

Smart Replay:

eラーニング動画における視覚的・時間的アクセシビリティの向上

SECHAYK Yotam * SHAMIR Ariel † 五十嵐 健夫 *

概要. eラーニングは、教材への幅広いアクセスを可能にすることを目的としている。しかし、動画コンテンツを多用することは、アクセシビリティに大きな課題をもたらす。多様な参加者を対象とした予備的調査に基づき、eラーニングのビデオコンテンツに存在するアクセシビリティの障壁を明らかにする。これらには、ユーザーの理解速度と動画速度の不一致、どこに注意を向けてよいかわからない視覚的複雑さ、ナビゲーションのしにくさなどが含まれる。また、我々の調査結果は、アクセシビリティの問題が、障害のある利用者となし利用者の両方に影響を及ぼすことを示している。さらに、既存のアクセシビリティツールには限界があり、さらなる対応が必要であることを示している。そこで我々は、eラーニングアクセシビリティツールである「Smart Replay」を提案する。私たちのツールは、学習ビデオのビジュアルとコンテンツに基づいた分析を行い、アクセシブルな再生オプションを生成する。視覚的分野と時間的分野の両方を強化したビデオセクションの復習を可能にする。

1 はじめに

eラーニングはインターネットを利用した教育で、世界中の学習者に利便性を提供している。近年、eラーニング、特にビデオコンテンツによる教育が人気を博している。しかし、eラーニングは、障がいのある利用者にとっては、学習能力に影響を与えるという課題がある。静的メディアとは異なり、動画は情報を時間的に分散させ [31]、検索を困難にし [17]、双方向性を十分に活用できていない [14][3]。これは、学習可能性を妨げ、アクセス障壁を提示し、多様な学習者にとっての有効性を低下させる。

アクセシビリティ（バリアフリー）とは、どのような能力状態にある人々にとっても、障壁を減らし、困難を軽減することである。アクセシビリティ [29] および eラーニング [13] 研究への関心が高まる中、ビデオベースの eラーニングの領域は、依然として重大なアクセス障壁を抱えたままである。本研究の目的は、このようなアクセシビリティの課題を解明し、改善策として新しいアクセシビリティツールを紹介し、この領域における今後の試みを促すことである。まず、 $N = 53$ の参加者を対象としたウェブ調査から始める。年齢、興味、能力状態（感覚、身体、認知など）の多様性を包含し、一般的なアクセス障壁とアクセシビリティ・ツールの利用パターンを精査する。この包括的な分析により、あらゆるタイプの学習者が遭遇する一般的な課題について、貴重な洞察を得ることができる。

その結果に基づき、動画学習のためのアクセシビ

リティツールである **Smart Replay** (SR) を提案する。SR は、ビデオ再生のアクセシビリティを向上させる要素を統合するために、視覚的領域と文脈的領域の両方でビデオ分析を採用している。このツールは、(1) どこに注意を向ければよいかわからないという問題、(2) 複雑なコンテンツナビゲーション、(3) 動作再生速度と利用者の理解速度の不一致をターゲットとしている。ウェブ・インターフェースを介して、ユーザーはアクセシビリティを向上させた eラーニング・ビデオを視聴、ナビゲート、再生することができる。

2 関連研究

eラーニングは、教育のためにインターネットを使用し、世界中の学習者にとって大きな利便性を提供している [30]。特に Covid-19 パンデミック以降、eラーニングの研究は成長している。以前の研究では、異なるプラットフォームの利用とその影響についての調査が行われている [35][23][43]、また、eラーニングの効果を評価する方法やアプローチについても研究が行われている [16][2][25]。さらに、異なるデザインのパラダイムに関する研究も行われており [28][45][24][12][1][10]、eラーニングアクセシビリティについての研究も進行中である [7][47]。

eラーニングアクセシビリティの研究では、ユニバーサルデザインの活用や特定の環境におけるアクセシビリティの分析が行われている [39][40][42]。また、新たなアクセシブルガイド生成技術に関する研究も行われている [22]。既存の研究とは異なり、私たちの研究では、さまざまな能力を持つ人々と共有されるアクセシビリティニーズを特定し、取り組ん

Copyright is held by the author(s).

* 東京大学

† Reichman University

でいる。

アクセシビリティ研究は進展しているが [29], まだ多くの課題が残されている。過去の研究では, ビデオのアクセシビリティ評価や, よりアクセス可能なビデオを生成するツールの提供に焦点が当てられている [27][26][15][32][48]。静的な視覚フィルターや視線ベースのズームングといった手法も探求されている [41][6]。また, 最近の機械学習の進歩により, ビデオの説明やコンテンツの簡素化が改善されている [49][44]。

一般のビデオと異なり, eラーニングビデオは固有の課題を提供している [18]。eラーニングの中核的なニーズは, 学習可能性 (理解) の促進である。これまでの研究では, ゲーミフィケーションの利点やコンテンツの説明 [50][21][51][19][44], スライドのアクセシビリティ分析 [38], 黒板のコンテンツ理解 [11], スクリーンリーダーに対応したスライド [36][37], そしてeラーニングビデオのナビゲーションに関する研究が行われている [31]。

しかし, 時間的な柔軟性と認知負荷の考慮は未だに解決されていない課題である。これまでの研究は, eラーニングビデオをスクリーンリーダーなどの特定のアクセシビリティツールに統合することに焦点を当ててきたが, 広範なアクセシビリティのニーズを見逃している可能性がある。私たちの研究では, 既存の障壁と核心的な問題を特定し, それらに取り組むための基盤を築くことを目指している。

