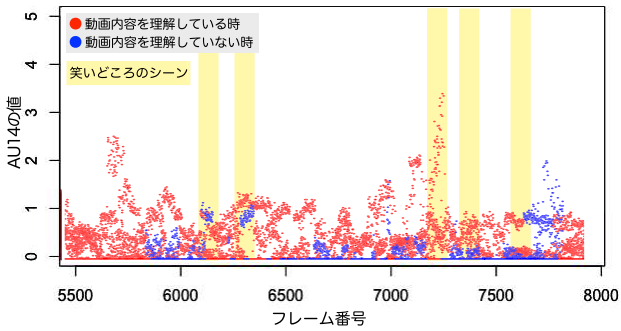


図 2 フレームに沿った AU14 の数値（文献 [1] より改変）
Fig. 2 Value of AU14 according to frames (modified from Ref. [1]).

3.4.3 録画映像の観察

解析結果およびアンケート結果をふまえ，実験参加者の笑顔に着目し，録画映像を観察し分かったこととして次の 2 点がある．

- 動画コンテンツ内の笑いどころに際し，参加者は内容を理解できる箇所において笑い，内容を理解できていない箇所において無表情を保つ．
- 参加者は動画コンテンツ内の笑いどころではない箇所において，無表情を保つ．

また，笑いどころ付近のフレーム（全実験参加者の 6,110–6,160，6,250–6,300，7,210–7,260，7,360–7,410，および 7,570–7,620 フレーム目）における AU14 の値のみ抽出し，理解ができている際の値と理解ができていない際の値について Hedges'g における効果量を測定した結果，効果量は 0.588 と高くなった．さらに，AU14 が理解度の有無によって示す値を笑いどころが出現する付近で可視化した（**図 2**）．この図から，理解できていない際にも AU14 の値が上がることがあるものの，理解できていると，理解できていないときよりも AU14 の値は大きくなる傾向にあることが分かった．

以上の結果から，次の 3 点が分かった：

- 実験参加者が笑顔であれば動画の内容を理解しており，笑顔でなければ動画の内容を理解していない傾向にある．
- 笑いどころにおける笑顔の有無に着目することにより，より精確に理解度の有無を判断できる．
- 理解していなくても笑いどころにおいて笑顔になることがあるものの，笑いどころにおいて理解している表情の群と理解していない表情の群に笑顔の指標に有意差と中程度の効果量がある．

4. 実装

予備実験の結果より，外国語学習者の理解度を測る指標として，外国語学習者が笑いどころで笑っているか否かを用いることにした．

つまり，本システムは外国語学習者が笑いどころにおい

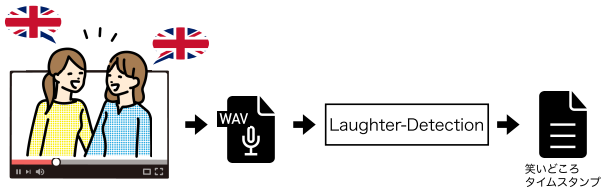


図 3 笑いどころの抽出手法
Fig. 3 Workflow of the extraction of punchline scenes.

て笑っていないと再生速度を落とし，逆に外国語学習者が笑いどころにおいて笑っていると再生速度を徐々に上昇させ，本来の動画再生速度に近づける．速度調節については，1.0 倍速，0.9 倍速，0.8 倍速，0.7 倍速，0.6 倍速の 5 段階で速度を段階的に変更することにした．理由としては以下になる：

- 後述する *Friends* の動画クリップにおいて音割れが起こらない下限の速度が 0.6 倍速であった．
- ネイティブスピーカが処理できることが担保されている 1.0 倍速の音声処理を素早く行うことが多視聴の目標である．
- 国内ワークショップのデモ発表 [14] において，0.1 倍速刻みが良いとフィードバックを受けた．

また，再生する動画については，コメディ動画は動画中に笑いどころが多いうえ，多視聴に用いられる動画はシチュエーションコメディ（以降，シットコム）が良いとされている [3]．よって，システムを用いて再生する動画はシットコムを想定した．YouTube に組み込まれているタイムストレッチを用いることにより，再生速度の変更からくるピッチの変更を減らし，違和感および不快感を生じさせにくくした．

笑いどころの識別には，シットコムの劇中の背景音に含まれる録音笑いを用いた．また，動画中で笑い声が聞こえる箇所のタイムスタンプを Laughter-Detection ライブラリ[*1]であらかじめ抽出し，筆者の手により手直しを施した（**図 3**）．

