Exploring the Usage of Thermal Imaging for Understanding Video Lecture Designs and Students' Experiences

Namrata Srivastava
The University of Melbourne
Melbourne, VIC
srivastavan@student.unimelb.edu.au

Sadia Nawaz
The University of Melbourne
Melbourne, VIC
nawazs@student.unimelb.edu.au

Jason M. Lodge
The University of Queensland
Brisbane, QLD
jason.lodge@uq.edu.au

Eduardo Velloso
The University of Melbourne
Melbourne, VIC
eduardo.velloso@unimelb.edu.au

Sarah Erfani
The University of Melbourne
Melbourne, VIC
sarah.erfani@unimelb.edu.au

James Bailey
The University of Melbourne
Melbourne, VIC
baileyj@unimelb.edu.au

ABSTRACT

Video is becoming a dominant medium for the delivery of educational material. Despite the widespread use of video for learning, there is still a lack of understanding about how best to help people learn in this medium. This study demonstrates the use of thermal camera as compared to traditional self-reported methods for assessing learners' cognitive load while watching video lectures of different styles. We evaluated our approach in a study with 78 university students viewing two variants of short video lectures on two different topics. To incorporate subjective measures, the students reported on mental effort, interest, prior knowledge, confidence, and challenge. Moreover, through a physical slider device, the students could continuously report on their perceived level of difficulty. Lastly, we used thermal sensor as an additional indicator of students' level of difficulty and associated cognitive load. This was achieved through, continuous real-time monitoring of students by using a thermal imaging camera. This study aims to address the following: firstly, to analyze if video styles differ in terms of the associated cognitive load. Secondly, to assess the effects of cognitive load on learning outcomes; could an increase in the cognitive load be associated with poorer learning outcomes? Third, to see if there is a match between students' perceived difficulty levels and a biological indicator. The results suggest that thermal imaging could be an effective tool to assess learners' cognitive load, and an increased cognitive load could lead to poorer performance. Moreover, in terms of the lecture styles, the animated video lectures appear to be a better tool than the text-only lectures (in the content areas tested here). The results of this study may guide future works on effective video designs, especially those that consider the cognitive load.

CCS CONCEPTS

• **Applied computing** → **E-learning**; *Interactive learning environments.*

KEYWORDS

Video lectures, Thermal Imaging, Cognitive load, Instructional design

1 INTRODUCTION

Digital learning environments are becoming pervasive especially in the higher education sector as they offer scalability, flexibility and convenience. Video lectures contribute as the main information delivery tool in these environments [26]. They are also often employed as an important teaching method in traditional, flipped and blended classrooms [25, 46]. In a recent study, students' perception of instructional strategies that use video lectures, was 'overwhelmingly positive'. Students in fully online and blended courses reported that they felt 'more connected' to instructors through the instructor-made videos [41]. Media based content delivery can positively impact students' actual and perceived learning [18], content delivered via this medium is also believed to promote deeper thought processes, communication and interaction among learners [14].

Although there is ample evidence that these multimedia tools are effective for students' learning, there are no set guidelines or standard as to how the videos be created [15]. Therefore, the challenge remains for the researchers to identify the conditions under which different media are most effective for learning [30]. While, recent works on multimedia design have found that lecture production decisions such as length of the lecture, narration and language, speaking rate of the instructor, presence of instructor's image and the instructional design can have a significant impact on students' engagement and learning experiences [12, 23, 32]; several aspects

of instructional videos are not adequately investigated yet. In particular, what are the merits and limitations of different video lecture designs [8]? How different video designs affect learners' experiences and outcomes? Can we analyse through digital, physiological or bio markers when a learner is facing difficulty during a video lecture? A better understanding of these questions could ultimately enable the design of instructional videos that can harness learners' motivation, improve their learning performance and outcomes and can also meet the needs of individual learners.

Existing research suggests that for an effective design of multimedia based videos, learners' cognitive abilities should be catered for. Therefore, a first step to achieve this goal requires an indicator of the cognitive load a student is experiencing. Previous attempts at inferring cognitive load incorporated self-report questionnaires [11, 40]. However, these indicators are subjective and are often collected under the assumption of constant cognitive load across a learning activity. Consequently, researchers are now attempting to measure cognitive load by combining subjective and physiological methods [24]. Particularly continuous real-time measures (multi-channel approaches), such as those collected through psycho-physiological instruments, have the potential to capture the variations in cognitive load throughout the learning session [4]. In this regard, some researchers suggest that skin temperature (ST) and the galvanic skin response (GSR) could be associated to the level of task difficulties [13, 34]. However, a major drawback of using such sensors and measurement techniques is that they often need to be physically attached to the learners, and therefore are potentially invasive.

With recent advances in psycho-physiological research, thermal imaging has shown great potential for estimating the cognitive load people are experiencing by measuring temperature fluctuations of different facial regions [2, 50]. Unlike traditional physiological techniques, thermal imaging is uniquely contact-free, making it a promising tool for ubiquitous computing environments.

However, being contact-free, thermal imaging also poses unique methodological challenges [42]. For example, data collection in previous studies has been carried out under highly controlled conditions where participants were asked to restrict their head movements in order to robustly track their facial temperature from defined facial regions (nose and forehead). This high level of physical restraint has perhaps constrained the adoption of thermal imaging in educational research.

Therefore, in this paper, we present a complete pipeline for automatically extracting the facial temperature of different regions based on computer-vision and image processing techniques. Later, we demonstrate the effectiveness of this tool in estimating the cognitive load of the students, while they watched a video lecture. In particular, through real-time continuous tracking using a thermal camera, we advance the traditional cognitive load measurement techniques by following contributions:

(1) *Methodological*: We demonstrate using information from visible spectrum images (RGB-images) and applying image-processing and computer-vision techniques, that we can reliably and robustly extract facial temperature profiles of the students without imposing any movement restriction on them.

(2) *Analytical* : We demonstrate how measures based on thermal imaging could be related to cognitive load by -
   (a) assessing the impact of different instructional designs on students' cognitive load
   (b) analysing the effects of cognitive load on learning outcomes of the students
   (c) comparing their perceived difficulty with the changes in their facial temperature

We evaluated our approach in an experiment with 78 students who watched two short video lectures on two different topics while continuously being monitored by a thermal camera. Additionally, the students self-reported their difficulty level throughout the lecture using a physical slider. Later, after watching the lecture, they were asked to report on their mental effort, challenge, confidence and interest levels. We aim to see how different video types affect students' perceived difficulty levels and physiology based difficulty levels. We also expect that such fine-grained analysis could reveal if certain parts of the lecture were more difficult for learners (across different video forms of the same content). Lastly, we aim to analyse if there is a match between learners' subjective and physiology based measures of task based difficulties.