3 予備的調査

現在のアクセス障壁を理解するために, 私たちは予備的調査を行った。障害のある利用者もない利用者も募集した。テルアビブ大学のバリアフリー支援室による参加者募集と, ソーシャルネットワークグループへの投稿を通じて募集を行った。10個の質問からなるウェブアンケートを実施し, すべての質問は任意である。そのため, 回答数が少ない質問もある (< 53)。主な対象は, (1)eラーニングの利用状況, (2) 学習ビデオのユーザー・エクスペリエンス, (3) 既存のツールと未対応の問題である。私たちの知る限り, これは混在する参加者を対象にアクセスの障壁を調査した最初の研究である。また, アクセシビリティ・ツールが, 障害の有無にかかわらず, 個人にどのようなプラスの影響を与えることができるかを, より詳細に明らかにした。

参加者の年齢はさまざままで, 26~35歳が27人(51%), 18~25歳が15人(28%), 46~60歳が7人(13%), 36~45歳と61歳以上がそれぞれ2人(4%)だった。性別や学歴に関する情報は収集していない。能力の多様性(すなわち, 障害の状態)も, 本研究の重要なポイントである(表1)。先行研究では, 特定のユーザー・グループ [39][8][4] に焦点を当てていたが, 我々は, アクセシビリティ・

ニーズの共通性を理解するために, 多様性を重視している。

表 1. 障害状態の概要, 複数選択可。合計: いくつ挙げたか。排他的: 排他的に言及した数。52の回答がある選択問題。

状態	合計	排他的
正常な視力への修正	12	6
視覚障害または低視力	11	8
色覚異常	1	1
聴覚障害または難聴	11	8
認知障害	3	2
ディスレクシア	2	0
ADHD (注意欠陥多動性障害)	11	5
身体的障害	7	5
なし	7	7

Part 1: eラーニングの利用状況 (Q1-Q6)。参加者はオンライン学習リソースを利用する傾向が高い(図1)。「読書」は49名, 「ビデオ (オンデマンド)」は45名, 「オンライン授業 (ライブ)」は37名が利用している。上記のすべてを利用しているのは合計30人である。さらに, 43人がインストラクションビデオ (ハウツー, ガイドなど) を, 34人が講義録を, 28人がビデオエッセイを視聴し, 17人がすべてを利用している。

よく視聴されるコンテンツの一般的なタイプには, 専門的なトピック (38), 料理 (24), 数学 (18), プログラミング (14) などが含まれる。人気のあるプラットフォームには, YouTube¹ (47), Khan Academy² (11), Coursera³ (11) などがある。

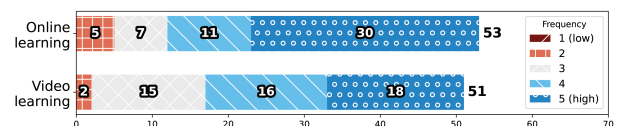


図 1. オンライン学習コンテンツ, 特にビデオを利用する可能性。

Part 2: 学習ビデオのユーザー・エクスペリエンス (Q7-Q8)。参加者はスマートフォン (9人) よりもPC (42人) を非常に好む。また, 能力の有無に関わらず, 多くの困難を経験している。我々は, これらを時間的, 視覚的, 音声的な困難と区別している(表2)。非障害者 (7名) 全員が, 学習ビデオを視聴する際に多くの困難に直面すると述べている。

¹ <https://www.youtube.com/>

² <https://www.khanacademy.org/>

³ <https://www.coursera.org/>

Smart Replay

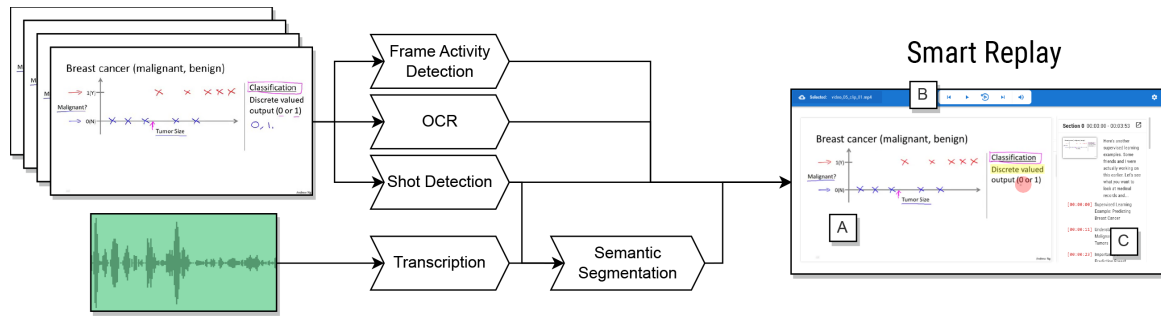


図 2. Smart Replay のシステム概要図。映像の分析、特徴の抽出、アクセス可能な再生のオンライン計算。スマート・リプレイ：(A) ビデオ表示，(B) 再生コントロール，(C) チャプター選択と概要。

表 2. 困難. (T.) 時間的, (V.) 視覚的, (A.) 音声的.
(* = 自由記述).