実装言語には Python 3.7.4 を用いた．内蔵カメラで取得したフレーム画像から参加者の笑い顔を検出するために，Perception for Autonomous Systems 内の分類モデルである MiniXeption を用いた [15]．MiniXeption で，Happy と判断された確率がすべての感情（Anger，Disgust，Fear，Happy，Sad，Surprise，および Neutral）の確率において最も高い場合，笑顔であると判断した．また，ブラウザ操作のために Selenium を用いた．

5. 実験

ユーザの表情を検出し動画の再生速度を調節するシステム（以降，システム）の有用性を評価する実験を行った．

[*1] https://github.com/jrgillick/laughter-detection（最終閲覧日：2022 年 9 月 3 日）

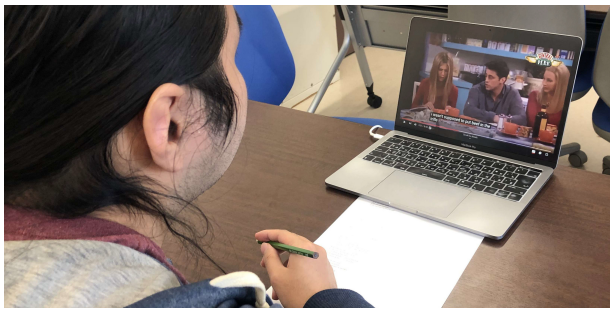


図 4 実験参加者が *Friends* を視聴している様子
Fig. 4 Participant watching *Friends*.

この実験は筑波大学の倫理審査委員会の承認（承認番号 2021R557）のもとで実施した実験である．

5.1 実験参加者

本実験には，情報科学専攻の学生 24 名（平均 22.70 歳，標準偏差 1.802 歳，20 名男性，4 名女性．P0，P1，...，P23 と以降表記）が参加した．10 名が著者が所属している研究室内からの参加者であり，14 名が研究室外からの参加者であった．参加に際し，報酬 880 円を全員に支払った．

本実験では，英語を学習対象言語として選定した．実験参加者には実験参加前に，カメラから表情を見えやすくするためにマスクを外す，および前髪を掻き分けるように指示した．また，実験参加に際して常用する視力矯正器具を装着するよう指示した（裸眼 11 名，メガネ 7 名，コンタクト 7 名）．実験参加者の対象としては，CEFR スコアで B1 以上英語に習熟している者を対象とした．理由としては，動画を用いた外国語学習は CEFR スコアで B1 以上（TOEIC の点数が 550 以上 [16]）の外国語学習者が対象と考えられることがあげられる．実験中に顔の位置を大きく動かした者はいなかった．

実験参加者の日常における英語の勉強時間について，週 2 時間未満と答えた者が 19 名，週 2 時間以上 5 時間未満と答えた者が 3 名，週 15 時間以上 20 時間未満と答えた者が 1 名，および週 20 時間以上 25 時間未満と答えた者が 1 名だった．実験参加者の普段の英語学習に対する取り組みについては，英語論文を読むこと，国際学会の英語発表を聴講すること，およびオンライン英会話があげられ，多視聴を日常的に行っている者はいなかった．

5.2 実験に用いた機材

実験に用いた機材，ならびに，実験参加者，光源およびディスプレイの位置関係は予備実験と同様だった（**図 4**）．ただし，PC を用いて参加者の表情を 30 fps で取得した．

5.3 実験構成

実験を行うにあたり，内容の構成をなるべく似せた動画を 2 本（以降，V1 および V2）作成した（**図 5** a，図 5 b）．

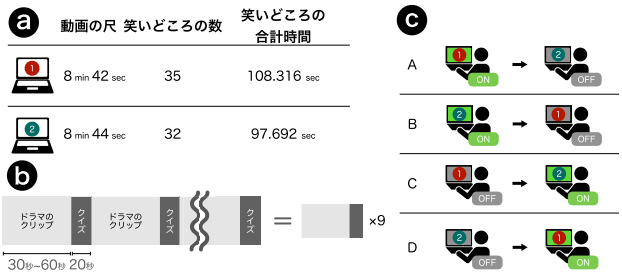


図 5 実験の統制．a）実験に使われた動画の統制．b）動画の構成．c）順序効果の統制（文献 [1] より改変）

Fig. 5 Experimental control. a) control of experimental videos. b) composition of a video. c) control of the order effect (modified from Ref. [1]).

