Accepted to *Information Systems Research*
ISR-2023-0718.R3

Online Appendix to "Attention Trap? How Visual Motion Shapes Children's Screen Time and Engagement on Video Platforms"

Sumeet Kumar (corresponding author), Madhu Viswanathan
Indian School of Business, Gachibowli, Hyderabad 500032, Telangana, India, sumeet_kumar@isb.edu, madhu_viswanathan@isb.edu

Ravi Bapna
Leavey School of Business, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053, USA, rbapna@scu.edu

This online appendix provides supporting information for the three studies reported in the main paper: feature-extraction procedures, full regression tables and the foreground/background motion decomposition for Study 1 (A1); viewership trends and the video-speed extension of the Study 2 field experiment (A2); lab setup, variable definitions and the re-engagement analyses for Study 3 (A3); the children-versus-grownups ad experiment (A4); and the list of YouTube channels in the Study 1 dataset (A5).

Online Appendices (A)

A1. Study 1: Supporting Information

We provide additional details on Study 1, including video features, summary statistics, and extended regressions.

A1.1. Definition and Process of Extracting Features from Video Data

Here we provide additional details about the video features used in the study and the process of extracting these features.

A1.1.1. Visual Features: We extract visual features directly from the videos in our dataset. A key methodological decision involves determining the portion of the video from which features should be derived. While it is possible to process entire videos, doing so is both time-intensive and computationally costly. Consistent with commonly cited platform guidance that a view is credited after sustained playback (often described as 30 seconds[1]), we focus on the first 30 seconds.

[1] https://www.tubics.com/blog/what-counts-as-a-view-on-youtube

From each video, we extract one frame per second for the first 30 seconds, resulting in 30 frames per video. This is accomplished using the OpenCV library, a widely used toolkit for computer vision tasks. We then compute a set of visual features based on these frames and aggregate them to the video level. All visual features are therefore based on a consistent sampling procedure from the initial segment of each video.

Optical Flow: Our main variable of interest is optical flow, a commonly used metric in computer vision applications for estimating the movement of brightness patterns across video frames (Gibson 1950). Optical flow provides a theoretically grounded and computationally tractable measure of how visual elements change over time. This makes it particularly well-suited for analyzing motion-based stimulation in video content targeted at children. The framework that we use to measure optical flow is summarized below:

Considering two consecutive frames in a video that are taken at t, and $t+\delta t$, the optical flow is calculated using the difference in location of a pixel in the first frame $I(x,y,t)$ to the pixel in the second frames at $I(x+\delta x, y+\delta y, t+\delta t)$.

$$I(x,y,t) = I(x+\delta x, y+\delta y, t+\delta t) \tag{1}$$

We can take the partial time derivative of $I(x,y,t)$, and represent them as I_x, I_y, and I_t, which are the derivatives with respect to x coordinate, y coordinate representing spatial intensity gradients, and t (temporal intensity gradient). Assuming that the intensity of a pixel represents a part of an object that doesn't change much in a short time, we have:

$$dI(x,y)/dt = 0 \tag{2}$$

Taking partial derivatives, we get:

$$\frac{\partial I}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial I}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial I}{\partial t} = 0 \tag{3}$$

This leads to invariance constraint for optical flow, which is represented as:

$$I_x u + I_y v + I_t = 0 \tag{4}$$

where, $u = \frac{\partial x}{\partial t}$ and $v = \frac{\partial y}{\partial t}$ are the optical flow components in x and y directions respectively. We can use this Equation 4 to measure the optical flow. However, the equation has two unknowns and cannot be solved without additional constraints. There are two popular approaches for addressing this and estimating optical flow. We discuss these next.

Lucas–Kanade Optical Flow: We estimate Lucas–Kanade (LK) optical flow using OpenCV's pyramidal implementation (Lucas and Kanade 1981). In practice, we focus on the first 30 seconds of each video and uniformly sample frames at 1 FPS, yielding a short sequence of frames $\{F_0, \ldots, F_T\}$ with $T \approx 30$. We convert sampled frames to grayscale and first detect a sparse set of salient feature points (Shi–Tomasi corners; up to 50 points) in an initial usable frame. LK then tracks these points between consecutive sampled frames, producing a displacement vector for each successfully tracked point. For each adjacent frame pair, we summarize motion using the sum of displacement magnitudes. If tracking quality degrades (e.g., too few points remain), we re-detect feature points in the current frame and continue tracking. Because LK motion measures are right-skewed, we aggregate them across the sampled window at the video level and apply a log transform (implemented as $\log(1+x)$ to accommodate zeros), yielding our LK motion measures (e.g., `LK_Motion (log)`).