困難	合計
T. Speech can be too fast	26
Reading on-screen text in time	14
The pace is too fast (or slow)	10
V. Pause frequently to see better	21
On-screen text is too small	17
Noticing small on-screen details	11
Bright colors are overwhelming	9
Distinguishing between colors	7
Unsure about attention placement	7
Unnoticeable non-central elements	5
Unreliable audio descriptions*	2
Visual overload*	1
A. Hard to hear the video clearly	24
Unreliable or missing captions	20

Part 3: 既存のツールと未対応の問題である (Q9-Q10). 参加者が使用しているツールはごく少数である。36 人がビデオ速度の変更を、27 人がキャプションを、20 人がノイズ低減ヘッドフォンを、8 人がスクリーン・リーダー・ソフトウェアを使用している。その他、画面拡大鏡、カラーフィルター、視線追跡カーソルなどが挙げられている。さらに、多くのアクセシビリティ・ニーズが、参加者から依然として提起されている。音声やオーディオの強化、カスタマイズによる信頼性の高い書き起こし、スクリーン・リーダーのサポート不足、より簡単なナビゲーション、双方向性などである。現在、研究は進んでいるものの、利用可能なツールは限られており、存在するツールに対する認識も不足している。

4 調査結果

学習者の能力に関係なく、アクセスの障壁は存在する。現在のツールの状況には多くの不満が残る。また、ユーザによるカスタマイズの有用性は [20][9] で議論されているが、アクセシビリティに関するカスタマイズの可能性については、さらなる調査が必要である。さらに、既存の研究にもかかわらず、多くのアクセシビリティの側面が未解明のままである。

我々の発見に基づき、ユーザーが直面する 3 つの重要な問題を区別し、これらに対処するための新しいアクセシビリティ・ツールを提案する。

- P1 (ついていくのが難しい) ビデオの時間的な再生速度がユーザーの理解速度と一致していない。
- P2 (見づらい) 多すぎる視覚的要素が注意を誘導する可能性がある。
- P3 (ナビゲーションが難しい) ビデオ内での検索は複雑で、多くのユーザーアクションに依存している。

5 Smart Replay

このような問題を解決するために、我々は e ラーニングビデオのアクセシビリティツールである *Smart Replay* (SR) を提案する。システム概要図を図に示す (図 2)。スライド、ナレーション、ライブドローイングを組み込んだ一般的なビデオのスタイル [46][5] に焦点を当てている (図 3)。

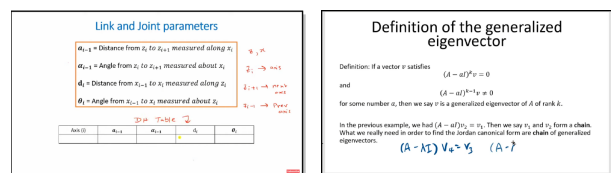


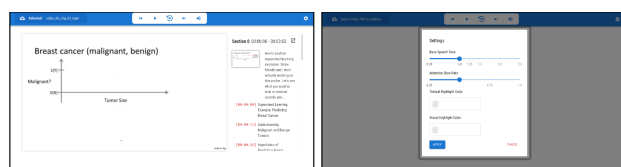
図 3. 対応コンテンツ例のフレーム。

映像解析により、複雑な映像部分でも利用しやすい再生機能を合成する。音声と視覚情報（図2）を利用し、既存の方法を用いて、光学的文字認識（OCR）、文字起こし、ショットセグメント情報を抽出する。

さらに、セマンティックセグメンテーション、OCR-音声相互参照、フレーム活性検出を導入する。後者は、eラーニングビデオのための新しいアプローチである。抽出された情報は再生システムに供給され、学習者は文脈上のセグメントを簡単にナビゲートしたり、アクセシビリティを強化して再生したり、学習体験をカスタマイズしたりするツールを得ることができる。

5.1 システムの特徴

ここでは、各機能の詳細とその問題点（P1～P3）について説明する。すべての機能にはカスタマイズが含まれる（図4-B）。ユーザーが体験をパーソナライズできるようにする。



(A) Main screen with video loaded.

(B) Customization screen.

図4. Smart Replayのウェブアプリケーション画面。

アクセシブル・リプレイ（P1）. 手動ナビゲーションを省略するリプレイボタン。アクセシビリティを強化したコンテンツセグメントを再生する。

時間調整（P1）. 認知負荷（発話速度、ハイライトなど）に応じて自動的に変化する、柔軟なビデオ速度。

OCR-音声クロスハイライト（P2）. 音声に基づいて画面上のテキストをハイライトする。学習者の注意を誘導し、フォローを助ける。

フレーム・アクティビティ・ハイライト（P2）. フレームアクティビティとは、ポインタの動きやスライド上の描画などのアクションを指す。これらはハイライトされ、ユーザーがどこに注意を向ければよいかを簡単に示すことができる。

セマンティックセグメンテーション（P3）. ビデオコンテンツからの意味情報に基づいた、きめ細かなセグメンテーション。ユーザーは簡単にコンテンツを選択し、ナビゲートし、概観することができる。

5.2 インタラクションの概要

図4-Aはメイン・ウェブ・インターフェースである。左側がビデオセクション、右側がセグメント化されたチャプターである。各メイン・セグメント（ショット）には、それぞれタイトルで定義されたセマンティック・セグメントのリストがある。ユーザーは興味のあるセグメントを見つけ、ビデオを簡単にナビゲートすることができる。

上部には、コントロールセクションがある。ユーザーは再生/一時停止▶、前の◀または次の▶セグメント、音量🔊、SRボタン🔍にアクセスできる。リプレイボタンを押すと、現在のセグメントが最初から再生され、ビデオにオーバーレイされた様々なアクセシビリティ機能が有効になる。

5.3 インタラクションの例

視覚障害を持つコンピュータ科学専攻のアリスの視点から、このシステムを動作を説明する。アリスは録画されたC++の講義を見ている。セマンティック・セグメンテーションによって、アリスは各章の概要を知り、簡単にナビゲートできる。