V1 および V2 は 9 つのドラマのクリップ，およびその動画内容に関する 4 択クイズ 9 問（以降，Q1 および Q2）で構成されており，V1 および V2 内のドラマのクリップ，および V1 および V2 間のドラマのクリップに前後関係はなかった．V1 内で発言される語数は 578 語，および V2 内で発言される語数は 469 語だった．

さらに，V1 および V2 に登場する語彙のテスト（以降，VT1 および VT2）をそれぞれに対し 16 問作成した．VT1 および VT2 の問題の選定については，Corpus of Contemporary American English に掲載されている語彙の出現頻度を参考に，話の流れを理解し，笑いどころにおける発言を理解するために必要な語彙を選んだ．具体的には，Fujii らの研究 [6] を参考に，出現頻度が少ないものを優先し 16 問を出題した．

Q1 および Q2 については，実用英語検定，TOEIC Listening&Reading，および TOEFL iBT のリスニング解答時間を参考にし，1 問あたり 20 秒提示した．

V1，V2，Q1，Q2，VT1，および VT2 の作成理由は以下のとおりである：

- 複数の動画，複数のクイズ，および複数の語彙テストを用いて実験参加者のパフォーマンスを測定することにより，評価に対する動画の内容の影響を減らすことができる．
- Q1 および Q2 を出題することにより，動画内容の把握具合を調べることができるとともに，実験参加者の意識を動画に対し集中させることができる．
- VT1 および VT2 をそれぞれ V1 および V2 の視聴前に行うことにより，動画において語彙が分からず話の流れが分からなくなる可能性を減らすことができる．
- VT1 および VT2 をそれぞれ V1 および V2 の視聴後に行うことにより，システムを用いることによる語彙やイディオムの定着効果も評価することができる．

また，P13 以外の実験参加者 24 名を A，B，C，および D の 4 グループに 6 名ずつ分けた（図 5 c）．すべての実験参加者に対し，システムの有無のどちらの条件下においてもパフォーマンスを測定することにより，語学力の個人差がシステムの評価に及ぼす影響を減らすことができる．また，4 グループに分けることにより，V1 および V2 の再生の順番からくる順序効果，およびシステムの使用/不使用の順番からくる順序効果を相殺する．A から D までのグループ分けについて，実験参加者は P13 を除き全員 TOEIC Listening&Reading Test [*2] を受験したことがあったため，筆者は参加者の TOEIC Listening&Reading Test の点数（以降，TOEIC の点数）をもとにグループ分けを行った．グループ分けは，グループ間の平均および標準偏差がなるべく均等になるように行った．また，TOEIC の点数が 700 点以上の者が 12 名，700 点未満の者が 12 名いたため，それぞれから 3 名ずつ選び 1 グループを構成した．TOEIC の点数およびグループ分けの内訳を**表 1** に示す．

表 1 それぞれのグループの実験参加者の TOEIC の点数．アスタリスクは先行研究 [1] の中間調査の後，本研究において追加した実験参加者を示す

Table 1 TOEIC L & R Scores of each participant. The asterisks indicate that they newly participated in our experiment after our previous work [1].

	TOEIC の点数						平均（標準偏差）
A	565	635	660*	700	755	855	695.0（100.9）
B	550	630*	690	755	760	815	700.0（97.31）
C	590	620*	670	730	790	845	707.5（99.08）
D	550*	625*	670	680	785	880	698.3（117.5）

V1 および V2 の内容については，実装においてシットコムの使用を想定したため，シットコムの代表的作品である *Friends* の動画クリップを用いた[*3]．V1 および V2 を見せる際，システムを用いた条件（実験条件）とシステムを用いない条件（対照条件）のどちらで行っているかは実験参加者には知らせなかった（単盲検法で実験を行った）．

5.4 実験手順

統制により，A，B，C，および D はシステムを用いる順番および動画を見る順番が異なる（図 5 c）．よって簡単のため，A グループを手本に説明する．

実験参加者はまず，ひととおりの実験手順について説明を受けた後，VT1 に解答した．VT1 に答えた後，実験参加者は正解を教えられた．次に，Q1 について，実験参加者は Q1 のすべての問題文および選択肢を V1 視聴前に読み，問題文および選択肢に分からない語彙および表現がないかを確かめた．これにより，実験参加者が語彙および表現が分からないために Q1 に間違えたり，解答時間内に問題文や選択肢を読みきれなかったりする可能性を減らした．その後，実験参加者は V1 をシステムを使用した状態で視聴

[*2] https://www.iibc-global.org/toeic/test/lr.html（最終閲覧日：2022 年 9 月 3 日）
[*3] https://www.youtube.com/watch?v=nvzkHGNdtfk（最終閲覧日：2022 年 9 月 3 日）