The Lucas-Kanade technique for optical flow estimation works well for minor motion fields, typical of recorded videos. Yet it falters when faced with extensive displacements, leading to imprecise estimates within uniform areas. To address this, employing a multi-scale (or pyramid) strategy or an iterative method can be advantageous in gauging flow across varied scales. Hence, alongside Lucas-Kanade, we also incorporate a widely-recognized method suggested by Farnebäck (2003), and report Farnebäck-based estimates as our primary motion measure in the main text, and treat Lucas–Kanade as a robustness check.

Farnebäck Optical Flow: Farnebäck optical flow estimates a dense motion field, producing a flow vector for each pixel in the image (Farnebäck 2003). We use OpenCV's implementation for each pair of consecutive sampled frames (1 FPS over the first 30 seconds)[2]. For each frame pair, Farnebäck

[2] https://docs.opencv.org/3.4/de/d9e/classcv_1_1FarnebackOpticalFlow.html

returns horizontal and vertical flow components at every pixel, which we convert to a per-pixel magnitude map. We then summarize frame-pair motion using the sum of magnitudes across all pixels. These frame-level summaries are aggregated to the video level across the sampled window, and we apply a log transform (implemented as $\log(1+x)$) to address skewness, yielding our primary motion measures (e.g., `FB_Motion (log)`).

Foreground–Background Decomposition: In addition to aggregate motion, we further decompose visual motion into foreground and background components to distinguish motion associated with focal actors from motion in visually secondary regions. For each sampled frame, we identify foreground regions using person bounding boxes detected by a pretrained YOLO object detector, with all pixels inside detected person boxes (expanded by a small padding) classified as foreground and all remaining pixels classified as background. Using the same optical-flow estimates described above, we compute motion magnitude and total motion separately within foreground and background regions at the frame level, and then aggregate these measures to the video level using the same averaging and log-transformation procedures. Full implementation details are provided in Appendix A1.3.

Count of Pixels: Pixel count serves as a proxy for video resolution quality. Higher pixel counts typically correspond to sharper and more visually appealing content, which may influence viewer perception. We compute pixel count using the maximum resolution available for each video retrieved from YouTube. The total number of pixels per frame is obtained through OpenCV and scaled by dividing by one million to yield the feature 'Pixel_m'.

Scene Cuts: We measure visual transitions across scenes using the PySceneDetect library[3], applying its 'AdaptiveDetector' algorithm. This detector dynamically adjusts sensitivity based on content characteristics and identifies abrupt visual changes by analyzing shifts in pixel intensities and color distributions. For each video, we compute both the absolute number of scene changes, calling it 'Scene_cuts'. Like for other visual features, we only use the first 30 seconds of videos to estimate the number of scene cuts.

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

Hue: Hue represents the dominant wavelength in a color spectrum and defines the fundamental color family perceived by viewers, such as red, green, or blue. We calculate hue by extracting the hue channel from each frame in the HSV (Hue, Saturation, Value) color space, as implemented in OpenCV. For each video, we compute the mean hue value across frames to generate a representative estimate of the predominant color tone.

Saturation: Saturation captures the intensity or purity of color. Higher saturation corresponds to more vivid, vibrant hues, while lower saturation results in more muted, grayish tones. We extract saturation values from the HSV color space for each frame and compute the average across the sampled frames. This measure reflects the overall visual vividness of a video.

Brightness: Brightness, or value, refers to the lightness of an image, independent of hue or saturation. It ranges from black to white and influences overall visual energy. We compute brightness using two complementary techniques: first, by averaging the value channel in the HSV space, and second, by calculating the mean pixel intensity after converting each frame to grayscale. These two estimates are closely aligned and provide robust measures of overall visual luminance.

Color Contrast: We assess color contrast through the variation in saturation across video frames. Specifically, we compute the coefficient of variation, defined as the standard deviation of saturation divided by the mean saturation. This metric captures how much visual intensity fluctuates within a video. High contrast values indicate frequent transitions between vivid and muted regions, while low values suggest uniform coloration.

Variety of objects: To account for variation in the visual content of videos, specifically, the types of objects depicted, which may influence viewership, we incorporate object-level analysis into our modeling framework. Recent advances in multimodal AI models, including those capable of extracting semantic content from images (e.g., Gemini, GPT-4o), offer a scalable approach to identify key visual entities. However, due to the computational and financial constraints associated with analyzing thirty thousand full-length videos, we adopt a more pragmatic approach grounded in prior literature.