再生中、講師はスライドに絵を描いている。聴きながら講師が描いている場所を特定することは、アリスにとって難しい作業である。アリスの視覚障害は、このセクションを理解する能力を妨げ、認知的負荷を増加させる。アリスがSRボタンを押すと、そのセクションが再生される。大きな視覚的ハイライトがどこを見るべきかを示し、スピーチはゆっくりになり、アリスは要求される認知的負荷を減らすことができる。

後の章では、多くの文字情報を含むスライドが提示される。これはアリスに要求される認知的負荷を増加させる。アリスはもう一度SRボタンを押す。今度は、講師が話すテキストの一部が強調表示され、アリスは楽についていける。

6 実装

概要に続いて、システム機能がどのように実装されているかについて説明する。フレームアクティビティ、セマンティックセグメンテーション、OCR-音声クロスハイライトに重点を置く。YouTubeから取得したビデオコンテンツを使用する[33]。

6.1 アクセシブル・リプレイと時間調整

SR（再生速度）機能は、ユーザーがアクティブなビデオセグメントを再生し直す。現在の再生時刻とセグメントの開始時刻との間隔がある閾値（ $th_{SR} = 2sec$ ）より小さい場合、リプレイは前のセグメントの開始時刻に移動する。

これにより、ユーザーの認知負荷を軽減するため、時間の柔軟性を組み込んでいる。再生中、ビデオの

時間経過速度が変わる．システムによって導入された視覚的ハイライトに関しては，ユーザー設定に基づいてビデオをさらに一時的に遅く再生する（デフォルトは $0.75 \cdot V_{speed}$ ）．

さらに，すべてのセンテンスに対して，所望の1秒あたりの単語数（WPS）値（ U^{wps} ）に基づいて再生速度を調整する．したがって，文字起こし行 T_n に対して次のようになる：

$$T_n^{wps} = \frac{|T_n^{words}|}{T_n^{end} - T_n^{start}} \quad (1)$$

ここで， T_n^{end} および T_n^{start} はそれぞれ終了時刻と開始時刻である．したがって，ビデオ再生速度 V_{speed} は次のようになる：

$$V_{speed} = \frac{U^{wps}}{T_n^{wps}} \quad (2)$$

さらに，非音声の長い箇所を高速化することができる．時間の調整を行うことで，学習者が追従し，集中力を保つ能力が向上すると期待できる．

6.2 OCR-音声クロスハイライト

スライドの関連部分にユーザーの注意を誘導する．まず，OCRのショット候補を収集するために低感度で *scenedetect* を実行する．次に，各候補ショットからテキストを抽出する（*EasyOCR*⁴ を使用）．同一のテキストを持つ候補を統合する（図5）．

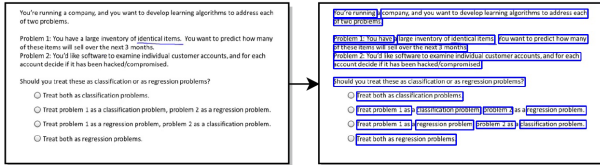


図 5. OCR 検出結果の例．

ビデオ再生中，トランスクリプションと適切なOCR要素をクロスリファレンスする（図6-A）．各OCRフラグメントについて，フラグメント fr_n とトランスクリプション行 T_n の間で共通の最も長い連続単語（ $W_n^{T,fr}$ ）を見つける．つまり， $|W_n^{T,fr}| \geq 2$ ．たとえば，トランスクリプションが "[...] try to predict the breast cancer [...]" であり，OCRフラグメントが "Breast cancer (malignant, benign)" の場合：

$$W_n^{T,fr} = ["breast", "cancer"] \quad (3)$$

その後，以下の手順に従う：

1. $|W_n^{T,fr}|$ が最大化される fr を検索する．
2. OCRフラグメントがフレーム $n-1$ に選択された場合，無視する．

残りのOCRフラグメントは，その境界ボックスに基づいてハイライト表示される．

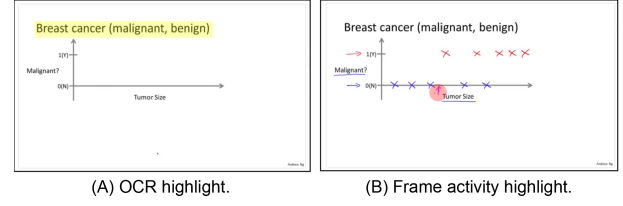


図 6. 異なる画面上のレンダリングハイライト，アクティビティ，および音声-OCR．

6.3 フレーム・アクティビティ・ハイライト

ビデオフレームで発生しているアクティビティを，2つのフレーム間の比較によって検出する．アクティビティとは，(1) マウス/ポインタの動き，(2) スライド上での描画，(3) 同じスライド上で新しいコンテンツの表示を指す（図6-B）．アルゴリズムの概要は図7に示されている．

ビデオフレームのウィンドウ F_w 上で作業する．ここで $|F_w| = V_{fps}$ (frames-per-second) である．最初と最後のフレームを比較する．OpenCV⁵の *absdiff* を使用してフレームの差分をチェックし，輪郭計算を適用する．各フレーム n の輪郭 i (つまり， $Ct_{n,i}$) は，次の簡略化された定義に変換される：

$$Ct_{n,i} = (A_{n,i}, P_{n,i}) \quad (4)$$

ここで， $A_{n,i}$ はそれがカバーする面積であり， $P_{n,i}$ は中心点である．これを使用して，中心から中心への距離に基づいて，見つかった輪郭をクラスターにグループ化するために連結コンポーネント（DFS）を使用する．すべての輪郭をノードとする完全に連結したグラフから始める．そして，中心点間の距離が与えられた閾値 $th_{RoI} = 0.1 \cdot \min(F_w, F_h)$ より大きい場合，エッジを反復的に削除する．ここで， F_w ， F_h はそれぞれフレームの幅と高さである．残った連結コンポーネントが注目領域（RoI）である．このようなRoIは次のように定義される：

$$R_{n,j} = (A_{n,j}, P_{n,j}, N_{n,j}) \quad (5)$$

ここで， $N_{n,j}$ はそれに含まれる輪郭の数， $P_{n,j}$ は輪郭の中心点， $A_{n,j}$ はその領域の面積で， $A_{n,j} = \sum_p A_{n,p}$ である， p はその領域内の輪郭である．