した．視聴中，ドラマのクリップが流れている場面において，笑いどころにおける参加者の表情に従ってシステムが作動し V1 の再生速度を調節した．また，Q1 の問題文が提示された場面において，実験参加者は 20 秒の解答時間内に問題を解いた（**図 6**）．V1 を見終わったのち，実験参加者は VT1 に再度解答し，システムを用いた V1 視聴後の語彙の定着度を測った．その後，実験参加者は作業負荷を測定するために NASA-TLX [17] に回答した．また，実験参加者は 5 分以上の休憩の後，先述した実験条件下における手順と同様の手順を対照条件下で行った．この条件下においては，実験参加者は V1 の代わりに V2 を，システムを使用せずに視聴した．その際，実験参加者は Q1 および VT1 の代わりに，Q2 および VT2 に答えた．最後に，実験参加者は System Usability Scale（以降，SUS）[18] に回答した．システムを用いた時点がグループにより異なるため，SUS の回答の時点はすべてのグループがシステムを使用終了している時点に一致させた．

録画をしていると参加者が緊張し顔がこわばる可能性があるため，実験を通して録画をしていないことを参加者に伝えた．また，参加者が緊張しないように，参加者が V1 および V2 を見ている最中，実験実施者は部屋から出た．

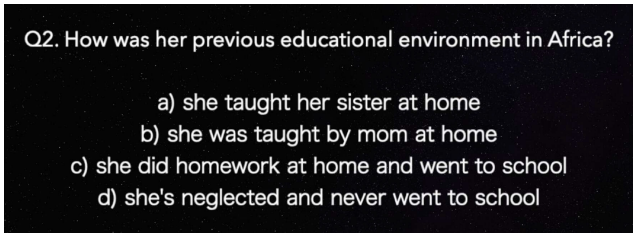


図 6 Q1 および Q2 の例
Fig. 6 Example of Q1 and Q2 in experimental videos.

表 2 VT1 および VT2 の解答時間，Q1 および Q2 の先読み時間，ならびにシステムを用いた際の視聴時間
Table 2 Time on answering VT1 and VT2, time on reading Q1 and Q2, and watching time with our system.

	VT の解答時間				Q の先読み時間	
	事前		事後			
	VT1	VT2	VT1	VT2	Q1	Q2
平均（秒）	114.5	130.8	45.95	48.61	203.8	190.0
標準偏差（秒）	32.64	51.92	11.84	17.19	124.9	108.5

そのほか，動画内容を想像することにより，Q1 および Q2 の点数に影響が出ないように，VT1 および VT2 に答える際，ならびに Q1 および Q2 の文を先読みする際，V1 お

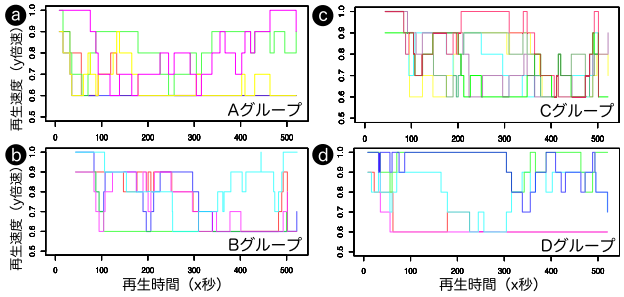


図 7 グループごとの再生速度の変化．a）A グループ．b）B グループ．c）C グループ．d）D グループ
Fig. 7 Speed logs by group. a) Group A. b) Group B. c) Group C. d) Group D.

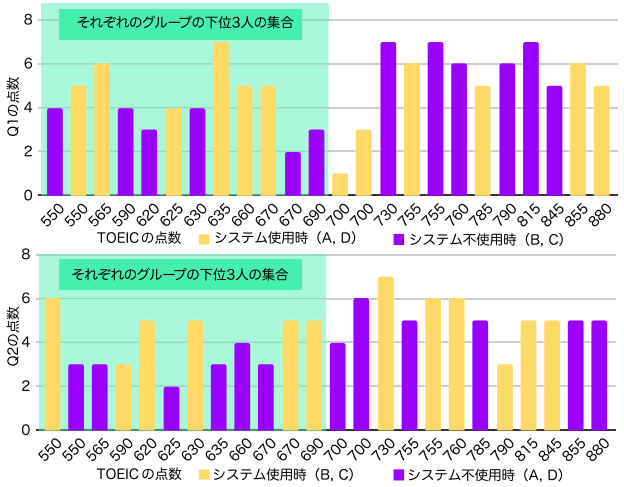


図 8 Q1 および Q2 の正答数と参加者の TOEIC の点数の関係．黄色のバーはシステム使用者の Q1 および Q2 の点数，ならびに紫色のバーがシステム不使用者の点数を示す．緑色部分は TOEIC の点数が 700 未満の者であり，それぞれのグループにおける下位 3 人の集合である
Fig. 8 Relationship between Q1 scores, Q2 scores and TOEIC scores of each participant. The yellow bars represent the scores of users with the system, and the purple bars represent the scores of users without the system. The green area represents the set of the bottom 3 TOEIC scores from each group as well as the set of TOEIC scores below 700.