2 THEORETICAL FRAMEWORK

Several competing theories have been forwarded to explain difficulties in the learning process in digital environments. Of these theories, 'cognitive load theory' (CLT) and 'cognitive theory of multimedia learning' (CTML) are the most dominant [28]. Researchers have used CLT and CTML as tools for describing the ways through which videos may assist or hinder students' learning. Based on these works, researchers have also suggested some of the key principles and recommendations for an effective video design [5, 29]. We can infer from the CTML theory that to improve students' learning, retention and to enhance their experience audio-video modalities should be combined [29].

Similarly, an implication of CLT is that to improve students' usage of working memory, multimedia should be designed to reduce the cognitive load [5]. Based on the CLT theory, cognitive load can be classified into three categories: intrinsic, extraneous and germane [44]. Intrinsic load pertains to the learning content; complex learning material may lead to an increased intrinsic load. Extraneous load is inherent to the learning activities and their design. Lastly, the germane load relates to the mental effort associated with the processing of new information and then its integration into existing schema. Each type of cognitive load competes for the limited resources of the working memory; and learning can only occur when the total load of the three stays within the capacity of the working memory [22]. An instructional designer should take the different types of cognitive load into account. They should try to find a right balance between keeping the learners engaged and challenged but still within their cognitive capacities [11] which also aligns with Vygotsky's zone of proximal development [49].

Previous studies have explored cognitive load in a variety of ways. Some authors have worked on the optimization of cognitive load and have suggested that the extraneous cognitive load should be reduced [9, 36, 47]. These works highlighted some of the non-cognitive factors (such as emotion, fatigue, vigilance) that can

affect learners' cognitive resources [6]. Learners' mental effort is then defined by taking together the cognitive and non-cognitive factors [17]. Physiology-based measures are currently being investigated for monitoring learners' cognitive load [45]. Examples include electroencephalography (EEG) [19], heart rate [38], electrodermal activity (EDA) [24] Although these techniques could provide biological indicators (less susceptible to user manipulation), these sensors need to be in direct contact with the learners, making them impractical for most environments [39].

Thermal imaging, however, is a strong candidate for inferring cognitive load due to its non-invasive physiological approach [2, 10, 50]. Being unobtrusive and small in size, thermal imaging cameras add sensing capabilities to the environment rather than burdening the user. Moreover, compared to other contact-less approaches, they are more robust, since temperature signature is more resistant to conscious manipulation than heart rate or EDA [16]. By measuring temperature fluctuations of different facial regions, these devices have shown great potential in inferring the physiological and cognitive states of the users [7]. For example, emotions like stress, fear, anxiety, startle responses and attention [1, 16, 42] could be detected by monitoring facial temperature changes. The reason why this is possible is that human skin temperature is modulated by ANS (autonomic nervous system) activities, and cognitive and affective states strongly influence how the blood flows through the body. For example, when we feel embarrassed, blood flows to our face, making our cheeks red. Because blood carries heat, it changes the temperature distribution of the skin, underlying tissues, and vessels.

However, apart from detecting emotional and affective states, there are few studies which have utilized this tool for measuring the mental workload of participants [2]. These studies found that there is a direct relation between workload and the facial temperature. Due to the involvement of ANS, increased brain activity correlates with a surge in blood supply. Tasks that require higher workload result in blood flowing from the adjacent facial areas to the brain leading to a detectable variation in facial temperature. Therefore, monitoring changes in different facial regions can provide insight into the changes in cognitive load. We particularly build upon previous work that inferred cognitive load based on the temperature variations in the forehead and nose regions [2]. We focused on these two facial regions, as they can be monitored even if the user is wearing glasses.

Another constraint, in the existing studies on mental effort is that cognitive load was induced artificially using an experimental manipulation [2]. The changes in facial temperature were measured while subjects completed tasks that were high in cognitive demand (such as a Stoop Test). Therefore, the feasibility of thermal imaging in real-educational scenarios (where cognitive load is not constant and may arise in episodes) still remains an open question. We aim to explore the usefulness of thermal imaging in educational domain. Further, there is little advice on how to analyze such continuous data in order to obtain actionable insights regarding the instructional and production designs of videos.

Therefore in this paper, through real time continuous tracking, we aim to answer the following research questions:

RQ1 : (a) Using thermal imaging, can we assess the impact of different instructional designs on the mental effort expended by students? (b) Can we analyse, through a thermal camera how (the learning processes or) the video content experience differ between the high and low achieving students?

RQ2 : Using an interactive sliding scale and a thermal camera can we analyse learners' perceived and biological responses to task difficulty and what is the relationship between these indicators?

3 STUDY SETUP AND PROCEDURE

Our analysis is based on the dataset collected in our prior study [43]. In our original study, we did not analyse the thermal data, restricting our scope to the analysis of the slider data. In this paper, we extend our previous work by also analysing the thermal imaging dataset. In this section, we briefly describe the dataset and its collection procedure.

3.1 Participants

Participants were 100 university students, aged between 18 and 42 (($Mean$ = 25.04, SD = 4.66). There were 52 females and 48 male. Based on their education level, 38 were undergraduate, 30 were masters and the remaining 32 were PhD students. Based on students' background and the learning material, 21 had previous knowledge about neuroscience only, 35 had previous knowledge about binary numbers only, and 22 had previous knowledge of both the topics. The experiment required 1 hour per participant and it was conducted in a lab-setting.All the participants were rewarded with a gift card for their time and contribution.

3.2 Video Lectures

In selecting our video material, we drew from the literature on educational psychology by reusing four video lectures designed to evaluate instructional design techniques [27, 48]. For our study, we selected two lecture topics - neuroscience (explaining the basic working of neurons in the human brain) and binary numbers (focusing on binary-number conversion, addition and 2's complement). We deliberately chose these two topics to have a diverse range of contents to counter the prior knowledge effect. Additionally, we selected the lectures, where these topics where presented in two different conditions -

(1) **Text-based** : The lectures in this condition employed voiceover production style,i.e. synchronized recording of slideware with instructor's voice. The neuroscience lecture contained slides with only text, no illustration, whereas binary lecture contained slides with equations.

(2) **Animation-based** : The lectures in this condition were presented using graphics and illustrations. For example, the neuroscience lecture contained the following: slides, annotated images and instructor's voice-over. On the other hand, the binary lecture used a digital ink style, where the lecture contained the same slides as the text version, but with additional examples recorded with digital pen. Moreover, before each example, there was also a Socratic dialogue between the instructor and a novice.

To summarise, all the participants were presented with two variants of short video lectures on the topics of neuroscience and binary numbers. The neuroscience lecture consisted of two different slides used to explain the same topic - one using animated video and the other through bullet point slides with no pictures. Whereas in the binary video lecture, the same slides were used in both the versions, however they differed in terms of their presentation styles (see Figure 1). All lectures were manually transcribed and annotated using Anvil[1].