ビデオ再生中，現在の再生時刻を基にすべての関連するフレームアクティビティRoIを収集する．これらのRoIに対して，いくつかのアクションを実行する：

1. $\#(R_n) > th_R$ の場合，ハイライトをスキップする．ここで $th_R = 8$ である．

⁴ <https://github.com/jaidedai/easyocr>

⁵ <https://opencv.org/>

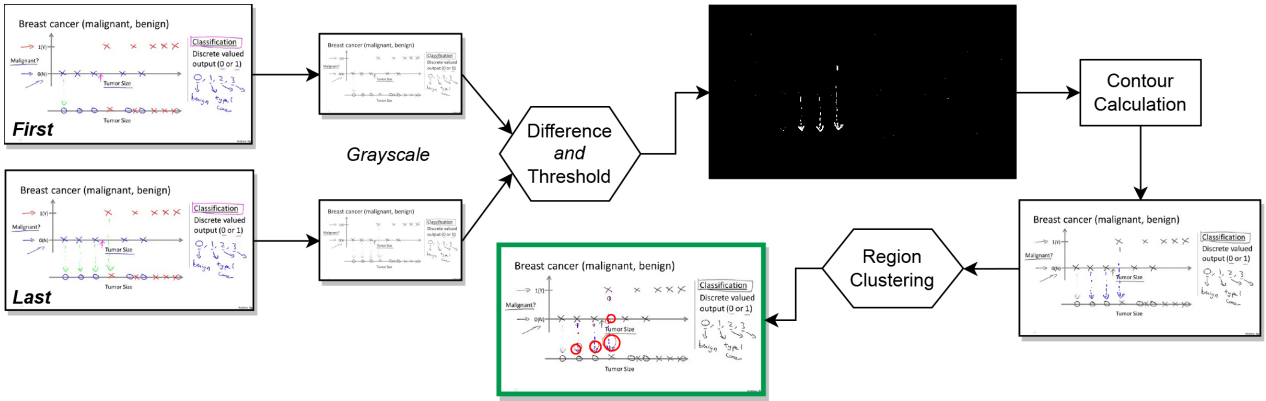


図 7. フレームアクティビティ検出アルゴリズムの概要. フレームのウィンドウで動作し, そのウィンドウ内の最初のフレームと最後のフレームを比較する.

2. 面積 ($A_{n,j}$) に基づいて RoI をソートする.
3. $A_{n,j} > 0.25 \cdot \max(F_w, F_h)$ または $A_{n,j} < 0.01 \cdot \min(F_w, F_h)$ の場合, RoI を無視する.
4. 以下のいずれかが前回のハイライトに対して正当化されているか確認する:
 - (a) 十分な時間が経過している $\Delta t \geq 10\text{sec}$.
 - (b) 中心点間の距離が大きい $P_{n,j} - P_{m,k} > 0.1 \cdot \min(F_w, F_h)$.

残りの RoI は, **ハイライト表示** され, その直径は $\sqrt{A_{n,j}} + A_b$ である. ここで A_b はユーザーによって設定された基本面積値である (デフォルトは 20px).

6.4 セマンティックセグメンテーション

まず, スライドのナビゲーションであるショット変更を *scenedetect*⁶ Python パッケージを使用して検出する. それから, OpenAI *Whisper*⁷ を使って文字起こしを行う. 意味的セグメンテーションアルゴリズムは, 各ショット S_i と文字起こし行 T_j に対して, $T_j \in S_i$ であるかどうかを計算する. 次に, 5つの異なるケースがある:

- C1 $T_j^{\text{start}} \geq S_i^{\text{start}}$ かつ $T_j^{\text{end}} \leq S_i^{\text{end}}$
- C2 $T_j^{\text{end}} \leq S_i^{\text{start}}$
- C3 $T_j^{\text{start}} \geq S_i^{\text{end}}$
- C4 $T_j^{\text{start}} < S_i^{\text{start}}$ かつ $S_i^{\text{start}} < T_j^{\text{end}} \leq S_i^{\text{end}}$
- C5 $T_j^{\text{end}} > S_i^{\text{end}}$ かつ $S_i^{\text{start}} \leq T_j^{\text{start}} < S_i^{\text{end}}$

C1 はショット内に完全に含まれているため, それに属する. C2 と C3 はショットの境界の外に完全に文字起こしに対応する. C4 と C5 は, 複数の

ショットにオーバーラップする中間の文字起こしに対応する. C4 と C5 に対して, 比率の閾値として $th_{ratio} = 0.5$ を導入する. C4 の検証:

$$\text{abs}\left(\frac{T_j^{\text{start}} - S_i^{\text{start}}}{T_j^{\text{end}} - T_j^{\text{start}}}\right) < th_{ratio} \quad (6)$$

C5 の検証:

$$\text{abs}\left(\frac{T_j^{\text{end}} - S_i^{\text{end}}}{T_j^{\text{end}} - T_j^{\text{start}}}\right) < th_{ratio} \quad (7)$$

各ショットセグメントごとに収集された文字起こし行を使用し, GPT-3.5[34] を活用して生成されたショットの文字起こしを文脈のある段落に分割する. それぞれにユーザーに提示される表題が付いている.

