SmartLearn: Visual-Temporal Accessibility for Slide-based *e-learning* Videos

Yotam Sechayk
The University of Tokyo
Japan, Tokyo
sechayk-yotam@g.ecc.u-tokyo.ac.jp

Ariel Shamir
Reichman University
Israel, Herzliya
arik@runi.ac.il

Takeo Igarashi
The University of Tokyo
Japan, Tokyo
takeo@acm.com

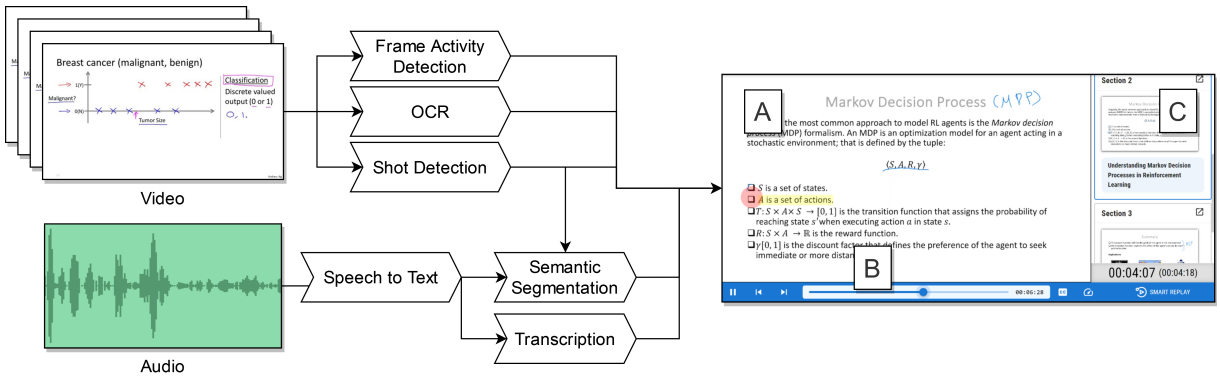

Figure 1: SmartLearn (SL) system diagram and pipeline. From Video analysis (left) to the video playback synthesis (right). A=main video area, B=video controls, C=chapter area.

ABSTRACT

In the realm of e-learning, video-based content is increasingly prevalent but brings with it unique accessibility challenges. Our research, beginning with a formative study involving 53 participants, has pinpointed the primary accessibility barriers in video-based e-learning: mismatches in user pace, complex visual arrangements leading to unclear focus, and difficulties in navigating content. To tackle these barriers, we introduced SmartLearn (SL), an innovative tool designed to enhance the accessibility of video content. SL utilizes advanced video analysis techniques to address issues of focus, navigation, and pacing, enabling users to interact with video segments more effectively through a web interface. A subsequent evaluation demonstrated that SL significantly enhances user engagement, ease of access, and learnability over existing approaches. We conclude by presenting design guidelines derived from our study, aiming to promote future efforts in research and development towards a more inclusive digital education landscape.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility systems and tools**; Web-based interaction; **User studies**; • **Applied computing** → **E-learning**; *Interactive learning environments.*

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3650883

KEYWORDS

E-learning, Online learning, Accessibility, Video Accessibility, Visual Accessibility, Temporal Accessibility, Universal Design

ACM Reference Format:
Yotam Sechayk, Ariel Shamir, and Takeo Igarashi. 2024. SmartLearn: Visual-Temporal Accessibility for Slide-based *e-learning* Videos. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3613905.3650883

1 INTRODUCTION

E-learning, facilitated by the internet, offers broad access to education. With the COVID-19 pandemic causing a surge in popularity. Particularly, with educational video content [2]. Videos are engaging and effective. Yet, pose unique challenges like temporal information scattering, and diverse visual language [16, 24, 44]. Despite growing focus on accessibility and e-learning, video-based e-learning still faces significant access barriers. Existing tools, like closed captioning (CC) and audio description (AD), address a limited subset of barriers [8, 14, 22, 32].

This paper introduces **SmartLearn** (SL), a tool designed to address slide-based video challenges. Based on a formative web-survey with $N = 53$ participants, we locate existing challenges and formulate the motivation. SL leverages video analysis to enhance playback by targeting unclear focus, content navigation, and pace disparities. A user study involving $N = 14$ participants demonstrated SL's effectiveness in addressing these challenges. Finally, we outline limitations, areas for future work, and fundamental design guidelines for accessible e-learning.

Our contributions are as follows:

- Formative study of accessibility challenges, with utilization patterns of existing tools.
- **SmartLearn**, a novel visual-temporal approach for enhanced accessibility.
- User evaluation with learnability metrics.
- Design guidelines for accessible e-learning systems.

2 BACKGROUND

2.1 E-learning

Alongside the growth of e-learning, platforms like Coursera[1] and Zoom[2] [23] gained popularity. And prior work, explored usability of these and other platforms [30, 40]. As well as methods to evaluate effectiveness of e-learning platforms [1, 15]. Additionally, suggesting different content creation and design strategies [20, 21, 43]. Accessibility research on e-learning mostly explores specific groups of users [19, 36, 39]. With some studies taking a more Universal Design (UD) approach [4, 37]. Similarly, our study aims to address a broad range of accessibility needs by exploring a universal approach. We argue that accessibility tools benefits individuals with and without disabilities, thus promoting inclusion and equity.

2.2 Accessibility for Video

As noted, videos present various access barriers. Prior work explored systems and methods to improve AD content creation [25, 45, 48], static visual filters [38] and gaze-estimated video magnification [3]. All aimed at blind and low vision (BLV) users. Other works explored temporal accessibility, using machine learning (ML) to simplify content [42], summarize videos [7, 11], and support navigation [6, 24]. Our approach combines both visual and temporal methods, along with other techniques to address and provide a more comprehensive approach.

2.3 Slide-based Video Accessibility

Slide-based videos are a subset of e-learning videos, where the content is presented in a slide-like format. Prior work supports BLV learners [10, 33, 34] and instructors [35]. Additionally, note-taking tools [47] and semantic navigation [18, 46, 49] have been explored. Our work approaches accessibility through universal design paradigms, incorporating techniques seen in these works, and expanding on them with novel approaches to offer an inclusive solution for general users. Utilizing both the visual and temporal dimensions.

3 FORMATIVE STUDY: UNDERSTANDING *E-LEARNING* CHALLENGES

3.1 Method

As a first step, we conducted a formative web-survey. We surveyed 53 diverse participants (education, ages, interests, and disability). Results indicate that available accessibility tools are insufficient, with several issues raised. Including timing discrepancies, visual complexities, unclear audio, and a lack of audio enhancements and controls. Participants were recruited through social media and mailing lists. The survey was conducted online, and included 10 fully-optional questions (multiple-choice and open-ended). Participation was uncompensated, with the survey taking 5–10 minutes. This study was approved by the university's ethics board.

[1]https://www.coursera.org/
[2]https://zoom.us/

3.2 Results

Among participants, 27 are 26–35, 15 are 18–25, 7 are 46–60, and 2 are 36–45 and 61+ respectively. Gender related information was not collected. Regarding disabilities, 7 identify as non-disabled, 11 as deaf or hard of hearing (DHH), 11 as blind or low vision (BLV), and more. See Table 3 in Appendix A for detailed information.