3.3 Measurement instruments

3.3.1 Subjective measurements.
We collected two types of subjective measures in our experiment – survey-based and slider-based.

Survey-based : The participants were asked to fill out a feedback questionnaire at the end of the video. These questions have been designed to reflect standard student evaluation instruments typically used in universities. The questionnaire consisted of following items:

- **Prior knowledge**: We asked the participants "have they ever studied or been exposed to [neuroscience principles or binary numbers] before?"
- **Mental-effort**: We asked the participants to rate "how much mental effort [they invested] while watching the lecture" on a 10-point bipolar scale ranging from "Not much" (1) to "Very"(10).
- **Challenge, Confidence and Interest**: For each item, we asked the participants to rate the lecture in terms of 'challenge', 'confidence' and 'interest' on a 10-point bipolar scale.
- **Lecture-style**: Participants were also asked to rate the lecture according to following five factors on a Likert agreement scale ranging from "Strongly disagree"(1) to "Neutral (3)" to "Strongly agree"(5):
  (1) **Organization**: "The video was well-organized"
  (2) **Clarity** : "The video was clear"
  (3) **Engagement**: "The video engaged me in learning"
  (4) **Helped to Learn**: "The teaching in this video helped me learn"
  (5) **Satisfaction**: "Overall, I am satisfied with video"

In addition, before and after watching each video lecture, participants were asked to report their Judgement of Learning (JOLs). For

---
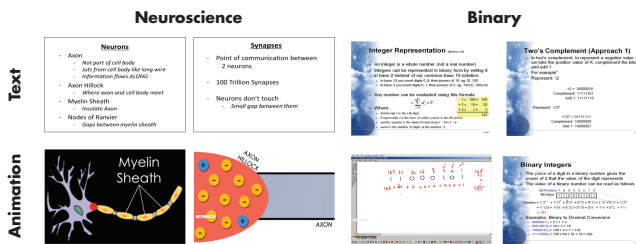[1]http://www.anvil-software.org/

Figure 1: Sample frames from each version of the lectures. Both versions of the neuroscience lecture had the same audio track. The binary lecture combined slides and digital ink, and had different audio tracks.

example, before watching the binary lecture they were asked - "How confident you are now that you can correctly answer questions about the binary conversion, binary addition and two's complement (0 being not confident, 100 being very confident)". The same question was asked again after they watched the lecture. In the educational psychology literature, these ratings have important implications for understanding how people learn and use memory [27, 48].

Slider-based: To capture continuous subjective reactions of participants in real-time, we asked our participants to rate the difficulty throughout the lecture using a physical slider. The slider was one of the faders available in the Numark Mixtrack PRO Midi Controller, which outputs a number between 0 and 127 (see Figure 2).

The survey-based self-reported mental effort and slider-based difficulty served as the "ground truth" for measuring the cognitive load of the students.

Figure 2: Experimental setup used during the study. It consists of a thermal camera, web-camera, eye-tracker and a slider.

3.3.2 Physiological measurements.
We used an Optris PI-400 [2] thermal camera to record participants' facial temperature while they were watching the video lectures. Being equipped with a frame rate of 80 Hz and an optical resolution of 382 x 288 pixels, these small cameras can capture real-time thermographic images in high speed unobtrusively (see Figure 2). These cameras operate with a thermal sensitivity of $0.04°C$ and can measure temperatures between $−20°C$ and $900°C$.

As this study sits within a wider research project about developing adaptive e-learning tools that respond to changes in biometric signals, we also recorded participants' facial expressions with a Logitech webcam, their eye movements with a Tobii Pro X2-30 eye-tracker. However, for the purpose of this paper, our focus is limited to facial temperature analysis.

3.3.3 Pre-test and Post-test scores.
To test the prior knowledge and lecture understanding of the participants, they were asked to complete two separate tests: one before and one after the lecture. The tests were also drawn from the same previous work as the lectures [27]. Each test contained 9 multiple choice questions with 4 answers and one "I don't know" (IDK) option. To counteract correct answers from random guesses, we instructed participants to select IDK if they were not sure which was the right answer. The

---
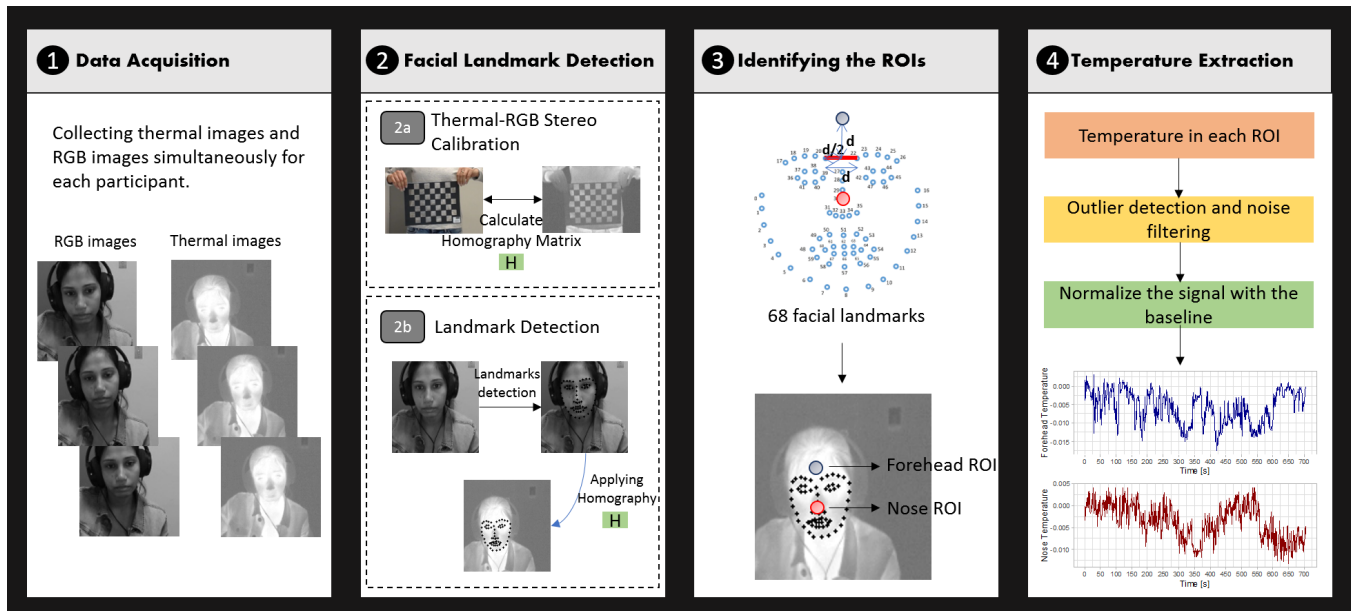[2]https://www.optris.global/thermal-imager-optris-pi400-pi450

Figure 3: Proposed facial temperature extraction pipeline shown for one of the participant (P19). It consists of four steps-(1) The thermal and RGB images are acquired. (2) [2a.] A homography matrix H is calculated using Thermal-RGB stereo calibration. [2b.] Facial landmarks are detected on RGB image, and mapped directly on thermal image by applying the same homography matrix H. (3) Two regions of interest (ROI)–forehead and and nose are defined. (4) Temperature signals are extracted from each ROIs, pre-processed and used for analysis.

tests were marked by awarding one point for each correct answer, while the incorrect, unselected or IDK options received zero points.