7 結論

私たちは, さまざまな能力を持つ多様なグループを対象に, eラーニングに関する広範な調査を実施した. その結果, 3つの重要な課題が明らかになった. 動画は (1) フォローしにくい, (2) 見づらい, (3) ナビゲートしにくい. 残念なことに, 既存のツールはこれらの課題に十分に対処できていない. 過去の研究では, これらの課題の側面について研究されているが, 認知的負荷や柔軟な時間の使い方といった本質的な考慮は, 顕著に欠けていた.

「Smart Replay」という革新的なソリューションを提案した. これは, eラーニングビデオを詳細に分析し, フレームアクティビティの拡張や OCR スピーチの統合などの機能を追加したものである. セマンティックセグメンテーションとリプレイ機能によって, コンテンツ内での移動が容易になり, 改善されたナビゲーションが実現される. 時間的な柔軟性とユーザーのカスタマイズを組み込むことで, 認知負荷のダイナミクスにより適合したインタラクティブで個人化された学習環境を作成することを目指している.

⁶ <https://www.scenedetect.com/>

⁷ <https://openai.com/research/whisper>

謝辞

本研究は Google の一部支援を受けて行われた。その経済的なサポートに感謝の意を表す。

参考文献

- [1] G. E. Acuna, L. A. Alvarez, J. Miraflores, and M. J. C. Samonte. Towards the Development of an Adaptive E-learning System with Chatbot Using Personalized E-learning Model. In *ICFET 2021: The 7th International Conference on Frontiers of Educational Technologies, Bangkok, Thailand, June 4 - 7, 2021*, pp. 120–125. ACM, 2021.
- [2] M. Agatha Priska, D. Aulia, E. Muslim, and L. Marcelina. Developing a framework to evaluate e-learning system at higher education in Indonesia. In *2020 The 4th International Conference on Education and E-Learning*, pp. 27–32, 2020.
- [3] M. Akram, M. W. Iqbal, M. U. Ashraf, E. Arif, K. Alsubhi, and H. M. Aljahdali. Optimization of Interactive Videos Empowered the Experience of Learning Management System. *Comput. Syst. Sci. Eng.*, 46(1):1021–1038, 2023.
- [4] W. Aljedaani, R. Krasniqi, S. Aljedaani, M. W. Mkaouer, S. Ludi, and K. Al-Raddah. If online learning works for you, what about deaf students? Emerging challenges of online learning for deaf and hearing-impaired students during COVID-19: a literature review. *Universal access in the information society*, pp. 1–20, 2022.
- [5] D. An. Introduction to Jordan Canonical Form, 2021. YouTube video.
- [6] A. S. Aydin, S. Feiz, V. Ashok, and I. V. Ramakrishnan. Towards making videos accessible for low vision screen magnifier users. In F. Paternò, N. Oliver, C. Conati, L. D. Spano, and N. Tintarev eds., *IUI '20: 25th International Conference on Intelligent User Interfaces, Cagliari, Italy, March 17-20, 2020*, pp. 10–21. ACM, 2020.
- [7] J. D. Basham, J. Blackorby, and M. T. Marino. Opportunity in crisis: The role of universal design for learning in educational redesign. *Learning Disabilities: A Contemporary Journal*, 18(1):71–91, 2020.
- [8] C. Batanero-Ochaíta, L. de Marcos, L. F. Rivera, J. Holvikivi, J. R. Hilera, and S. O. Tortosa. Improving Accessibility in Online Education: Comparative Analysis of Attitudes of Blind and Deaf Students Toward an Adapted Learning Platform. *IEEE Access*, 9:99968–99982, 2021.
- [9] X. A. Chen, C. Wu, T. Niu, W. Liu, and C. Xiong. Marvista: A Human-AI Collaborative Reading Tool. *CoRR*, abs/2207.08401, 2022.
- [10] N. A. Dahan, M. Al-Razgan, A. Al-Laith, M. A. Alsoufi, M. S. Al-Asaly, and T. Alfakih. Meta-verse framework: A case study on E-learning environment (ELEM). *Electronics*, 11(10):1616, 2022.
- [11] K. Davila, F. Xu, S. Setlur, and V. Govindaraju. FCN-LectureNet: Extractive Summarization of Whiteboard and Chalkboard Lecture Videos. *IEEE Access*, 9:104469–104484, 2021.
- [12] H. Deng. Strategies to Improve the Online-Teaching Effect of Colleges and Universities in the Post-Epidemic Era. In *CIPAE 2021: 2nd International Conference on Computers, Information Processing and Advanced Education, Ottawa, ON, Canada, 25-27 May, 2021*, pp. 786–789. ACM, 2021.
- [13] E. Djeki, J. Dégila, C. Bondiombouy, and M. H. Alhassan. E-learning bibliometric analysis from 2015 to 2020. *Journal of Computers in Education*, 9:727–754, 12 2022.
- [14] H. Haga. Combining video and bulletin board systems in distance education systems. *The Internet and Higher Education*, 5(2):119–129, 2002.
- [15] T. Han, M. Bain, A. Nagrani, G. Varol, W. Xie, and A. Zisserman. AutoAD: Movie Description in Context. *CoRR*, abs/2303.16899, 2023.
- [16] N. Harrati, I. Bouchrika, A. Tari, and A. Ladjailia. Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. *Comput. Hum. Behav.*, 61:463–471, 2016.
- [17] L. He, E. Sanocki, A. Gupta, and J. Grudin. Comparing presentation summaries: slides vs. reading vs. listening. In T. Turner and G. Szwillus eds., *Proceedings of the CHI 2000 Conference on Human factors in computing systems, The Hague, The Netherlands, April 1-6, 2000*, pp. 177–184. ACM, 2000.
- [18] K. Hornbæk, D. T. Engberg, and J. Gomme. Video lectures: HCI and e-learning challenges. In *Workshop on Human-computer interaction and E-learning*, 2002.
- [19] W. Huang, T. Xiao, Q. Liu, Z. Huang, J. Ma, and E. Chen. HMNet: a hierarchical multimodal network for educational video concept prediction. *International Journal of Machine Learning and Cybernetics*, pp. 1–12, 2023.
- [20] V. Ilin. The role of user preferences in engagement with online learning. *E-Learning and Digital Media*, 19(2):189–208, 2022.
- [21] A. Khaldi, R. Bouzidi, and F. Nader. Gamification of e-learning in higher education: a systematic literature review. *Smart Learn. Environ.*, 10(1):10, 2023.
- [22] J. Kong, D. Sabha, J. P. Bigham, A. Pavel, and A. Guo. TutorialLens: Authoring Interactive Augmented Reality Tutorials Through Narration and Demonstration. In F. R. Ortega, R. J. Teather, G. Bruder, T. Piumsomboon, B. Weyers, A. U. Batmaz, K. Johnsen, and C. W. Borst eds., *SUI '21: Symposium on Spatial User Interaction, Virtual Event, USA, November 9-10, 2021*, pp. 16:1–16:11. ACM, 2021.