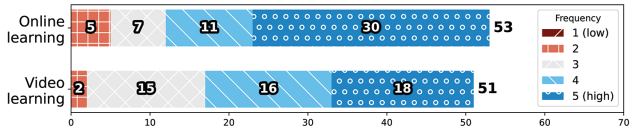

Figure 2: Frequency of participants to use online resources for studying, as well as specifically video resources for studying. Questions are optional, therefor the total number of participants is not consistent.

Part 1: E-learning usage (Q1-Q6). Participants are highly likely to use online learning (Figure 2). Reading content is used by 49 participants, videos by 45, and 37 join online classes. Total of 30 utilize all. Additionally, 43 watch instructional videos, 34 lecture recordings, and 28 video essays, while 17 consume all. Common types of content watched include, specialized topics (38), cooking (24), math (18), and programming (14) among others. YouTube[3] has emerged as the most used platform (47). Social media platforms also emerged as learning resource for some [12, 13, 40].

Part 2: Learning video user experience (Q7-Q8). Participants prefer a PC (42) over smartphones (9). With various difficulties experienced by most participants. We split these to temporal, visual, and audio difficulties. See Table 5 in Appendix A for details. Importantly, we note that all non-disabled participants face numerous difficulties.

Part 3: Existing tools and unaddressed issues (Q9-Q10). Only a few tools are being used, 36 use video speed changes, 27 use captions, 20 use noise-reducing headphones, and 8 use screen reader software. Other mentions include screen magnifier, color filters, and eye-tracked cursor, among others. Many accessibility needs still exist. Lack of voice and audio enhancements (P13, P14, P17), and unreliable captioning (P11, P12, P16, P20, P23). Additionally, limited caption customization (P23), and speedy captions (P28). For BLV users, lack of screen reader support (P19, P23), and shallow descriptions (P5, P19). Easier navigation (P6, P37, P53), and more interactive videos (P7, P25, P45, P50) are desired.

3.3 Key Findings

Access barriers persist for all learners. The current tool landscape leaves much to be desired. And while customization is explored [5, 17], the impact it can have on accessibility, demands further analysis.

Based on our findings, we distinguish three key problems:

[3]https://www.youtube.com/

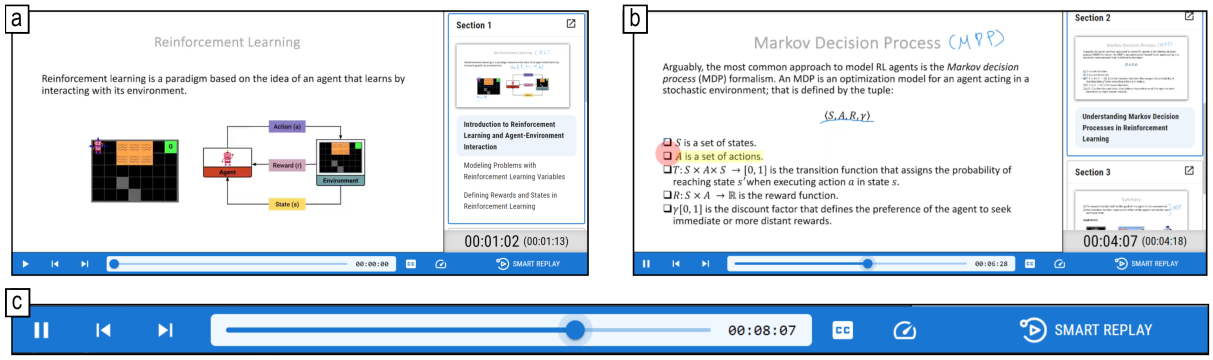

Figure 3: SmartLearn application interface. (a) Main screen, (b) Presenting visual highlights, (c) Video control bar.

X1 *(hard to follow)* Video **temporal** pace is not matching with user abilities.
X2 *(hard to see)* Overwhelming **visuals** can lead to misplaced attention.
X3 *(hard to navigate)* **Searching** in videos is complicated, relying on many user actions.

4 SMARTLEARN

In order to address X1-X3, we propose *SmartLearn* (SL), an e-learning system for accessible video playback. Targeting slide-based videos with narration and live sketching. SL is a web application built using Flask[4] and React[5].

Our approach works with audiovisual data. Using Optical Character Recognition (OCR), transcription, and shot detection with existing tools. Alignment of speech and OCR, similar to existing work [18, 33, 35]. With chapter segmentation using Large Language Models (LLM), and frame-activity detection as novel approaches. See Figure 1 for an overview, and Figure 3 for the interface.

4.1 System Features

- **Quick Replay (X1).** Replay a recent content segment, omitting the need for manual navigation.
- **Temporal Adjustments (X1).** Flexible speed that changes automatically based on the desired speed of speech.
- **OCR-Speech Cross Highlight (X2).** Highlighting on-screen text based on speech.
- **Frame-Activity Highlight (X2).** Highlighting actions such as pointer movements, and sketching on the slides.
- **Semantic Segmentation (X3).** Automatic segmentation of shots based on content.

4.2 Implementation

4.2.1 Temporal Adjustments. We adjust the playback speed for each transcription line. This is based on a desired words-per-second (WPS) value selected by the user. For each line we calculate the speed based on the ratio between the WPS for that line and the desired WPS value. Implementation details available in Appendix B.1.

4.2.2 OCR-Speech Cross Highlight. Example output is seen in Figure 3-b. Similar to existing work [18, 33, 35]. We use *EasyOCR*[6] to extract slide text. During video playback, match speech with OCR fragments. Using the longest consecutive words in common ($W_n^{T,fr}$) between the fragment fr_n and the transcription T_n, such that $|W_n^{T,fr}| \geq 2$ and is maximizes. The result is highlighted based on its bounding-box. Implementation details available in Appendix B.2.

4.2.3 Frame Activity Highlight. Example output is seen in Figure 3-b. Activity refers to (1) mouse/pointer movements, (2) sketching on the slide, (3) revealing new content on the same slide. For a window of frames F_w (s.t., $|F_w| = V_{fps}$), we compare the first and last frames. Using OpenCV[7] `absdiff`, find the contour of differences. We create regions of interest (RoI) which are used to define highlights. During playback, we use a highlighted circle with customized diameter (default $50px$). See Figure 4 for an overview of the algorithm. Implementation details available in Appendix B.3.

4.2.4 Semantic Segmentation. Example output is visible in Figure 3-a. We first detect shot changes (i.e., slides navigation) using the *scenedetect* python package. Using transcription generated with OpenAI *Whisper*[8] (with model `medium.en`), we collect transcription lines for each shot. Per shot, GPT-3.5 [29] is used to split the text into contextual paragraphs. Each with an easy-to-understand title. Implementation details available in Appendix B.4.

5 EVALUATION

Our formative study highlighted a demand for an engaging learning experience, improving on challenging navigation, and video pace disparities. We hypothesize that SL enhances the learning experience by streamlining these elements.

Specifically, we hypothesize that SmartLearn will:

H1 Promote a more active and engaging learning process.
H2 Increase the learnability without negatively impacting the workload.
H3 Improve the user experience of slide-based video learning.