表 3 VT1 および VT2 の点数
Table 3 Scores of VT1 and VT2 by group.

	平均値（標準偏差）				中央値			
	V1		V2		V1		V2	
	視聴前	視聴後	視聴前	視聴後	視聴前	視聴後	視聴前	視聴後
A	12.83（1.169）	16.00（0.000）	9.500（2.345）	15.83（0.408）	12.50	16.00	9.500	16.00
B	12.80（1.303）	16.00（0.000）	10.80（2.490）	15.80（0.447）	14.00	16.00	10.00	16.00
C	13.17（0.752）	15.67（0.516）	9.5（2.739）	16.00（0.000）	13.00	16.00	11.00	16.00
D	14.4（1.673）	16.00（0.000）	9.200（2.168）	16.00（0.000）	14.00	16.00	9.000	16.00

よび V2 の内容を極力思い浮かべないように参加者に伝えた．実験の合計時間は 1 人あたり平均 64 分だった．V1 および V2 の解答時間，Q1 および Q2 を事前に見た時間（以降，設問の先読み時間）を**表 2** に示す．また，システム使用時の動画視聴時間の平均は 624.1 秒であり，標準偏差は 46.31 秒だった．

5.5 結果

5.5.1 VT1 および VT2 の点数

V1 および V2 の視聴前後に，実験参加者は VT1 および VT2 に答えた．各グループの点数を**表 3** に示す．システムを用いた際も用いていない際も，V1 および V2 視聴後の VT1 および VT2 の点数は全員ほぼ満点であり，システムを用いることによる語彙の定着度への差は見られなかった．

5.5.2 システムを用いて動画を見た際の速度の変更記録

システムを用いて V1 および V2 を見た際，再生速度の変更記録を取得した．グループごとの再生速度の変化を**図 7** に示す．再生速度の変更記録に特徴は見られなかったが，システムが外国語学習者の表情に合わせて再生速度を適宜調節したことが分かる．

5.5.3 Q1 および Q2 の点数

Q1 および Q2 の正答数と参加者の TOEIC の点数の関係を**図 8** および**表 4** に示す．図 8 および表 4 は，システム不使用時の点数（紫）はおおよそ TOEIC の点数と正比例にあるが，システム使用時の点数（黄色）は TOEIC の点数と比例関係にないことを示す．また，**表 5** はシステムの有無が Q1 および Q2 に与える影響を基本統計量および 2 群間の差により示したものである．検定手法および効果量としては，Shapiro-Wilk 検定，Kolmogorov-Smirnov 検定，Student の t 検定，Wilcoxon の順位和検定，および Hedges'g を用いた．結果から，TOEIC の点数が下位 12 名の Q1 および Q2 の点数については，有意に約 2 点の差が現れることが示された．

5.5.4 作業負荷およびユーザビリティの評価

NASA-TLX の結果を**図 9** に示す．タイムプレッシャについては，システムを用いた際の方が有意に負荷が低く，そのほかの項目については有意差はなかった．

SUS については，平均が 73.6，標準偏差が 13.92，および中央値が 75 であった．Aaron らによると，SUS の点数が示すユーザビリティの評価は Good である [19]．

6. 議論および今後の展望

実験結果から，Q1 および Q2 の結果より，TOEIC の点数がおよそ 700 未満（CEFR スコア B1 下位以下）の学習者は，本システムを用いることにより動画内容を理解しや

表 4 システムを使用したグループおよび不使用のグループにおける，TOEIC の点数と Q1 および Q2 の点数について相関係数および無相関検定の結果．Spearman の順位相関係数および Spearman の無相関検定を用いた

Table 4 Result of the correlation coefficient and the test of no correlation between the group which used our system and the control group. We used Spearman's rank correlation coefficient and its test of no correlation.

システム	視聴した動画/グループ	相関係数	無相関検定の有意水準
有	V1/(A, D)	−0.026	0.937
	V2/(B, C)	0.004	0.991
無	V1/(B, C)	0.619	0.032
	V2/(A, D)	0.813	0.002

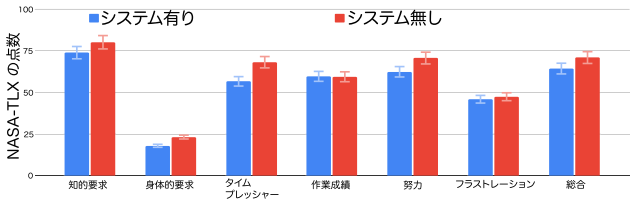


図 9 NASA-TLX の結果．エラーバーは両方向の 95%信頼区間を示す

Fig. 9 Result of NASA-TLX. The error bars represent 95% confidential interval in both directions.