- [23] A. Krouska, C. Troussas, and C. Sgouropoulou. Usability and Educational Affordance of Web 2.0 tools from Teachers' Perspectives. In N. N. Karanikolas, A. Voulodimos, C. Sgouropoulou, M. Nikolaidou, and S. Gritzalis eds., *PCI 2020: 24th Pan-Hellenic Conference on Informatics, Athens, Greece, 20-22 November, 2020*, pp. 107–110. ACM, 2020.
- [24] S. N. Kurbakova, Z. N. Volkova, and A. V. Kurbakov. Developing Students' Cognitive Abilities in E-Learning Environment. In *IC4E 2021: 12th International Conference on E-Education, E-Business, E-Management, and E-Learning, Tokyo Japan, 10 - 13 January, 2021*, pp. 124–130. ACM, 2021.
- [25] Z. Li and Z. Zhan. Integrated infrared imaging techniques and multi-model information via convolution neural network for learning engagement evaluation. *Infrared Physics & Technology*, 109:103430, 2020.
- [26] X. B. Liu, R. Wang, D. Li, X. A. Chen, and A. Pavel. CrossA11y: Identifying Video Accessibility Issues via Cross-modal Grounding. In M. Agrawala, J. O. Wobbrock, E. Adar, and V. Setlur eds., *The 35th Annual ACM Symposium on User Interface Software and Technology, UIST 2022, Bend, OR, USA, 29 October 2022 - 2 November 2022*, pp. 43:1–43:14. ACM, 2022.
- [27] X. Liu, P. Carrington, X. A. Chen, and A. Pavel. What Makes Videos Accessible to Blind and Visually Impaired People? In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, and S. M. Drucker eds., *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pp. 272:1–272:14. ACM, 2021.
- [28] Y. Liu. Design and Application of Hybrid Teaching Platform Based on Internet +. In *CIPAE 2021: 2nd International Conference on Computers, Information Processing and Advanced Education, Ottawa, ON, Canada, 25-27 May, 2021*, pp. 1490–1493. ACM, 2021.
- [29] K. Mack, E. McDonnell, D. Jain, L. L. Wang, J. E. Froehlich, and L. Findlater. What Do We Mean by "Accessibility Research"? A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, and S. M. Drucker eds., *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pp. 371:1–371:18. ACM, 2021.
- [30] I. A. Mastan, D. I. Sensuse, R. R. Suryono, and K. Kautsarina. Evaluation of distance learning system (e-learning): a systematic literature review. *Jurnal Teknoinfo*, 16(1):132–137, 2022.
- [31] G. N. V. P. and P. S. Interactive Audio Indexing and Speech Recognition based Navigation Assist Tool for Tutoring Videos. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, pp. 1678–1682, 2022.
- [32] R. Natalie, E. Jarjue, H. Kacorri, and K. Hara. ViScene: A Collaborative Authoring Tool for Scene Descriptions in Videos. In T. J. Guerreiro, H. Nicolau, and K. Moffatt eds., *ASSETS '20: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, Greece, October 26-28, 2020*, pp. 87:1–87:4. ACM, 2020.
- [33] A. Ng. Introduction to Machine Learning — ML-005 Lecture 1 — Stanford University — Andrew Ng, 2017. YouTube video.
- [34] OpenAI. OpenAI: Introducing ChatGPT, 2022. (accessed on July 3, 2023).
- [35] D. Pal and V. Vanijja. Perceived usability evaluation of Microsoft Teams as an online learning platform during COVID-19 using system usability scale and technology acceptance model in India. *Children and youth services review*, 119:105535, 2020.
- [36] Y. Peng, J. P. Bigham, and A. Pavel. Slidecho: Flexible Non-Visual Exploration of Presentation Videos. In J. Lazar, J. H. Feng, and F. Hwang eds., *ASSETS '21: The 23rd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, USA, October 18-22, 2021*, pp. 24:1–24:12. ACM, 2021.
- [37] Y. Peng, P. Chi, A. Kannan, M. R. Morris, and I. Essa. Slide Gestalt: Automatic Structure Extraction in Slide Decks for Non-Visual Access. In A. Schmidt, K. Väänänen, T. Goyal, P. O. Kristensson, A. Peters, S. Mueller, J. R. Williamson, and M. L. Wilson eds., *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI 2023, Hamburg, Germany, April 23-28, 2023*, pp. 829:1–829:14. ACM, 2023.
- [38] Y. Peng, J. Jang, J. P. Bigham, and A. Pavel. Say It All: Feedback for Improving Non-Visual Presentation Accessibility. In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, and S. M. Drucker eds., *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pp. 276:1–276:12. ACM, 2021.
- [39] G. Ravichandran, D. J. Sujathamalini, and D. K. Gunasekaran. E-Learning-Accessibility of Students with Visual Impairment in Higher Education. *International Journal of Research and Review*, 9(5):27–31, 2022.
- [40] S. Russ and F. Hamidi. Online learning accessibility during the COVID-19 pandemic. In S. R. Vázquez, T. Drake, D. Ahmetovic, and V. Yaneva eds., *W4A '21: 18th Web for All Conference, Virtual Event / Ljubljana, Slovenia, April 19-20, 2021*, pp. 8:1–8:7. ACM, 2021.
- [41] A. Sackl, F. Graf, R. Schatz, and M. Tschelegi. Ensuring Accessibility: Individual Video Playback Enhancements for Low Vision Users. In T. J. Guerreiro, H. Nicolau, and K. Moffatt eds., *ASSETS '20: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, Greece, October 26-28, 2020*, pp. 67:1–67:4. ACM, 2020.