[4] https://flask.palletsprojects.com/
[5] https://reactjs.org/
[6] https://github.com/jaidedai/easyocr
[7] https://opencv.org/
[8] https://openai.com/research/whisper

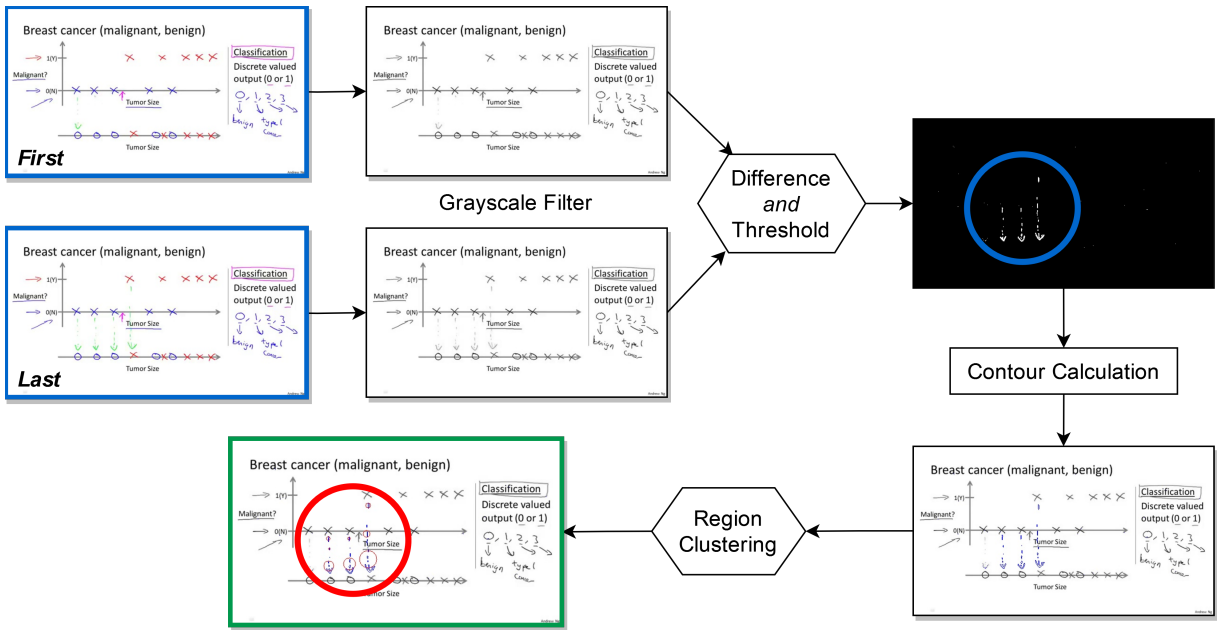

Figure 4: Overview of Frame Activity detection algorithm. Working over a window of frames, comparing the *first* and *last* frames in that window.

5.1 Method

We conducted a within-subject user study with 14 participants. The study was conducted using Zoom and included three parts: (1) introduction task, (2) two main tasks, and (3) subjective evaluation with semi-structured interview. Each task presented a learning video followed by a quiz. Participants are recruited through mailing lists and social media. The participants were compensated with a $20 gift card for their time[9]. The study has been approved by the university's ethics board. Additional information is available in Appendix C. Statistical significance uses *single-tailed T-test p-value.*

5.2 Results

5.2.1 Interaction logs. Task completion time between SmartLearn and the baseline was not significantly different ($\mu = 941.19$, $\sigma = 87.43$ vs. $\mu = 864.36$, $\sigma = 126.17$; $p = 0.0572$). However, the video playback time slightly increased using the proposal ($\mu = 686.80$, $\sigma = 52.91$ vs. $\mu = 610.03$, $\sigma = 69.63$; $p = 0.0045$), while the quiz time was virtually indistinguishable ($\mu = 258.93$, $\sigma = 73.54$ vs. $\mu = 254.32$, $\sigma = 105.37$; $p = 0.4533$). See Figure 10 in Appendix D.

Interaction logs reveal usage differences. The average interaction count is not significantly different ($\mu = 107.07$, $\sigma = 233.83$ vs. $\mu = 113.92$, $\sigma = 239.23$; $p = 0.5303$), but increased slightly. The different is more evident when separating videos. The first has a relatively similar usage pattern, while the second presents clear incline for interaction using SmartLearn. See Figure 5 and Figure 6. Additional details in Appendix D. This supports our first hypothesis.

A total of 8 interaction types exist. Some are shared, while others are unique to SL. See Table 1 for a list of all interaction types and the corresponding usage.

[9]For some participants, a similar amount in their currency of choice was distributed

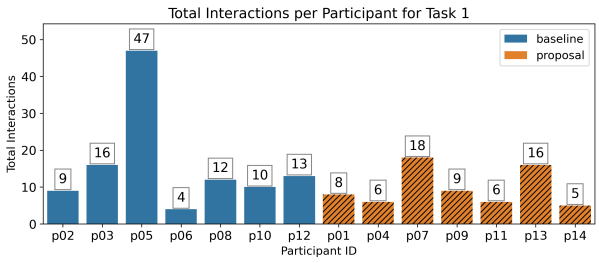

Figure 5: Average interactions for the video of Task 1 (baseline vs. SmartLearn).

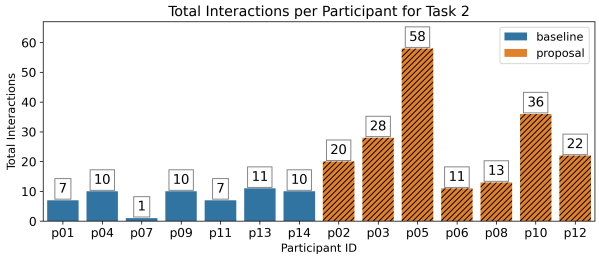

Figure 6: Average interactions for the video of Task 2 (baseline vs. SmartLearn).

Looking at `Chapter Click` and Seek operations, SL has increased usage. While seek interactions remained largely the same,

Table 1: Total of each interaction type when comparing SmartLearn to the baseline. Items marked with * are not available in the baseline.

	Baseline	SmartLearn
Play/Pause	53	73
Speed Change	39	54
Seek	62	63
Chapter Click	3	28
Next*	–	3
Previous*	–	1
Replay*	–	21

participants greatly utilized more detailed chapters. Especially during Task 2, which some argued is more challenging. See Table 1. This supports our first hypothesis.

5.2.2 Quiz scores. The overall score was not significantly different ($\mu = 3.42$, $\sigma = 1.16$ vs. $\mu = 3.42$, $\sigma = 0.93$; $p = 0.5$). However, 6 participants increased their score, 3 participants maintained it, and only 5 participants lowered it. See Table 2. This supports our second hypothesis, although larger evaluation is required for a stronger correlation.

5.2.3 Subjective evaluation. Overall, participants rated SmartLearn as significantly **less** mentally demanding ($\mu = 2.64$, $\sigma = 1.08$ vs. $\mu = 3.5$, $\sigma = 0.85$; $p = 0.0142$), **more** easy to use ($\mu = 4.57$, $\sigma = 0.51$ vs. $\mu = 3.78$, $\sigma = 0.97$; $p = 0.0074$), and **more** useful ($\mu = 4.28$, $\sigma = 0.72$ vs. $\mu = 3.35$, $\sigma = 0.84$; $p = 0.0022$). Moreover, most participants prefer using SL in the future (10 vs. 4). See Figure 7. This supports our third hypothesis.