表 5 システムを使用したグループと不使用のグループについて，Q1 および Q2 の点数に関する基本統計量ならびに 2 群の差の有意確率および効果量．TOEIC の点数順に上位，下位および全体で区分けした

Table 5 Statistics and difference between the groups with the system and the groups without the system.

TOEIC の点数による区分け		中央値		平均値（標準偏差）		有意確率	効果量
		システム有	システム無	システム有	システム無		
下位 12 名	Q1	5.000	3.500	5.333（1.032）	3.333（0.816）	0.004	1.983
	Q2	5.000	3.000	4.833（0.983）	3.000（0.632）	0.011	2.047
上位 12 名	Q1	5.000	6.500	4.333（1.966）	6.333（0.816）	0.032	1.226
	Q2	5.500	5.000	5.333（1.366）	5.000（0.632）	0.389	0.289
全体	Q1	5.000	4.500	4.833（1.586）	4.833（1.749）	0.930	0.000
	Q2	5.000	4.000	5.083（1.165）	4.000（1.206）	0.036	0.882

すくなったと分かった．しかし，それ以上の TOEIC の点数を取得している学習者にとってはシステムを使った効果があまり見られなかった．よって，TOEIC の点数が 700 以上の学習者の理解が不十分になる理由としては，速度以外のほかの要因があると考えられる．実際に，事後アンケートにおいて得た TOEIC の点数が 700 以上の学習者の自由記述において，突飛なストーリーについていけない，および動画内で用いられている英語の文法で分からないものがあったと指摘されていた．前者については，実験参加者が米国ドラマおよび米国文化に慣れていない可能性がある．後者については，米国の日常会話における文法知識を実験参加者が持っていなかった可能性がある．このように，より熟練の外国語学習者に対しては，再生速度以外の問題点に着目し，支援を行うことが必要である．

制限および今後の展望としては 6 点ある．

まず，本実験においては，実験の規模を限定するため，ブラウザの操作などの既存の動画速度変更手法との比較，および静的に動画全体の再生速度を変化させた条件との比較を行っていない．これらの手法および条件との比較を行えば，実験参加者の動画内容理解の促進について，本システムの評価をより相対的に行うことができる．ゆえに，今後上記手法および条件との比較検討を行う必要がある．

表情をシステムの入力とすることについて，多視聴は日常的に行う必要がある学習法であり，学習準備にかかる作業を極力減らすことが望ましいため，本システムでは PC 以外の機器を用いず，PC 内蔵カメラから取得可能な特徴を入力として用いた．しかし，実験参加者は映像および背景音に含まれる録音笑いにつられて笑う可能性もあるため，表情は理解度を示す完全な指標とはならない．実際，予備実験では理解できていない際にも AU14 の値は上昇しうること，つまり理解できていない際にも笑う可能性があることが示されている（図 2）．また，本実験の事後アンケートにおいて，理解していなくても映像につられて笑った場面があると申告した者が 1 名いた．つられ笑いへの対策として今後，PC 内蔵カメラ以外から取得可能な特徴を調べ，理解度を示すのにより適した指標を調べる必要がある．2 点候補をあげれば，瞳孔径および脳波は e ラーニングにおいて集中度または理解度を測定する実例が存在するため，指標となりうる [20], [21], [22].

ほかに，本システムでは再生速度の変更範囲を 0.6 倍速から 1.0 倍速に 0.1 倍速刻みに設定していた．1.0 倍速を上限とした理由は次になる：

- 笑顔を入力とした本システムは誤判断が起こる可能性があり，必要以上に速くなる可能性がある．
- 聞き取りの目標の速さ（1.0 倍速）以上にして試聴時間の短縮の評価を行う前に，理解度の促進を行えたかの評価を行うべきである．
- コメディドラマが題材であるため，1.0 倍速より速い場合にドラマを面白く感じるかが不確定である．

しかし 1.0 倍速以上へ速度変化の上限を変更した場合，1.0 倍速で動画内容を十分理解できる学習者は動画時間を短縮できるため，本システムは本実験における TOEIC の点数 700 以上の者に対しても有用なシステムとなる可能性がある．今後は動画速度を 1.0 倍速よりも速くした場合の評価を行う必要がある．

実験に用いた動画は 2 本であり，どちらもシットコムであった．数と種類に制限があるため，今後，より多種の動画を用いたケースおよび学習者の理解度を測定できる特徴量を調べ，システムの汎用性および拡張性を調べる．