- [42] C. Smith. Challenges and opportunities for teaching students with disabilities during the COVID-19 pandemic. *International Journal of Multidisciplinary Perspectives in Higher Education*, 5(1):167–173, 2020.
- [43] B. Song, Y. Sun, J. Guo, D. Zhao, and J. Tan. Application of case teaching method and Microteaching Method in experimental teaching based on wechat platform. In *BIC 2021: International Conference on Bioinformatics and Intelligent Computing, Harbin, China, January, 2021*, pp. 407–411. ACM, 2021.
- [44] T. Sukardiyono, M. I. Luthfi, and N. D. Septiyanti. Breaking Down Computer Networking Instructional Videos: Automatic Summarization with Video Attributes and Language Models. *Elinvo (Electronics, Informatics, and Vocational Education)*, 8(1):26–37, 2023.
- [45] J. C. Sun and K. Chang. Design and development of a location-based mobile learning system to facilitate English learning. *Univers. Access Inf. Soc.*, 15(3):345–357, 2016.
- [46] ThatEngineering. Denavit - Hartenberg (DH) Tables For Robotic Systems - Direct Kinematics II, 2020. YouTube video.
- [47] F. Vindigni, T. Lasala, C. Giraud, C. Furiassi, M. Caramagna, M. Bosco, and G. Baratto. Implementing Universal Design for Learning: An Online Training Course for Peer Tutors to Students with Disabilities and Specific Learning Disorders: Implementing UDL: An Online Training Course for Peer Tutors to Disabled and SLD Students. In *Proceedings of the 2022 6th International Conference on Education and E-Learning*, pp. 157–161, 2022.
- [48] Y. Wang, W. Liang, H. Huang, Y. Zhang, D. Li, and L. Yu. Toward Automatic Audio Description Generation for Accessible Videos. In Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjørn, and S. M. Drucker eds., *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pp. 277:1–277:12. ACM, 2021.
- [49] B. F. Yuksel, P. Fazli, U. Mathur, V. Bisht, S. J. Kim, J. J. Lee, S. J. Jin, Y. Siu, J. A. Miele, and I. Yoon. Human-in-the-Loop Machine Learning to Increase Video Accessibility for Visually Impaired and Blind Users. In R. Wakkary, K. Andersen, W. Odom, A. Desjardins, and M. G. Petersen eds., *DIS '20: Designing Interactive Systems Conference 2020, Eindhoven, The Netherlands, July 6-10, 2020*, pp. 47–60. ACM, 2020.
- [50] D. Zhang, L. Zhou, R. O. Briggs, and J. F. N. Jr. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Inf. Manag.*, 43(1):15–27, 2006.
- [51] H. Zhou, R. Martín-Martín, M. Kapadia, S. Savarese, and J. C. Niebles. Procedure-Aware Pretraining for Instructional Video Understanding. *CoRR*, abs/2303.18230, 2023.

未来ビジョン

我々のロードマップには、効果、インタラクションパターン、および使いやすさに関する標準的なユーザースタディと、実世界のeラーニングのパフォーマンスに関する探索的なスタディが含まれる。このアプローチにより、ユーザー・エクスペリエンスを包括的に理解することができる。

さらに、今後、アクセシビリティのさらなる向上を探求したいと考えている。認知的負荷の軽減という観点に焦点を当て、理解を深めるために、ビデオの特定の領域を拡大（ズーム）することを検討する。

拡大表示は以前の研究 [6] でも検討されたツールであるが、提案された視線ベースのソリューションは、eラーニングコンテンツに適用すると失敗する。加えて、提示されたインタラクション手法は認知負荷の軽減には逆効果であり、改善できると考えている。

また、クリエイターのためのオーサリングツールの開発を模索している。「ヒューマン・イン・ザ・ループ」アプローチを採用し、eラーニング・ビデオのアクセシビリティを高めるために活用することができるツール。SRシステムを活用することで、カスタマイズされたバリアフリーな学習体験を生み出すことを目指している。