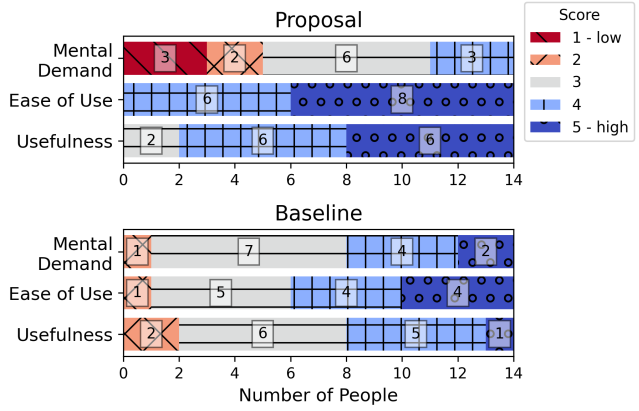

Figure 7: Subjective evaluation results of participants using SmartLearn and the baseline. Note that for *Usefulness* and *Ease of Use* higher is better, while for *Mental Demand* lower is better (1 = low, 5 = high). All metrics are statistically significant.

5.2.4 Semi-structured interview. Generally, participants reported positive impact of SL on their learning, particularly highlighting its effectiveness in making complex content more accessible. The chapter feature was favored by most for its navigational ease (P1-P6, P8, P9, P13), followed by text and pointer highlights for enhancing engagement and attention (P1, P2, P5, P6, P7, P11, P12, P13).

Visual enhancements. Pointer highlights were deemed beneficial for maintaining focus, especially for those dependent on captions or struggling with cursor visibility (P1, P6, P7, P11, P12). P1 likened the pointer highlight to *"a laser pointer in class,"*. Some participants, however, found them unnecessary, citing a natural focus on the instructor's writing (P8, P13). Text highlights were appreciated for enhancing video content immersion, especially in text-heavy presentations (P1, P4, P5, P13), though some viewed them as redundant, preferring only captions (P7).

Temporal interactivity. The chapter function was valued for its navigational benefits, allowing easy revisiting of video sections (P1, P3, P5-P8, P10-P13), with some users desiring even more detailed segmentation (P6, P11, P12). Despite the slider tool's precision, its use was noted as cumbersome by some (P1-P4, P6, P7, P9). Concerns about the replay function's clarity were raised, with a preference for more predictable controls, such as replaying specific sentences (P6, P8, P11-P13). Speech speed adjustments received mixed reviews. Some found it helpful for maintaining attention or reducing boredom by varying the pace (P4, P14), whereas others felt the changes were unpredictable and unnatural (P1, P3, P5-P8, P10-P14).

6 DISCUSSION

6.1 Reflections

6.1.1 Chapter Generation Trade-off. Utilizing LLMs for chapter generation highlighted their summarization strengths, as seen with participants finding LLM-generated chapters clearer than other platforms (P8). While the current chapter amount was generally sufficient (P1, P8, P13), some participants believed additional chapters could offer better detail and possibly negate the slider's necessity (P6, P11, P12). However, too many chapters could confuse users, indicating a balance is needed between detail and simplicity.

6.1.2 Replay Button Underutilization. The `Replay` feature was underused, potentially due to its overlap with chapter functionality, or unclear outcome expectations. For example, P8 was uncertain about what `Replay` would do. This suggests that more transparent design for the replay function, or integrating more granular subchapters into the `Replay` feature could increase its usage and utility.

6.1.3 Speed Change Strategies. Typically, speed adjustment use a multiplier (e.g., x0.5, x1.5). SL, matches actual speed with a desired WPS rate. This approach resulted in unnatural speech. P1 stated, *"It feels unnatural and distracting [...] I feel more comfortable if I was notified about speed changes before they occur."* P14 mentioned, *"it [WPS] doesn't indicate much for me."* However, some found value in these modifications; P4 observed, *"lowered speed signals an important part, making me more attentive."* Thus, the changes could be utilized to enhance engagement among viewers. Therefor, speech-pattern preserving speed modifications could be a potential future improvement. And while research exists on the effect of speed on learning [27, 28, 31], the impact of speed changes on speech patterns and the relation to attention could be further explored.

Table 2: Quiz scores per participant comparing the baseline to SmartLearn. B marks the baseline score, SL marks the SmartLearn score and C marks the type of change, positive +, negative −, or unchanged ∘

PID	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14
B	4	4	5	2	4	2	4	4	2	3	3	4	3	4
SL	2	3	2	2	4	3	5	3	4	5	4	2	5	4
C	−	−	−	∘	∘	+	+	−	+	+	+	−	+	∘

6.1.4 Learning with SmartLearn. While the evaluation scenario of SL is limited, the potential our approach has on making learning more accessible is visible. Similar to other works, it is challenging to see clear benefit based on quiz results in such learning scenario [18]. The act of learning is done out of interest, necessity or desire, all of which could be considered as a form of investment. Such investment must originate with the learner, to form motivation. We believe that a longer exploratory study as future work could provide more insights into the learning process and the impact of SL on the results of learning.

6.2 Limitation

6.2.1 Participant Distribution. The study's findings are constrained by its small sample of 14 participants. Expanding and diversifying the sample could enhance the validity of the findings.

6.2.2 Time Constraints. SmartLearn's novel interface posed a learning curve. Participants' reliance on familiar features like the seek slider indicates a comfort with traditional tools. The short study period also limited insights into SmartLearn's long-term impact, contrasting with the likely use in authentic learning scenarios. Future studies over a longer period could capture the platform's efficacy in real-world conditions.

6.3 Design Guidelines

We present design guidelines for universal accessibility of e-learning systems. These guidelines are based on the results of our study, participant feedback, and what we learned. It is not meant to be a complete list, but rather a starting point for future work to further explore and refine.

We argue that universal accessibility for e-learning could be achieved when:

- G1 The design is *responsive* to both user actions and the content.
- G2 The user experience is *customizable* on all dimensions.
- G3 The design is *transparent* and clearly communicates the current state.

6.3.1 Responsiveness (G1). Responsive design is the idea responding to actions of the user, and is a staple of web-design. But we believe responsiveness should relate to the content as well. P13 commented, *"Speed changes can work with highlights, slowing down (the video) when some text is highlighted."* P4 commented, *"It will be useful if I can know if the upcoming part of the video is important."* P13 added, *"I expected for the video to speed up when the instructor was silent."* Both P4 and P13 are referring to the idea of responsiveness to the content.

6.3.2 Customization (G2). Customization empowers users by allowing them to tailor the system to their preferences, which is crucial for students with disabilities. It should be a design goal across all system dimensions, including visual, audio, and interactivity. Users expressed a desire for more control over replay functionality (P3, P13) and chapter markers (P3, P6, P11, P12), emphasizing the importance of customization in enhancing user control, which rewards users for their engagement.

6.3.3 Transparency (G3). Transparency is not a new concept [9, 41], and similarly we stress the importance of transparency in e-learning systems. P6 commented, *"I want to be notified when the speaker mentions key points."* Additionally, P6, P8, P11, P12, and P13 all expressed the need for a more transparent design.

7 CONCLUSION

E-learning enhances educational accessibility but confronts unique challenges, particularly with video content, a primary e-learning medium. Traditional accessibility tools like captions fall short in addressing all user needs, revealed in our formative study with 53 participants. This study uncovered universal access barriers across diverse user abilities, highlighting issues such as cognitive pace disparities, unclear attention focus, and complex content navigation.

Addressing these challenges, we developed SmartLearn, a video player that leverages video analysis to adapt the playback experience to user needs. By highlighting key information and simplifying navigation with LLM-generated chapters, SmartLearn improved learning outcomes and user satisfaction in a study with 14 participants. Our findings underscore the potential of customized, intelligent e-learning tools to break down accessibility barriers, paving the way for future research and design strategies in e-learning video accessibility.