さらに，参加者の TOEIC の点数と Q1 および Q2 の点数の比較については，検定内において，参加者という変数と，システムを使ったか否かという変数が異なるため厳密には比較ができていない．しかし，今後さらにサンプル数を増やして検証すれば，参加者の個人差のバイアスを打ち消すことは可能である．

また，速度変更の実装について，本システムにおいては過去の再生部分に対する反応に基づいて，未来の再生部分の速度を変化させている．これは次の理由に基づく：

- 多視聴は日常的に継続する必要がある勉強方法のため，日常的にストレスなく学習を続けるためには，特定の動画の場面を繰返し見るよりも多くの新鮮な動画に触れる方が良い．
- 通常速度で理解できない場面の周辺の場面も，通常速度で理解できない場面の可能性が高い．
- 表情をシステムへの入力として用いた実装は誤判断が生じうることを示す予備実験の結果があり，誤判断が生じても学習者にストレスがかかりにくい実装にするべきである．
- 対象となる動画の笑いどころのタイムスタンプを Laughter–Detection ライブラリを用いて抽出すればシステムが適用可能であるため，前処理が簡単である．

しかし，ほかの実装方法として，実験参加者が動画内容を理解できていないとシステムが判断した場面を巻き戻す実装，または再生速度を遅くせず，単語間の無声部分を長くし Words Per Minutes を落とす実装が考えられる．実際に，先行研究例として，文章を反復して聞くことおよび単語間の無声区間を長くとることにより聴解の理解度が深まると Carvantes らおよび Blau は示した [7], [23]．さらに，Blau は単語間の無声区間を長くする方が音声全体を引き伸ばすよりも聴解の理解はしやすくなると示している [7]．今後，これらの実装方法と本システムの実装について，学習者が感じる作業負荷，および学習者の動画内容の理解促進に与える影響を調べ，システム改良をする余地がある．

7. 結論

本研究では，多視聴を行う際に学習者が動画コンテンツ

の再生速度についていけない場合に備え，学習者の理解度に合わせて再生速度を自動調節するシステムを実装し，評価した．学習者の理解度については，予備実験の結果およびデモ発表におけるフィードバックに基づき，笑いどころにおける外国語学習者の表情から判断することにした．

評価実験において，本システムはすべての実験参加者の作業負荷を高めず，英語初学者から中級者の層（実験環境においては TOEIC Listening&Reading スコア 550 以上 700 未満の層）の参加者において，内容に対する理解を有意に向上させることが分かった．

よって，本システムを用いれば，作業負荷を高めることなく，外国語学習者は自身の語学能力を超えた教材を理解できるようになると示された．つまり，本システムは多視聴のための動画教材の選択肢を広げることができると示された．

参考文献

[1] Nishida, N., Nozaki, H. and Shizuki, B.: Laugh at Your Own Pace: Basic Performance Evaluation of Language Learning Assistance by Adjustment of Video Playback Speeds Based on Laughter Detection, *Proc. 9th ACM Conference on Learning @ Scale*, pp.368–373, Association for Computing Machinery (online), DOI: 10.1145/3491140.3528299 (2022).

[2] Ivone, F. and Renandya, W.: Extensive Listening and Viewing in ELT, *Teflin Journal*, Vol.30, No.2, pp.237–256 (online), DOI: 10.15639/teflinjournal.v30i2/237-256 (2019).

[3] Webb, S.: Extensive Viewing: Language Learning through Watching Television, *Language Learning Beyond the Classroom*, pp.159–168 (online), DOI: 10.4324/9781315883472-24 (2015).

[4] Song, S., Hong, J., Oakley, I., Cho, J.D. and Bianchi, A.: Automatically Adjusting the Speed of E-Learning Videos, *Proc. 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp.1451–1456, Association for Computing Machinery (online), DOI: 10.1145/2702613.2732711 (2015).

[5] Hu, S.H. and Willett, W.J.: Kalgan: Video Player for Casual Language Learning, *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pp.1–6, Association for Computing Machinery (online), DOI: 10.1145/3170427.3188498 (2018).

[6] Fujii, K. and Rekimoto, J.: SubMe: An Interactive Subtitle System with English Skill Estimation Using Eye Tracking, *Proc. 10th Augmented Human International Conference 2019*, pp.1–9, Association for Computing Machinery (online), DOI: 10.1145/3311823.3311865 (2019).

[7] Blau, E.K.: The Effect of Syntax, Speed, and Pauses on Listening Comprehension, *TESOL Quarterly*, Vol.24, No.4, pp.746–753 (online), DOI: 10.2307/3587129 (1990).