3.4 Study Procedure

Upon arrival, participants were asked to read a plain language statement describing the purpose of the study and the procedure, to sign a consent form and to fill in a short demographics questionnaire asking about their age, gender, department and degree of study at the university. We then calibrated the eye tracker using the manufacturer's default 9-point procedure. We also verbally explained the purpose of the experiment and how the participants should operate the slider during the experiment. The experimenter was present in the same room during the course of the whole experiment, but outside the field-of-view of the participant.

The procedure was split into two learning tasks, where each task corresponded to each of the lecture topics (neuroscience and binary). For each learning task, participants first completed the pre-test with 9 multiple-choice questions and rated their confidence-level about the lecture topic (JOLs). Then, participants watched the corresponding video lecture without pausing, rewinding or note-taking. While they watched the lecture, they continuously rated the difficulty level of the lecture using the slider. After the lecture, participants were first asked to fill out the questionnaire asking them about their engagement and opinion of each lecture. Then they answered a post-test questionnaire which assessed their retention and understanding of the material. The same procedure was then repeated for the second topic. Throughout the procedure, we recorded their facial temperature using an unobtrusive thermal camera, and also recorded their eye-movements and facial expressions.

To summarise, all the participants viewed one lecture in one style condition and then viewed the other in the other style condition. For example, if the first lecture was the animated version of the neuroscience lecture, then the second lecture was the text version of the binary lecture. The order of the topics and the videos styles were counter-balanced between participants. The complete session lasted approximately 60 minutes.

4 FACIAL TEMPERATURE EXTRACTION AND ANALYSIS

A block diagram of the employed feature extraction framework is displayed in Figure 3. In each extracted thermal image frames, 68-facial landmarks were detected by using a two-step framework which combined the techniques from image processing and computer-vision (Section 4.2). Next, for each participant, we defined two regions of interest (ROI)–center of the forehead and tip of the nose (Section 4.3). Finally, the mean temperature in each of the ROIs was extracted from each frame, and these two temperature time series, after pre-processing, were used to assess the cognitive load of the participants.

4.1 Data Acquisition

For extracting the thermal facial landmarks, we utilized the information from visible spectrum images (RGB). For our data collection, we built an application in C#, which can collect data from all the sensors simultaneously and in real-time Next, we used the thermal and RGB images of each participant to extract the facial landmarks.

4.2 Facial Landmark Detection

To investigate the effects of lecture difficulty on participants' facial temperature, accurate identification of facial landmarks in thermal images was important. However, detecting these facial regions in thermal images is a challenging task for several reasons. Firstly, the traditional facial landmark detection techniques which are trained on RGB images fail to work for thermal images as these images lack the details present in the visible spectrum. Secondly, most previous research involved acquiring data under highly controlled conditions where participants were asked to restrict their head movements in order to robustly track their facial temperature from defined regions of interest (ROIs). We believe, a computer-assisted learning scenario should not impose such restrictions on the participants. Therefore, we propose to use a two-step feature extraction pipeline, using the techniques from image processing (i.e. stereo calibration based approach), and computer-vision (68-facial landmarks detection).

4.2.1 Thermal-RGB Stereo Calibration. Before each participant's recording, we performed a small calibration step. The step involved capturing pictures from multiple angles within the focus region of the cameras holding a masked checkerboard (a checkerboard whose white colored regions are masked out) on top of the body (as shown in Figure 3). As heat is generated from our body, both the thermal and RGB checkerboard images looked similar in the visible domain. We calculated the corresponding points in both the images by detecting the checkerboard corners using MATLAB's Stereo Camera Calibrator App [3]. The software returned the list of corresponding detected points in both the images.

4.2.2 Landmark Detection. The corresponding points detected in both the images were used to construct a 3x3 homography matrix. Next, we detected the facial landmarks on the RGB images using deep-learning based open source free library (OpenFace [3]) and later applied the homograhy matrix on the detected landmarks, to map the corresponding facial points on the thermal image.

4.3 Identifying the Facial Regions

Temperature changes in different facial areas such as tip of the nose, centre of the forehead have been shown to be linked to changes in mental workload [2]. Therefore, we considered only these two important facial regions for our analysis. The midpoint between the two medial eyebrow landmarks was used to define the forehead region, while the center of the nose landmark was used as the nasal region. Instead of considering a single temperature point, a radius of 3-pixels was considered to define a region. Figure 3 illustrates the location of these ROIs.

4.4 Facial Temperature extraction

For extracting the temperature profiles, we followed a protocol used in previous research [42]. First, the mean temperature across all pixels in two ROIs—forehead and nose was calculated per each thermal frame. There were few frames where participants' faces were not detected or detected with lesser confidence (*Confidence*<95%), temperature values corresponding to the ROIs of previous frames were then used. Next, each time series was corrected for outliers using

mean-value imputation method. To reduce the heteroscedasticity [4] effect, we log-transformed the signals for each participant. Finally, the resulting signal values were subtracted from the mean-baseline to reduce the inter-subject variability. We used the first 10 seconds on the experiment as the baseline condition of each participant.

Landmarks could not be reliably and accurately estimated for 12 participants. Further, we discarded 7 participants thermal data as there forehead was not visible either because of their hairstyle (6) or because of wearing a cap (1). Moreover, we found that 3 participants' landmark estimation was lost at critical points during the lecture, so we discarded them as well. Overall, this study analysis thermal data of 78 participants.

5 RESULTS AND DISCUSSION

Consistent with existing works [20, 33], we adopted an 'iterative' analytics approach where the analysis is presented at different levels of granularity: (a) lecture design level (b) group level and (c) individual level (Figure 4). At the lecture design level, the effects of different video styles on students' mental effort were analysed using a thermal camera. Next, at the group level, a comparison of the high and low achieving student groups is presented in terms of their learning processes and experiences with particular video style. Lastly, we analysed how an individual's subjective experence and physiology based experience varies while watching video lectures. This deeper analysis is conducted on the neuroscience video only, as the participants found its contents to be more difficult.
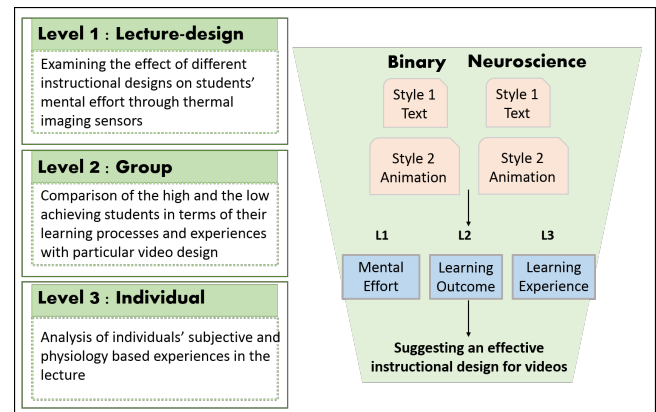

Figure 4: Study Dimensions. As shown, we conducted analysis at three different levels of granularity—lecture-level, group level and individual level.