ACKNOWLEDGMENTS

Special acknowledgment must be given to the Barrier-free Offices at both the University of Tokyo and Tel-Aviv University. Their assistance was crucial in facilitating the user studies that this research hinged upon, and their support has been invaluable.

Moreover, I am deeply indebted to the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) for their generous scholarship, affording me the extraordinary opportunity to pursue advanced studies in Japan.

My gratitude extends to Google for the endorsement provided through the Google AIR Grant, without which this study could not have reached its full potential.

REFERENCES

[1] Meilia Agatha Priska, Dilla Aulia, Erlinda Muslim, and Lidya Marcelina. 2020. Developing a framework to evaluate e-learning system at higher education in Indonesia. In *2020 The 4th International Conference on Education and E-Learning*. 27–32.

[2] Shakeel Ahmad, Ahmad Shukri Mohd Noor, Ali A Alwan, Yonis Gulzar, Wazir Zada Khan, and Faheem Ahmad Reegu. 2023. eLearning acceptance and adoption challenges in Higher Education. *Sustainability* 15, 7 (2023), 6190.

[3] Ali Selman Aydin, Shirin Feiz, Vikas Ashok, and I. V. Ramakrishnan. 2020. Towards making videos accessible for low vision screen magnifier users. In *IUI '20: 25th International Conference on Intelligent User Interfaces, Cagliari, Italy, March 17-20, 2020*, Fabio Paternò, Nuria Oliver, Cristina Conati, Lucio Davide Spano, and Nava Tintarev (Eds.). ACM, 10–21. https://doi.org/10.1145/3377325.3377494

[4] James D Basham, Jose Blackorby, and Matthew T Marino. 2020. Opportunity in crisis: The role of universal design for learning in educational redesign. *Learning Disabilities: A Contemporary Journal* 18, 1 (2020), 71–91.

[5] Xiang 'Anthony' Chen, Chien-Sheng Wu, Tong Niu, Wenhao Liu, and Caiming Xiong. 2022. Marvista: A Human-AI Collaborative Reading Tool. *CoRR* abs/2207.08401 (2022). https://doi.org/10.48550/arXiv.2207.08401 arXiv:2207.08401

[6] Ioana Croitoru, Simion-Vlad Bogolin, Samuel Albanie, Yang Liu, Zhaowen Wang, Seunghyun Yoon, Franck Dernoncourt, Hailin Jin, and Trung Bui. 2023. Moment Detection in Long Tutorial Videos. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*. IEEE, 2594–2604. https://doi.org/10.1109/ICCV51070.2023.00245

[7] Kenny Davila, Fei Xu, Srirangaraj Setlur, and Venu Govindaraju. 2021. FCN-LectureNet: Extractive Summarization of Whiteboard and Chalkboard Lecture Videos. *IEEE Access* 9 (2021), 104469–104484. https://doi.org/10.1109/ACCESS.2021.3099427

[8] Essohanam Djeki, Jules Dégila, Carlyna Bondiombouy, and Muhtar Hanif Alhassan. 2022. E-learning bibliometric analysis from 2015 to 2020. *Journal of Computers in Education* 9 (12 2022), 727–754. Issue 4. https://doi.org/10.1007/s40692-021-00218-4

[9] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. *23rd International Conference on Intelligent User Interfaces* (2018). https://api.semanticscholar.org/CorpusID:3788634

[10] Danyang Fan, Sasa Junuzovic, John C. Tang, and Thomas Jaeger. 2023. Improving the Accessibility of Screen-Shared Presentations by Enabling Concurrent Exploration. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2023, New York, NY, USA, October 22-25, 2023*. ACM, 44:1–44:16. https://doi.org/10.1145/3597638.3608411

[11] Solomon E. Garber, Aaditya Prakash, Nick Moran, Maria Altebarmakian, Richard Alterman, Antonella Di Lillo, and James A. Storer. 2017. A two tier approach to chalkboard video lecture summary. In *2017 IEEE Frontiers in Education Conference, FIE 2017, Indianapolis, IN, USA, October 18-21, 2017*. IEEE Computer Society, 1–9. https://doi.org/10.1109/FIE.2017.8190536

[12] Manuel B Garcia, Irish C Juanatas, and Roben A Juanatas. 2022. TikTok as a Knowledge Source for Programming Learners: a New Form of Nanolearning?. In *2022 10th International Conference on Information and Education Technology (ICIET)*. IEEE, 219–223.

[13] Sourojit Ghosh and Andrea Figueroa. 2023. Establishing TikTok as a Platform for Informal Learning: Evidence from Mixed-Methods Analysis of Creators and Viewers. In *56th Hawaii International Conference on System Sciences, HICSS 2023, Maui, Hawaii, USA, January 3-6, 2023*, Tung X. Bui (Ed.). ScholarSpace, 2431–2440. https://hdl.handle.net/10125/102931

[14] Tengda Han, Max Bain, Arsha Nagrani, Gül Varol, Weidi Xie, and Andrew Zisserman. 2023. AutoAD: Movie Description in Context. *CoRR* abs/2303.16899 (2023). https://doi.org/10.48550/arXiv.2303.16899 arXiv:2303.16899

[15] Nouzha Harrati, Imed Bouchrika, Abdelkamel Tari, and Ammar Ladjailia. 2016. Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. *Comput. Hum. Behav.* 61 (2016), 463–471. https://doi.org/10.1016/j.chb.2016.03.051

[16] Liwei He, Elizabeth Sanocki, Anoop Gupta, and Jonathan Grudin. 2000. Comparing presentation summaries: slides vs. reading vs. listening. In *Proceedings of the CHI 2000 Conference on Human factors in computing systems, The Hague, The Netherlands, April 1-6, 2000*, Thea Turner and Gerd Szwillus (Eds.). ACM, 177–184. https://doi.org/10.1145/332040.332427

[17] Vladislav Ilin. 2022. The role of user preferences in engagement with online learning. *E-Learning and Digital Media* 19, 2 (2022), 189–208.

[18] Hyeungshik Jung, Hijung Valentina Shin, and Juho Kim. 2018. DynamicSlide: Exploring the Design Space of Reference-based Interaction Techniques for Slide-based Lecture Videos. In *Proceedings of the 2018 Workshop on Multimedia for Accessible Human Computer Interface* (Seoul, Republic of Korea) *(MAHCI'18)*. Association for Computing Machinery, New York, NY, USA, 33–41. https://doi.org/10.1145/3264856.3264861

[19] Junhan Kong, Dena Sabha, Jeffrey P. Bigham, Amy Pavel, and Anhong Guo. 2021. TutorialLens: Authoring Interactive Augmented Reality Tutorials Through Narration and Demonstration. In *SUI '21: Symposium on Spatial User Interaction, Virtual Event, USA, November 9-10, 2021*, Francisco R. Ortega, Robert J. Teather, Gerd Bruder, Thammathip Piumsomboon, Benjamin Weyers, Anil Ufuk Batmaz, Kyle Johnsen, and Christoph W. Borst (Eds.). ACM, 16:1–16:11. https://doi.org/10.1145/3485279.3485289

[20] Svetlana N. Kurbakova, Zlata N. Volkova, and Alexander V. Kurbakov. 2021. 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*. ACM, 124–130. https://doi.org/10.1145/3450148.3450180

[21] Yi Liu. 2021. 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*. ACM, 1490–1493. https://doi.org/10.1145/3456887.3459706

[22] Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. 2021. What Do We Mean by "Accessibility Research"?: A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi, Pernille Bjørn, and Steven Mark Drucker (Eds.). ACM, 371:1–371:18. https://doi.org/10.1145/3411764.3445412

[23] Ignatius Adrian Mastan, Dana Indra Sensuse, Ryan Randy Suryono, and Kautsarina Kautsarina. 2022. Evaluation of distance learning system (e-learning): a systematic literature review. *Jurnal Teknoinfo* 16, 1 (2022), 132–137.