Previous research uses video thumbnails as representative proxies for overall video content (Choi et al. 2025). Following the approach, we extract and analyze YouTube video thumbnails of all videos

present on the platform. Video thumbnails are single image frames often selected by creators to reflect the theme of the video. To identify objects within each thumbnail, we employ the "GPT4o-mini"[4] vision-language model released by OpenAI, which supports both text and vision in the API.

For each thumbnail, we query the model (using API) with the prompt: "You are an expert at analyzing YouTube thumbnail images. Identify all entities, objects, and visual elements present in the image. Focus on people, characters, animals, cartoons, and other notable elements. Return your analysis as a JSON array of entities/objects without any explanation.". For the first few thumbnails, we did a manual evaluation check, finding that the results conformed to what one would expect by visually analyzing them. From the output, we extract the frequently occurring objects across the dataset and select the top few for inclusion as categorical variables in our regression models. We find these top few related to whether the thumbnail has 'text', 'person', 'cartoon', 'animal', and 'logo'. Consequently, we use these in regressions as binary variables, whether they are present or not, e.g., 'Has_text', implying whether text is present in the thumbnail or not. This approach allows us to efficiently capture salient visual content features at scale while maintaining interpretability.

A1.1.2. General Video Features

Topics: YouTube API allows access to many YouTube video attributes. In particular, two attributes of our interest are the 'tags', and 'categoryId'[5]. Tags are keywords that creators add to their videos to help viewers find them, and are supposed to be a representation of the topics included in the videos[6]. Therefore, we extract video tags from metadata provided by the YouTube API. We filter these tags based on frequency to retain the most representative topical indicators for each video. Some of these tags include 'babies', 'family', 'educational', 'music_songs', etc. We converted these tags into binary variables, e.g., 'Has_cartoons_tag', so that they could be included in the regressions.

[4] https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/

[5] `https://developers.google.com/youtube/v3/docs/videoCategories/list`

[6] https://backlinko.com/hub/youtube/tags

YouTube Category: Each YouTube video is assigned a platform-defined category, which reflects its genre or content type. Categories include music, gaming, animals, travel, and others[7]. These predefined fields provide a coarse-grained classification of videos, and we include them as fixed effects in our statistical models.

A1.1.3. Video Transcript Features

Sentiment: We analyze the emotional tone of each video's spoken content by applying the VADER (Hutto and Gilbert 2014) sentiment analysis tool to the transcript. This tool assigns positive and negative sentiment scores based on lexical and syntactic features. The resulting measures capture the affective valence of verbal communication in the video transcript.

Speech Rate: Speech rate captures the tempo of verbal delivery and is computed as the number of transcript characters divided by the total video duration in seconds. This metric reflects how densely information is packed in time and may influence both cognitive load and viewer retention.

A1.1.4. Audio (Vocal) Features We extract a set of relevant audio features that characterize the vocal and musical qualities of each video. These include tempo, pitch, volume, timbre, harmony, and dynamics, which are commonly studied in research on voice and audio processing (Sivasankar et al. 2005). These features allow us to capture both structural and expressive elements of the audio track, which may influence children's engagement with video content.

Tempo: Tempo refers to the speed of the audio and is measured in beats per minute. Faster tempo often conveys excitement and urgency, while slower tempo may indicate calmness or seriousness. Pitch estimates the fundamental frequency of a sound and is a perceptual property that helps distinguish between high and low tones. This is particularly relevant for character voices and musical elements in children's content.

Timbre: Timbre, also known as tone color, reflects the quality of a sound that allows listeners to differentiate between sources such as voices, instruments, or ambient sounds. It provides a nuanced measure of auditory richness and variation.

[7] https://mixedanalytics.com/blog/list-of-youtube-video-category-ids/

Harmony: Harmony captures the tonal structure and consonance of the audio. It reflects how well different musical notes and chords blend together, contributing to the overall pleasantness of the sound. We derive harmony by calculating chroma features, which represent the distribution of spectral energy across the twelve pitch classes used in Western music. Higher average chroma values indicate greater harmonic richness.

Dynamics: Dynamics measures the variation in loudness over time and captures how audio energy fluctuates throughout a video. These variations can create dramatic contrast and emphasize key moments. We compute dynamics using the root mean square (RMS) energy of the audio signal, which quantifies the average power of the waveform. Higher RMS values indicate more intense and expressive audio content.