[8] Kao, C., Liu, Y. and Hsu, A.: Speeda: Adaptive Speed-up for Lecture Videos, *Proc. Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology*, pp.97–98, Association for Computing Machinery (online), DOI: 10.1145/2658779.2658794 (2014).

[9] Renandya, W. and Jacobs, G.: Extensive Reading and Listening in the L2 Classroom, *English Language Teaching Today: Linking Theory and Practice*, chapter 8, pp.97–110, Springer (2016).

[10] Bohlke, D., Dummett, P., Lansford, L. and Stephenson, H.: *Keynote, American English*, CENGAGE Learning (2016).

[11] Baltrusaitis, T., Zadeh, A., Lim, Y.C. and Morency, L.-P.: OpenFace 2.0: Facial Behavior Analysis Toolkit, *13th IEEE International Conference on Automatic Face & Gesture Recognition*, pp.59–66 (online), DOI: 10.1109/FG.2018.00019 (2018).

[12] Ekman, P.: Facial Expression and Emotion, *American psychologist*, Vol.48, No.4, pp.384–392 (online), DOI: 10.1037/0003-066X.48.4.384 (1993).

[13] Cohen, J.: *Statistical Power Analysis for the Behavioral Sciences*, Routledge (2013).

[14] 西田直人，野崎陽奈子，志築文太郎：表情に基づく再生速度自動調節機能を備えた外国語学習支援の提案，第 29 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 (2021).

[15] Arriaga, O., Valdenegro-Toro, M., Muthuraja, M., Devaramani, S. and Kirchner, F.: Perception for Autonomous Systems (2020).

[16] Richard, T. and Caroline, W.: Linking English-Language Test Scores Onto the Common European Framework of Reference: An Application of Standard-Setting Methodology, *ETS Research Report Series*, Vol.1, pp.1–75 (online), DOI: 10.1002/j.2333-8504.2008.tb02120.x (2008).

[17] Hart, S. and Staveland, L.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, Hancock, P. and Meshkati, N. (Eds.), *Human Mental Workload*, Vol.52, pp.139–183 (online), DOI: 10.1016/S0166-4115(08)62386-9 (1988).

[18] Brooke, J.: SUS: A “Quick and Dirty” Usability Scale, *Usability Evaluation In Industry*, pp.189–194, Taylor and Francis (1996).

[19] Bangor, A., Kortum, P. and Miller, J.: Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale, *Journal of Usability Studies*, Vol.4, No.3, pp.114–123 (online), DOI: 10.5555/2835587.2835589 (2009).

[20] Wang, H., Li, Y., Hu, X.S., Yang, Y., Meng, Z. and min Kevin Chang, K.: Using EEG to Improve Massive Open Online Courses Feedback Interaction, *AIED Workshops* (2013).

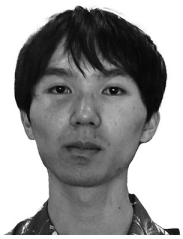
[21] Schneegass, C., Kosch, T., Baumann, A., Rusu, M., Hassib, M. and Hussmann, H.: BrainCoDe: Electroencephalography-Based Comprehension Detection during Reading and Listening, *Proc. 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, pp.1–13, Association for Computing Machinery (online), DOI: 10.1145/3313831.3376707 (2020).

[22] 野間慶子，小川賀代：学習支援システムに向けた計算時における瞳孔反応のパターン化，技術報告 1 (2011).

[23] Cervantes, R. and Gainer, G.: The Effects of Syntactic Simplification and Repetition on Listening Comprehension, *TESOL Quarterly*, Vol.26, pp.767–770 (online), DOI: 10.2307/3586886 (2012).

西田 直人 （学生会員）

1998 年生．2022 年筑波大学情報学群情報メディア創成学類卒業．現在，東京大学大学院学際情報学府博士前期課程在学中．ヒューマンインタフェースに関する研究に興味を持つ．ACM，IEEE 各学生会員．



横山 海青 （学生会員）

1998 年生．2021 年筑波大学情報学群情報科学類卒業．現在，同大学院システム情報工学研究群情報理工学位プログラム博士前期課程在学中．ヒューマンインタフェースに関する研究に興味を持つ．

志築 文太郎 （正会員）

1971 年生．1994 年東京工業大学理学部情報学科卒業．2000 年同大学大学院情報理工学研究科数理・計算科学専攻博士課程単位取得退学．博士（理学）．現在，筑波大学システム情報系教授．ヒューマンインタフェースに関する研究に興味を持つ．日本ソフトウェア科学会，ACM，IEEE Computer Society，ヒューマンインタフェース学会各会員．