[24] Gareshma N, Vasuki P, and Preethi S. 2022. 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)*. 1678–1682. https://doi.org/10.1109/ICSCDS53736.2022.9760784

[25] Rosiana Natalie, Ebrima Jarjue, Hernisa Kacorri, and Kotaro Hara. 2020. ViScene: A Collaborative Authoring Tool for Scene Descriptions in Videos. In *ASSETS '20: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, Greece, October 26-28, 2020*, Tiago João Guerreiro, Hugo Nicolau, and Karyn Moffatt (Eds.). ACM, 87:1–87:4. https://doi.org/10.1145/3373625.3418030

[26] Andrew Ng. 2017. Introduction to Machine Learning | ML-005 Lecture 1 | Stanford University | Andrew Ng. https://youtu.be/6QRpDLj8huE YouTube video.

[27] Cuong Nguyen and Feng Liu. 2016. Gaze-based Notetaking for Learning from Lecture Videos. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016). https://api.semanticscholar.org/CorpusID:6225105

[28] Naoto Nishida, Hinako Nozaki, and Buntarou Shizuki. 2022. Laugh at Your Own Pace: Basic Performance Evaluation of Language Learning Assistance by Adjustment of Video Playback Speeds Based on Laughter Detection. *Proceedings of the Ninth ACM Conference on Learning @ Scale* (2022). https://api.semanticscholar.org/CorpusID:249205272

[29] OpenAI. 2022. OpenAI: Introducing ChatGPT. https://openai.com/blog/chatgpt (accessed on July 3, 2023).

[30] Debajyoti Pal and Vajirasak Vanijja. 2020. 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 (2020), 105535.

[31] Ray Pastore. 2012. The effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. *Comput. Educ.* 58 (2012), 641–651. https://api.semanticscholar.org/CorpusID:39162270

[32] Amy Pavel, Gabriel Reyes, and Jeffrey P. Bigham. 2020. Rescribe: Authoring and Automatically Editing Audio Descriptions. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 747–759. https://doi.org/10.1145/3379337.3415864

[33] Yi-Hao Peng, Jeffrey P. Bigham, and Amy Pavel. 2021. Slidecho: Flexible Non-Visual Exploration of Presentation Videos. In *ASSETS '21: The 23rd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, USA, October 18-22, 2021*, Jonathan Lazar, Jinjuan Heidi Feng, and Faustina Hwang (Eds.). ACM, 24:1–24:12. https://doi.org/10.1145/3441852.3471234

[34] Yi-Hao Peng, Peggy Chi, Anjuli Kannan, Meredith Ringel Morris, and Irfan Essa. 2023. Slide Gestalt: Automatic Structure Extraction in Slide Decks for Non-Visual Access. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI 2023, Hamburg, Germany, April 23-28, 2023*, Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, Anicia Peters, Stefanie Mueller, Julie R. Williamson, and Max L. Wilson (Eds.). ACM, 829:1–829:14. https://doi.org/10.1145/3544548.3580921

[35] Yi-Hao Peng, JiWoong Jang, Jeffrey P. Bigham, and Amy Pavel. 2021. Say It All: Feedback for Improving Non-Visual Presentation Accessibility. In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi, Pernille Bjørn, and Steven Mark Drucker (Eds.). ACM, 276:1–276:12. https://doi.org/10.1145/3411764.3445572

[36] G Ravichandran, Dr J Sujathamalini, and Dr K Gunasekaran. 2022. E-Learning-Accessibility of Students with Visual Impairment in Higher Education. *International Journal of Research and Review* 9, 5 (2022), 27–31.

[37] Shanna Russ and Foad Hamidi. 2021. Online learning accessibility during the COVID-19 pandemic. In *W4A '21: 18th Web for All Conference, Virtual Event / Ljubljana, Slovenia, April 19-20, 2021*, Silvia Rodríguez Vázquez, Ted Drake, Dragan Ahmetovic, and Victoria Yaneva (Eds.). ACM, 8:1–8:7. https://doi.org/10.1145/3430263.3452445

[38] Andreas Sackl, Franziska Graf, Raimund Schatz, and Manfred Tscheligi. 2020. Ensuring Accessibility: Individual Video Playback Enhancements for Low Vision Users. In *ASSETS '20: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, Virtual Event, Greece, October 26-28, 2020*, Tiago João Guerreiro, Hugo Nicolau, and Karyn Moffatt (Eds.). ACM, 67:1–67:4. https://doi.org/10.1145/3373625.3417997

[39] Clinton Smith. 2020. Challenges and opportunities for teaching students with disabilities during the COVID-19 pandemic. *International Journal of Multidisciplinary Perspectives in Higher Education* 5, 1 (2020), 167–173.

[40] Bo Song, Yan Sun, Junjie Guo, Dan Zhao, and Jiayin Tan. 2021. 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*. ACM, 407–411. https://doi.org/10.1145/3448748.3448988

[41] Aaron Springer and Steve Whittaker. 2019. Progressive disclosure: empirically motivated approaches to designing effective transparency. *Proceedings of the 24th International Conference on Intelligent User Interfaces* (2019). https://api.semanticscholar.org/CorpusID:67769608

[42] Totok Sukardiyono, Muhammad Irfan Luthfi, and Nisa Dwi Septiyanti. 2023. Breaking Down Computer Networking Instructional Videos: Automatic Summarization with Video Attributes and Language Models. *Elinvo (Electronics, Informatics, and Vocational Education)* 8, 1 (2023), 26–37.

[43] Jerry Chih-Yuan Sun and Kai-Yu Chang. 2016. Design and development of a location-based mobile learning system to facilitate English learning. *Univers. Access Inf. Soc.* 15, 3 (2016), 345–357. https://doi.org/10.1007/s10209-014-0392-x

[44] Khalaf Mohammed Tahat, Walaa Al-Sarayrah, Said A Salloum, Mohammed Habes, and Sana Ali. 2021. The influence of YouTube videos on the learning experience of disabled people during the COVID-19 outbreak. In *Advances in Data Science and Intelligent Data Communication Technologies for COVID-19: Innovative Solutions against COVID-19*. Springer, 239–252.