In the following sections, we have focused on two Region of Interests (ROIs)—forehead (center of the forehead) and nose (tip of the nose). Also, we define the term **"Change in Temperature"** as the difference between the absolute values of the temperature in these ROIs and the mean temperature of the first 10*s* of the lecture. We represent it in terms of percentage.

---

[3]https://au.mathworks.com/help/vision/ug/stereo-camera-calibrator-app.html

[4]Heteroscedasticity refers to the condition in which the variability of a variable is unequal across the range of values of a second variable that predicts it
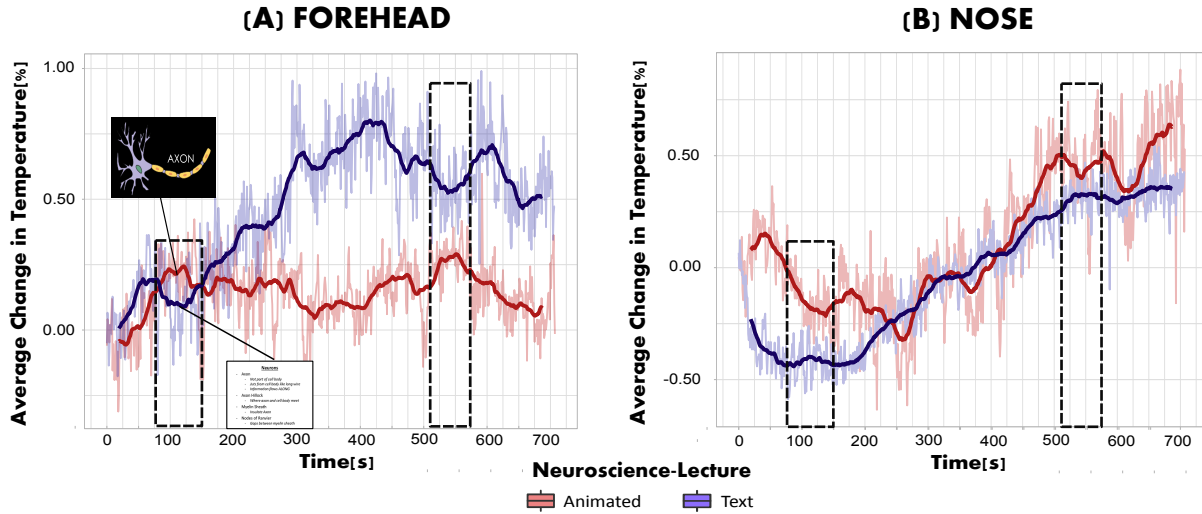
Figure 5: Effect of different instructional designs on (a) forehead and (b) nose temperature. Time is indicated on x-axis, and average change in temperature(%) on y-axis. The highlighted regions refers to the time where the forehead temperature in the text-group showed a decreasing trend, while animated-group had an increasing trend. Shading indicates non-filtered signals.

5.1 Lecture-design level analysis: Effect of instructional design

To begin our analysis, we divided the participants into two subgroups (1) participants who watched the animated version of the lecture ($n = 43$), and (2) participants who viewed the text version of the same lecture ($n = 35$). Next, we examined the changes in facial temperature in the two ROIs—forehead and nose for the two groups. For the neuroscience video, the audio track was the same in both versions, only the visual content differed. This allowed us to perform time-frame by time-frame comparison of the responses to the videos – text version with the animated version. Figure 5 shows the temperature changes in the forehead and nose region averaged across all the participants.

This comparison allowed us to identify interesting points about the lectures. First, we see that the nose temperature for both the groups starts to converge at $t = 246s$. However, before that time, we can observe that the group watching the text-based version of the lecture had lower nose-temperature than the group watching the animation-lecture. Upon inspection of the video material, we found that during the first 4 minutes of the lecture, the participants in the text-based were presented with slides containing the definition for

all the terms such as *Axon, Axon Hillock, Myelin Sheath, Nodes of Ranvier* , whereas the animated version of the video presented one concept at a time. In line with previous research, the decrease in nose temperature here could be an indicator of higher cognitive load [2].

Second, we could observe similar changes in the forehead temperature of the two groups. The participants who watched the text-based lecture showed higher temperature changes than the participants watching the animated version of the lecture. Again, previous research suggests that an increase in forehead temperature here could also indicate higher cognitive load [2]. However, we can also notice there are two regions in the time-series signals, where text-version showed a relative decreasing trend in temperature while the animated-version showed an increasing trend. Through visual inspection of the lecture, we found that at $t = 74s$, the text-based lecture has all the terms present in a single slide, whereas the animated version contained moving annotated images. Similarly, the segment starting at $t = 520s$, which corresponded to the part where ions flowed through synapses, we can observe this trend. After the lecture, participants reported that it was difficult for them to focus on the audio of the lecture with the moving ions. We can

Table 1: Statistical analysis results after comparing neuroscience text-based and animation-based lectures, ***p < 0.001

	Text-based Lecture n = 35		Animation-based Lecture n = 43				
	M	SD	M	SD	T - stat	p value	Effect size
Self-reported Measures [1-10]							
Mental Effort	7.54	1.93	6.77	2.02	-1.73	0.09	-0.39
Challenge	6.77	2.16	6.14	2.28	-1.25	0.21	-0.28
Confidence	5.66	2.47	6.35	2.09	1.32	0.19	0.30
Post score	4.23	2.25	5.00	2.26	1.50	0.14	0.34
Facial Temperature Changes [%]							
Mean Forehead Temperature	0.47	0.26	0.14	0.12	-30.32	0.000***	-1.61
Mean Nose Temperature	-0.03	0.30	0.12	0.32	9.15	0.000***	0.49

also, see a decrease in the nose-temperature for these intervals in the nose-temperature time-series signals, although the differences are quite small. These physiological responses could be indicating that during these segments in the lecture, participants preferred the text version more compared to the animated version.

We also performed a statistical comparison (Welch two sample t-test) on the facial temperature changes. Consistent with the previous work around cognitive load [31] and cognitive theory of multimedia learning [35, 37], we found that, when participants were exposed to text-based lectures, their forehead temperature was significantly higher and nose temperature was significantly lower than the animation-based lectures (Table 1).