[45] Yujia Wang, Wei Liang, Haikun Huang, Yongqi Zhang, Dingzeyu Li, and Lap-Fai Yu. 2021. Toward Automatic Audio Description Generation for Accessible Videos. In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi, Pernille Bjørn, and Steven Mark Drucker (Eds.). ACM, 277:1–277:12. https://doi.org/10.1145/3411764.3445347

[46] Chengpei Xu, Wenjing Jia, Ruomei Wang, Xiangjian He, Baoquan Zhao, and Yuanfang Zhang. 2023. Semantic Navigation of PowerPoint-Based Lecture Video for AutoNote Generation. *IEEE Trans. Learn. Technol.* 16, 1 (2023), 1–17. https://doi.org/10.1109/TLT.2022.3216535

[47] Chengpei Xu, Ruomei Wang, Shujin Lin, Xiaonan Luo, Baoquan Zhao, Lijie Shao, and Mengqiu Hu. 2019. Lecture2Note: Automatic Generation of Lecture Notes from Slide-Based Educational Videos. In *IEEE International Conference on Multimedia and Expo, ICME 2019, Shanghai, China, July 8-12, 2019*. IEEE, 898–903. https://doi.org/10.1109/ICME.2019.00159

[48] Beste F. Yuksel, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Soo Jung Kim, Joshua Junhee Lee, Seung Jung Jin, Yue-Ting Siu, Joshua A. Miele, and Ilmi Yoon. 2020. Human-in-the-Loop Machine Learning to Increase Video Accessibility for Visually Impaired and Blind Users. In *DIS '20: Designing Interactive Systems Conference 2020, Eindhoven, The Netherlands, July 6-10, 2020*, Ron Wakkary, Kristina Andersen, Will Odom, Audrey Desjardins, and Marianne Graves Petersen (Eds.). ACM, 47–60. https://doi.org/10.1145/3357236.3395433

[49] Baoquan Zhao, Songhua Xu, Shujin Lin, Ruomei Wang, and Xiaonan Luo. 2019. A New Visual Interface for Searching and Navigating Slide-Based Lecture Videos. In *IEEE International Conference on Multimedia and Expo, ICME 2019, Shanghai, China, July 8-12, 2019*. IEEE, 928–933. https://doi.org/10.1109/ICME.2019.00164

A FORMATIVE STUDY: *DETAILED RESULTS*

Table 3: Overview of disabilities and conditions. Multiple choice, with 52 responses. Some (10) acknowledged having two or more conditions. *Total* Mentions, describes how many mentioned this item. *Exclusive* Mentions, describes how many mentioned this item exclusively.

Condition	Total	Exclusive
Corrected vision (refractive errors)	12	6
Blind or low vision (BLV)	11	8
Color blindness	1	1
Deaf or hard of hearing (DHH)	11	8
Cognitive disability	3	2
Dyslexia	2	0
ADHD	11	5
Physical disability	7	5
Non-disabled	7	7

Table 4: Types of video contents participants watch. *Others:* Software support, TED talks, DIY, History, Dance, Music. Items with * indicate free-form user input items. Question is optional and multiple choice, with 50 total replies.

Topic	Count
Specialized subjects	38
Cooking	24
Math	18
Art	15
Programming	14
Languages*	2
Travel & culture*	2
Others*	1

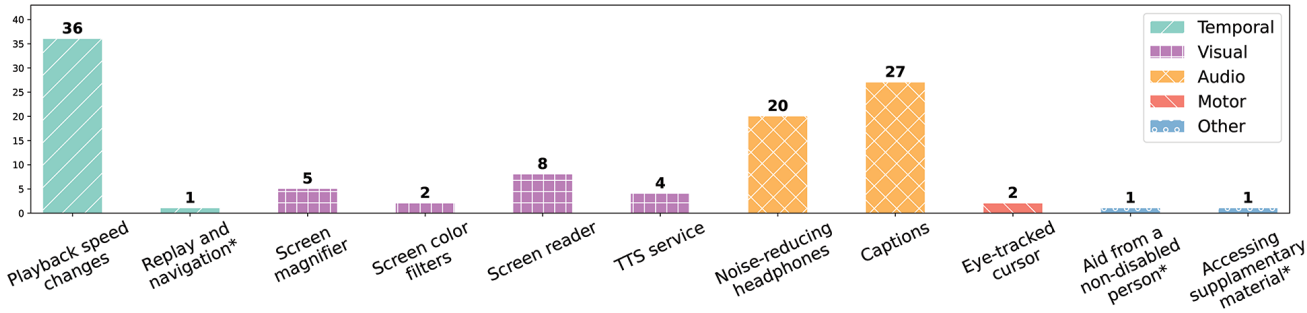

Figure 8: Accessibility tools being used by participants. Two or more tools are used by 34 (64%) participants. Items with * indicate free-form user input items. Question is optional and multiple choice, with 52 total replies.

Table 5: Difficulties participants face, grouped by category. T=Temporal, V=Visual, A=Audio. The right side shows for all non-disabled participants (P7, P9, P42, P44, P49, P50, P51) whether they experience the relevant difficulty. Items with * indicate free-form user input items. Question is optional and multiple choice, with 50 total replies.

Cat.	Difficulty	Total	P7	P9	P42	P44	P49	P50	P51
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 some colors	7							
	Unsure where to place attention	7							
	Missing elements out of center	5							
	Unreliable audio descriptions*	2							
	Overwhelming on-screen details*	1							
A.	Hard to hear the video clearly	24				✓	✓	✓	✓
	Unreliable or missing captions	20	✓	✓	✓		✓		

B DETAILED ALGORITHMS

B.1 WPS-based Speed Modifications

For a given transcription line T_n:

$$T_n^{wps} = \frac{|T_n^{words}|}{T_n^{end} - T_n^{start}} \tag{1}$$

Where T_n^{end} and T_n^{start} are end and start time respectively. Thus, the video playback speed V_{speed} would be:

$$V_{speed} = \frac{U^{wps}}{T_n^{wps}} \tag{2}$$

B.2 OCR-highlight

For the transcription [26] *"[...] try to predict the breast cancer [...]"* and the OCR fragment *"Breast cancer (malignant, benign)"*:

$$W_n^{T,fr} = \texttt{["breast", "cancer"]} \tag{3}$$

We then:

(1) Search for fr such that $|W_n^{T,fr}|$ is maximized.
(2) Ignore if OCR fragment has been selected for frame $n-1$.

B.3 Frame-activity detection

Using visual analysis we locate contours of changes between the first and last frame of a window of frames.

Each contour i of frame n (i.e., $Ct_{n,i}$) is then converted into the simplified definition:

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

Where $A_{n,i}$ is the area it covers and $P_{n,i}$ is the center point. Using this, we employ connected-components (DFS) to group found contours into clusters, based on center-to-center distance. We start with a fully connected graph of all contours as nodes. And iteratively remove edges if the distance between the center points is larger than a given threshold $th_{RoI} = 0.1 \cdot \min(F_w, F_h)$, where F_w, F_h are the width and height of the frame respectively. The remaining connected components are the regions of interest (RoI). Each such

RoI is defined as:

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

Where $N_{n,j}$ is the number of contours it includes, $P_{n,j}$ is the center point of the contours, and $A_{n,j}$ is the area of that region such that $A_{n,j} = \sum_p A_{n,p}$, where p is a contour in the region.

During video playback, we collect all relevant frame activity RoI based on the current playback time. Over these we perform several actions:

(1) Skip highlight if $\#(R_n) > th_R$ where $th_R = 8$.
(2) Sort RoI based on area ($A_{n,j}$).
(3) Ignore RoI if $A_{n,j} > 0.25 \cdot \max(F_w, F_h)$ or $A_{n,j} < 0.01 \cdot \min(F_w, F_h)$.
(4) Check if one of the following is justified, against prior *highlight*:
   (a) Enough time passed ($\Delta t \geq 10$ sec).
   (b) Distance between centers is large ($P_{n,j} - P_{m,k} > 0.1 \cdot \min(F_w, F_h)$).

B.4 Semantic segmentation using LLMs

The semantic segmentation algorithm first calculates for each shot S_i and transcription line T_j, if $T_j \in S_i$.

There are five distinct cases:

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

C1 is fully contained in the shot, therefore belongs to it. C2 and C3 describes transcriptions which are fully outside the shot boundaries. C4 and C5 describe in-between shot transcriptions, which overlap on more than one shot. For cases C4 and C5 we introduce $th_{ratio} = 0.5$ as ratio threshold. Verifying on C4:

$$abs(\frac{T_j^{start} - S_i^{start}}{T_j^{end} - T_j^{start}}) < th_{ratio} \tag{6}$$

And on C5:

$$abs(\frac{T_j^{end} - S_i^{end}}{T_j^{end} - T_j^{start}}) < th_{ratio} \tag{7}$$

C EVALUATION STUDY METHOD

C.1 Participants

Participants have various backgrounds. However, in this study, we did not probe participants' ability status. Future research should explore this factor and include specific groups of target users. There is no participant overlap with the formative study. See Table 6 for detailed information.

Table 6: Background information for each participant.

PID	Age	Gender	English Level	Topic Familiarity
P01	29	Male	Fluent	1
P02	27	Male	Intermediate	3
P03	28	Male	Fluent	4
P04	27	Female	Advanced	2
P05	25	Male	Intermediate	3
P06	24	Male	Advanced	1
P07	24	Male	Beginner	2
P08	25	Male	Intermediate	4
P09	23	Male	Intermediate	3
P10	23	Male	Advanced	3
P11	22	Male	Beginner	2
P12	23	Male	Fluent	2
P13	31	Female	Advanced	3
P14	23	Male	Fluent	3

There are 14 participants, aged 22 to 31. With 12 male and 2 female. Most participants are active university students (13), while 1 (P13) is a post graduate. Participants are from an international background, with 7 from China, 4 from Japan, and 1 from Italy, South Korea and Israel respectively.

C.2 Baseline

The baseline used in this study is a standard video player, with a seek bar, play/pause button, caption on/off button, multiplier-based video speed change button, and shot detection-based chapter markers. The baseline was designed to be as similar as possible to the proposal, with the only difference being the lack of the proposal's features.

C.3 Materials

The video content as well as quiz questions were outsourced to an instructor with several years of teaching experience both face-to-face and online. The instructor was asked to create the videos similar to content in available online resources. The videos were recorded using a screen recording software, with the instructor's voice over, utilizing a pen to draw on the slides. The instructor was compensated with a $50 gift card for their time and effort. Figure 9 shows screenshots of the video content used in the study.

C.4 Procedure

We conducted the study remotely via Zoom, using English, and required participants to use a computer with a headset and silence their mobile devices. Initially, they completed a demographic and

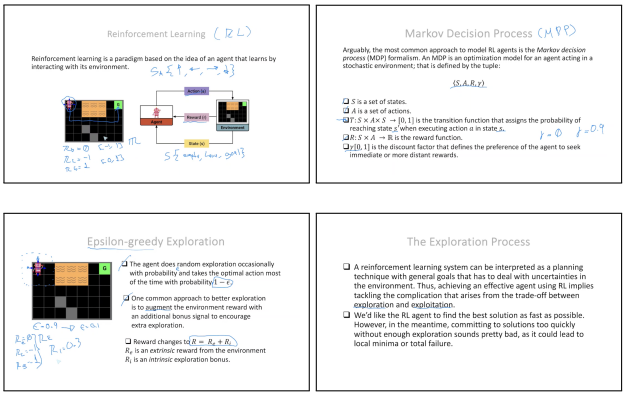

Figure 9: Screenshots of the video content used in the study.

background questionnaire. The study consisted of three parts, averaging one hour to complete.

C.4.1 Training (5 minutes). Participants watched a 3-minute video on Machine Learning and Probability to acquaint themselves with the SmartLearn interface. Following the video, they took a 2-minute, 2-question quiz. This phase used the SmartLearn system exclusively.

C.4.2 Main Tasks (40 minutes). Two instructional videos—one on Reinforcement Learning and the Markov Decision Model, and the other on Reinforcement Learning Policy—were presented, each concluding with a 5-question quiz designed to assess content comprehension. The quizzes included questions of varying difficulty and were randomized in both order and answer options. The tasks alternated between the baseline system and SmartLearn, with the sequence randomized and counterbalanced among participants.

C.4.3 Post-Study Questionnaire and Interview (15 minutes). A subjective evaluation questionnaire measured mental demand, ease of use, and usefulness for both systems. Future preference for usage was also asked. The accompanying semi-structured interview recorded participant experiences and perceptions, with the recordings later transcribed for analysis.

C.5 Analysis

Both quantitative and qualitative analysis were conducted. The quantitative analysis included the interaction logs of users with both the proposal and the baseline, and quiz scores. As for the qualitative analysis, it included the subjective evaluation and semi-structured interview. Statistical test used the Welch's T-test single tail *p-value*. When significance is presented the first value is the proposal as the second is the baseline.

C.5.1 Interaction Logs. Interaction logs were gathered with a custom logging system detailing timestamps of each interaction, corresponding video timestamps, interaction types, and any existing values.

C.5.2 Subjective Evaluation. The subjective evaluation included 3 metrics for each task. The mental demand, ease of use, and usefulness. The values range from 1 to 5, with 1 being the lowest and 5 being the highest.

- Mental Demand: How mentally demanding was the task?
- Ease of Use: How easy was it to use the system?
- Usefulness: How useful was the system to you?

C.5.3 Semi-structured Interview. The interview was anonymized and transcribed. The interview was analyzed to identify common themes and patterns. The analysis was done manually, by reading the transcripts and identifying common themes and patterns. The analysis was done by the author.

D EVALUATION RESULTS

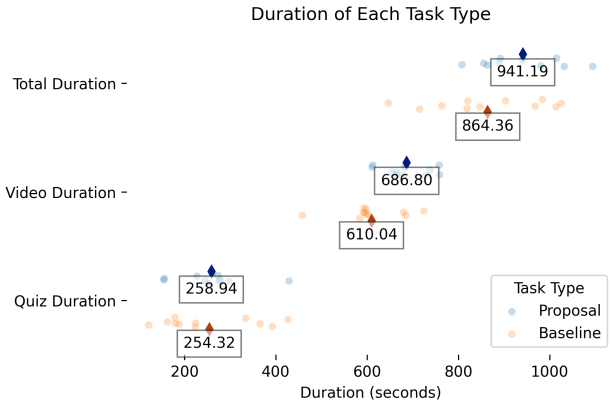

Figure 10: Average time spent on each task (video, quiz and overall) for the baseline and SmartLearn. The value represents the mean time in seconds.

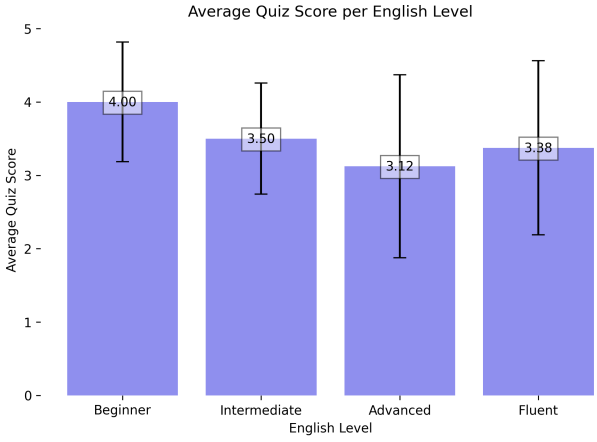

Figure 11: Average quiz score by English proficiency with standard deviation. No significant correlation was found between quiz scores and English proficiency.