Further, to interpret these differences, we evaluated participants' subjective experiences about the two lecture designs. We compared their perceived difficulty (mental-effort), their confidence and challenge levels and their post-test scores in both conditions. We found that participants who viewed the text-based lectures reported higher mental effort, higher challenge, lower confidence and achieved poorer learning outcomes compared to the participants who watched the animation-based lecture, although the differences were not significant (*p-value>0.05, effect size<0.5*) (see Table 1).

5.2 Group level analysis: High-achievers vs Low-achievers

To understand the impact of cognitive load on learning outcomes, we divided the participants into two sub-groups: high-achieving and low-achieving. We used the mean post-test score of all the participants as the cut-off point. Following, we present the analysis for the neuroscience text-based lecture. We chose this lecture, as this was the most-difficult lecture (Mental-Effort: $mean = 7.54, SD = 1.93$) among all, and therefore, could help us understand the effects of difficulty on facial temperature.

Out of 35 participants, 16 participants scored higher marks in the post-test, while the remaining 19 participants scored lower marks. The self-report measures for mental-effort, challenge and confidence for both the groups are presented in Table 2. We found that high-achieving group reported significantly higher confidence (*Cohen's d* = −1.70), than the low-achieving group. This group also reported less mental-effort and challenge levels, however, the difference was not significant.

Comparing these changes with the facial temperature changes, we found that the high-achieving group showed less variation in their forehead temperature ($M = −0.14\%, SD = 0.08\%$). The participants in this group seem to have a constant temperature relative
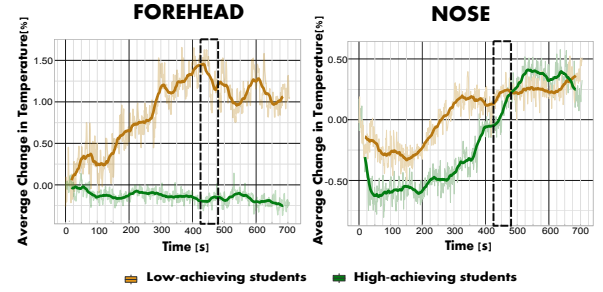

Figure 6: Forehead and Nose temperature of high-achieving and low-achieving groups. Time is indicated on x-axis, and average change in temperature(%) on y-axis. The highlighted regions refers to the time where the forehead temperature for the low-achieving group showed a decreasing trend. Shading indicates non-filtered signals.

to the participants in the low-achieving group ($M = 0.81\%, SD = 0.43\%$). We found that the forehead temperature of the low-achieving group increased until $t = 425s$ and then started to decrease. Temperature further increased only at $t = 475s$. A visual inspection of the video lecture then revealed that around these two time stamps the lecture-slide contents and audio were not well-synchronised. The words spoken by the instructor were not present on the slides. This mismatch in audio and video channels could have resulted in higher difficulty levels for the participants – as apparent by an increase in their nose temperature around the same time stamp (Figure 6).

5.3 Individual level analysis: Comparison between perceived difficulty and changes in facial temperature

As mentioned, in this study the participants could report on the perceived difficulty of the video lecture, in real-time using a physical slider. The usage of thermal camera then allowed us to continuously monitor the physiological changes related to difficulty. We compared how individuals' perception about the lecture differs from their physiological measurements. It was noticed that some participants forgot to move the slider during certain sections of the lecture, which made it difficult to present a statistical viewpoint for the above analysis. Therefore, we present here a case study for one participant watching the text version of the neuroscience lecture - P58 (Figure 7). The participant aged 22 is a male

Table 2: Statistical analysis results after comparing high-achieving students and low-achieving students groups, ***p < 0.001

	High-Achieving Group n = 16		Low-Achieving Group n = 19				
	M	SD	M	SD	T - stat	p value	Effect size
Self-reported Measures [1-10]							
Mental Effort	7.50	2.10	7.58	1.83	-0.12	0.91	0.04
Challenge	6.19	2.29	7.26	1.97	-1.48	0.15	0.50
Confidence	7.38	1.54	4.21	2.17	5.02	0.000***	-1.70
Facial Temperature Changes [%]							
Mean Forehead Temperature	-0.14	0.08	0.88	0.44	-61.32	0.000***	3.26
Mean Nose Temperature	-0.15	0.40	0.05	0.23	-11.89	0.000***	0.63

undergraduate Commerce student, with no knowledge about the neuroscience topic. His subjective measures for confidence, challenge and mental-effort and interest are 2, 9, 9 and 2 respectively. Of the total achievable score of 9, the participant's pre-test score was zero, and post-test score was 4. These measures indicate he is low in confidence, less interested in the lecture but invested high mental-effort while watching the lecture. Although he found the lecture challenging, he showed an improvement of 44.4% in the post-test.
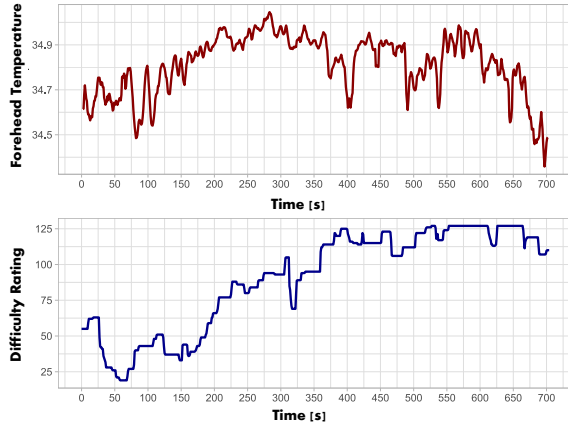

Figure 7: Changes in forehead temperature and perceived difficulty values for participant (P58) throughout the lecture.

When comparing his absolute slider ratings with the absolute values of forehead temperature, we found that the perceived difficulty of the lecture correlated with the changes in participant's forehead region (*Pearson − correlation* = 0.23). Although there will be some latency between the slider-movement and forehead temperature change, the focus of this analysis is to understand the trend, not the absolute values.

We found that, consistent with increasing difficulty ratings for the slider, the forehead temperature increased. However, at one particular instance, we can start to observe a decrease in participant's forehead temperature. We believe this might be an indication of increasing task difficulty. We can refer to this point as the *"point of deviation"*. We noticed that slider-ratings and temperature values positively correlated before this point(*cross-correlation* = 0.71 ± 0.017), and negatively correlated after this point(*cross-correlation* = −0.18 ± 0.042), considering a maximum of 10s lag. This probably suggests a clear shift in their attitude to the learning task, i.e. they are getting frustrated or feel like giving up. [21]

To summarise, when the related works indicated a higher cognitive load by an increase in the forehead temperature and a decrease in the nose temperature, their hypothesis was based on experiments where participants completed high cognitive-demand tasks (e.g. Stroop tests). In our analysis, we tested the same hypothesis but in a more naturalistic task. As the task was not deliberately designed to elicit high cognitive load, we didn't observe a similar trend throughout the experiment. However, during the difficult parts of the lecture (as determined by the average slider-based difficulty ratings [43]), our results matched with the existing studies.

5.4 Limitations & Future Work

This study did not incorporate participants' interaction patterns since in the lab environment they were not permitted to rewind, pause or replay the videos. The authors understand that such 'trace data' could provide meaningful interactions especially those related to learning behaviours and processes. However, the main focus of this study was the usage of physiological sensors towards understanding students' mental effort. This trace data could be incorporated into future study designs.

Another possible future direction could be to explore *why* we observed certain peaks and troughs in the facial temperature signals. For example, a decrease in forehead temperature or an increase in nose temperature (lower cognitive load) could indicate two conflicting factors: the participants might have found the lecture content easy to understand, or they might have disengaged from the task.

We believe video-stimulated recall or physiological responses from other sensors (e.g. eye-trackers) could help us address these questions and complement our quantitative findings.

6 CONCLUSIONS

In this paper, we demonstrated that thermal imaging is a promising tool to assess learners' cognitive load. This work has two major contributions: first, we provided a complete pipeline for facial temperature extraction. Second, this work provides advice on how to analyse continuous data for obtaining insights about video lecture designs. We believe that our work could guide future studies towards reliable and efficient extraction of facial temperature, which thus far had limited adoption in educational research.

While the findings of this study do not depart from previous research in educational technology, our results reaffirm the previous findings on learning from multimedia and add additional support in the form of psycho-physiological data. Furthermore, we have demonstrated that these principles can be seen in simulated real world learning environments without the use of invasive instruments.

REFERENCES

[1] Yomna Abdelrahman, Anam Ahmad Khan, Joshua Newn, Eduardo Velloso, Sherine Ashraf Safwat, James Bailey, Andreas Bulling, Frank Vetere, and Albrecht Schmidt. 2019. Classifying Attention Types with Thermal Imaging and Eye Tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 69 (Sept. 2019), 27 pages. https://doi.org/10.1145/3351227

[2] Yomna Abdelrahman, Eduardo Velloso, Tilman Dingler, Albrecht Schmidt, and Frank Vetere. 2017. Cognitive Heat: Exploring the Usage of Thermal Imaging to Unobtrusively Estimate Cognitive Load. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 33 (Sept. 2017), 20 pages. https://doi.org/10.1145/3130898

[3] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018).* 59–66. https://doi.org/10.1109/FG.2018.00019

[4] Monique Boekaerts. 2017. Cognitive load and self-regulation: Attempts to build a bridge. *Learning and Instruction* 51 (2017), 90 – 97. https://doi.org/10.1016/j.learninstruc.2017.07.001

[5] Paul Chandler and John Sweller. 1991. Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction* 8, 4 (1991), 293–332. https://doi.org/10.1207/s1532690xci0804_2

[6] Lucile Chanquoy, André Tricot, and John Sweller. 2007. *Cognitive load: Theory and applications.* Armand Colin.

[7] Youngjun Cho and Nadia Bianchi-Berthouze. 2019. Physiological and Affective Computing through Thermal Imaging: A Survey. *arXiv preprint arXiv:1908.10307* (2019).

[8] Konstantinos Chorianopoulos and Michail N. Giannakos. 2013. Usability Design for Video Lectures. In *Proceedings of the 11th European Conference on Interactive TV and Video (EuroITV '13).* ACM, New York, NY, USA, 163–164. https://doi.org/10.1145/2465958.2465982

[9] Michelle Patrick Cook. 2006. Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education* 90, 6 (2006), 1073–1091. https://doi.org/10.1002/sce.20164

[10] Veronika Engert, Arcangelo Merla, Joshua A. Grant, Daniela Cardone, Anita Tusche, and Tania Singer. 2014. Exploring the Use of Thermal Infrared Imaging in Human Stress Research. *PLOS ONE* 9, 3 (03 2014), 1–11. https://doi.org/10.1371/journal.pone.0090782

[11] Inga Glogger-Frey, Katharina Gaus, and Alexander Renkl. 2017. Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? *Learning and Instruction* 51 (2017), 26 – 35. https://doi.org/10.1016/j.learninstruc.2016.11.002

[12] Philip J. Guo, Juho Kim, and Rob Rubin. 2014. How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S '14).* ACM, New York, NY, USA, 41–50. https://doi.org/10.1145/2556325.2566239

[13] Eija Haapalainen, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey. 2010. Psychophysiological Measures for Assessing Cognitive Load. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp '10).* ACM, New York, NY, USA, 301–310. https://doi.org/10.1145/1864349.1864395

[14] Karim HAJHASHEMI, Neil ANDERSON, Cliff JACKSON, and Nerina CALTABIANO. 2015. ONLINE LEARNING: CAN VIDEOS ENHANCE LEARNING? *The Eurasia Proceedings of Educational & Social Sciences* 2 (2015), 238–241.

[15] Christina Ilioudi, Michail N Giannakos, and Konstantinos Chorianopoulos. 2013. Investigating differences among the commonly used video lecture styles. (2013).

[16] Stephanos Ioannou, Vittorio Gallese, and Arcangelo Merla. 2014. Thermal infrared imaging in psychophysiology: Potentialities and limits. *Psychophysiology* 51, 10 (2014), 951–963. https://doi.org/10.1111/psyp.12243

[17] Steve M Jex. 1998. *Stress and job performance: Theory, research, and implications for managerial practice.* Sage Publications Ltd.

[18] Robin H. Kay. 2012. Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior* 28, 3 (2012), 820 – 831. https://doi.org/10.1016/j.chb.2012.01.011

[19] Yufeng Ke, Hongzhi Qi, Feng He, Shuang Liu, Xin Zhao, Peng Zhou, Lixin Zhang, and Dong Ming. 2014. An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task. *Frontiers in Human Neuroscience* 8 (2014), 703. https://doi.org/10.3389/fnhum.2014.00703

[20] Gregor Kennedy and Terry Judd. 2004. Making sense of audit trail data. *Australasian Journal of Educational Technology* 20, 1 (2004). https://doi.org/10.14742/ajet.1365

[21] Gregor Kennedy and J LODGE. 2016. All roads lead to Rome: Tracking students' affect as they overcome misconceptions. ASCILITE. http://hdl.handle.net/11343/121990

[22] Femke Kirschner, Liesbeth Kester, and Gemma Corbalan. 2010. Cognitive load theory and multimedia learning, task characteristics, and learning engagement: The current state of the art. (2010).

[23] René F Kizilcec, Jeremy N Bailenson, and Charles J Gomez. 2015. The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology* 107, 3 (2015), 724.

[24] Charlotte Larmuseau, Pieter Vanneste, Piet Desmet, and Fien Depaepe. 2019. Multichannel Data for Understanding Cognitive Affordances During Complex Problem Solving. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK19).* ACM, New York, NY, USA, 61–70. https://doi.org/10.1145/3303772.3303778

[25] Shih-Yin Lin, John M Aiken, Daniel T Seaton, Scott S Douglas, Edwin F Greco, Brian D Thoms, and Michael F Schatz. 2016. Exploring University Students' Engagement with Online Video Lectures in a Blended Introductory Mechanics Course. *arXiv preprint arXiv:1603.03348* (2016).

[26] Tharindu Rekha Liyanagunawardena, Andrew Alexandar Adams, and Shirley Ann Williams. 2013. MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distributed Learning* 14, 3 (2013), 202–227.

[27] Jason Lodge, Jared Cooney Horvath, Alex Horton, Gregor Kennedy, Sven Venema, and Shane Dawson. 2017. Designing videos for learning: Separating the good from the bad and the ugly. (01 2017).

[28] Jason M. Lodge, Gregor Kennedy, Lori Lockyer, Amael Arguel, and Mariya Pachman. 2018. Understanding Difficulties and Resulting Confusion in Learning: An Integrative Review. *Frontiers in Education* 3, 49. https://doi.org/10.3389/feduc.2018.00049

[29] Richard E. Mayer. 2002. Multimedia learning. , 85 - 139 pages. https://doi.org/10.1016/S0079-7421(02)80005-6

[30] Richard E. Mayer and Roxana Moreno. 2002. Aids to computer-based multimedia learning. *Learning and Instruction* 12, 1 (2002), 107 – 119. https://doi.org/10.1016/S0959-4752(01)00018-4

[31] Richard E Mayer and Roxana Moreno. 2003. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist* 38, 1 (2003), 43–52.

[32] Roxana Moreno and Richard E Mayer. 2000. Engaging students in active learning: The case for personalized multimedia messages. *Journal of educational psychology* 92, 4 (2000), 724. https://doi.org/10.1037/0022-0663.92.4.724

[33] Sadia Nawaz, Gregor Kennedy, James Bailey, Chris Mead, and Lev Horodyskyj. 2018. Struggle town? Developing profiles of student confusion in simulation-based learning environments. *Open Oceans: Learning Without Borders* (2018), 224.

[34] Nargess Nourbakhsh, Yang Wang, Fang Chen, and Rafael A. Calvo. 2012. Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference (OzCHI '12).* ACM, New York, NY, USA, 420–423. https://doi.org/10.1145/2414536.2414602

[35] Fred Paas, Juhani E Tuovinen, Huib Tabbers, and Pascal WM Van Gerven. 2003. Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist* 38, 1 (2003), 63–71.

[36] Fred G Paas. 1992. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology* 84, 4 (1992), 429. https://doi.org/10.1037/0022-0663.84.4.429

[37] Fred GWC Paas and Jeroen JG Van Merriënboer. 1993. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors* 35, 4 (1993), 737–743.

[38] Fred GWC Paas and Jeroen JG Van Merriënboer. 1994. Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of educational psychology* 86, 1 (1994), 122. https://doi.org/10.1037/0022-0663.86.1.122

[39] Colin Puri, Leslie Olson, Ioannis Pavlidis, James Levine, and Justin Starren. 2005. StressCam: Non-contact Measurement of Users' Emotional States Through Thermal Imaging. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05).* ACM, New York, NY, USA, 1725–1728. https://doi.org/10.1145/1056808.1057007

[40] Steven F. Raaijmakers, Martine Baars, Lydia Schaap, Fred Paas, and Tamara van Gog. 2017. Effects of performance feedback valence on perceptions of invested mental effort. *Learning and Instruction* 51 (2017), 36 – 46. https://doi.org/10.1016/j.learninstruc.2016.12.002

[41] Katherine Kensinger Rose. 2009. Student perceptions of the use of instructor-made videos in online and face-to-face classes. *MERLOT Journal of Online Learning and Teaching* 5, 3 (2009), 487–495.

[42] Saurabh Sonkusare, David Ahmedt-Aristizabal, Matthew J Aburn, Vinh Thai Nguyen, Tianji Pang, Sascha Frydman, Simon Denman, Clinton Fookes, Michael Breakspear, and Christine C Guo. 2019. Detecting changes in facial temperature induced by a sudden auditory stimulus based on deep learning-assisted face tracking. *Scientific reports* 9, 1 (2019), 4729.

[43] Namrata Srivastava, Eduardo Velloso, Jason M. Lodge, Sarah Erfani, and James Bailey. 2019. Continuous Evaluation of Video Lectures from Real-Time Difficulty Self-Report. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* ACM, New York, NY, USA, Article 586, 12 pages. https://doi.org/10.1145/3290605.3300816

[44] John Sweller. 1999. Instructional design. In *Australian educational review.* Citeseer.

[45] John Sweller. 2011. Cognitive load theory. In *Psychology of learning and motivation.* Vol. 55. Elsevier, 37–76.

[46] Peter Tiernan. 2015. An inquiry into the current and future uses of digital video in University teaching. *Education and Information Technologies* 20, 1 (2015), 75–90.

[47] Jeroen J G Van Merriÿnboer and John Sweller. 2010. Cognitive load theory in health professional education: design principles and strategies. *Medical Education* 44, 1 (2010), 85–93. https://doi.org/10.1111/j.1365-2923.2009.03498.x

[48] Sven Venema and Jason Lodge. 2013. Capturing dynamic presentation: Using technology to enhance the chalk and the talk. *Australasian Journal of Educational Technology* 29, 1 (2013). https://doi.org/10.14742/ajet.62

[49] Lev Vygotsky. 1978. Interaction between learning and development. *Readings on the development of children* 23, 3 (1978), 34–41.

[50] Qiushi Zhou, Joshua Newn, Namrata Srivastava, Tilman Dingler, Jorge Goncalves, and Eduardo Velloso. 2019. Cognitive Aid: Task Assistance Based On Mental Workload Estimation. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19).* ACM, New York, NY, USA, Article LBW2315, 6 pages. https://doi.org/10.1145/3290607.3313010

Author/s:
Srivastava, N;Nawaz, S;Lodge, JM;Velloso, E;Erfani, S;Bailey, J

Title:
Exploring the usage of thermal imaging for understanding video lecture designs and students' experiences

Date:
2020-03-23

Citation:
Srivastava, N., Nawaz, S., Lodge, J. M., Velloso, E., Erfani, S. & Bailey, J. (2020). Exploring the usage of thermal imaging for understanding video lecture designs and students' experiences. LAK '20: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, pp.250-259. ACM. https://doi.org/10.1145/3375462.3375514.

Persistent Link:
http://hdl.handle.net/11343/241529