Volume: Volume is operationalized as the median decibel level of the audio track, which represents the typical sound intensity experienced by the viewer. This is calculated from the audio spectrogram, which visualizes sound amplitude over time. Using the median reduces the influence of brief spikes or outliers and yields a more stable measure of average loudness. This measure is particularly useful for identifying systematic volume choices that might affect attention.

We use the Librosa library[8], a widely used Python package for audio analysis to extract these features. As we did for video features, each audio track is divided into one-second segments, and feature values are computed for each segment individually. We then calculate the average across segments to obtain a stable estimate of each feature for the full 30-second clip.

A1.1.5. Emotion Features: We retrieve emotions from faces present in the video frames. Our approach comprises a face detection model and an emotion classification model. For face detection, which involves detecting all human faces in an image, we use a popular Python package RetinaFace[9]. Retinaface provides one of the state-of-the-art methods for detecting faces and has been used in many prior works on image analysis (Deng et al. 2020). The output of face detection is the regions

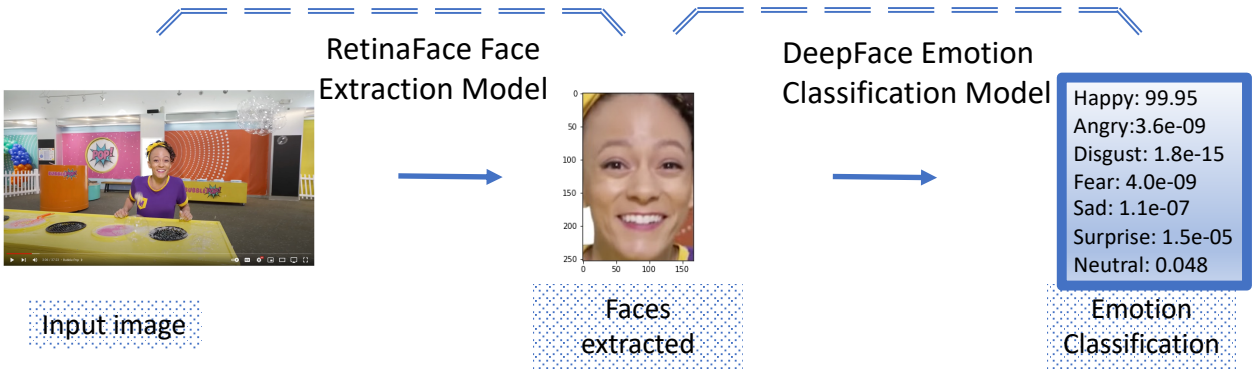
[8] https://librosa.org/doc/main/generated/librosa.piptrack.html

[9] `https://github.com/serengil/retinaface`

of the image that contain faces, which is then passed to another popular Python package named DeepFace[10] to retrieve emotions expressed in these faces (Serengil and Ozpinar 2020). Trained on four million images uploaded by Facebook users, the model employs a deep neural network and achieved an accuracy of 97.35% ± 0.25% on a well-known benchmark dataset where human beings had 97.53% (Taigman et al. 2014). The DeepFace algorithm classifies emotions into seven categories: 1) Happy, 2) Neutral, 3) Surprise, 4) Sad, 5) Angry, 6) Fear, and 7) Disgust.

$$Emotions\_focal\_ratio = \frac{focal\_emotion\_count}{count\_of\_all\_emotions} \tag{5}$$

Figure A1 **Illustration of emotion extraction from a video frame**



We show an example of face detection and emotion classification in Fig. A1. As we can observe, the proposed approach can find faces in the video frame and correctly identify the emotion expressed by the lady in the frame. For estimating facial emotions in a video, we sample frames at 1 FPS over the first 30 seconds and compute dominant emotions expressed in all faces present in these frames. Using the dominant emotions for all frames, we estimate the ratio of each emotion for the video, which is the fraction of the count of the focal emotion, to the sum of the count of all emotions as described in Equation 5.

[10] `https://github.com/serengil/deepface`

A1.2. Dataset Statistics and Effect Size

We report detailed summary statistics for all variables in Table A1 and the full set of regression results in Table A2. To assess practical significance, we interpret the elasticity implied by the log–log specification using the Farnebäck measure. In Column (10) of Table A2, the coefficient on *FB_Motion (log)* is $\beta = 0.106$, implying that a 1% increase in visual motion is associated with a 0.106% increase in video views (and a 10% increase corresponds to approximately a 1.1% increase in views).

Equivalently, a 10% increase in visual motion corresponds to approximately a 1.1% increase in viewership. Expressed differently, achieving a 1% increase in video views requires roughly a 9.4% increase in optical flow intensity. While this elasticity is modest in magnitude, it is economically meaningful in the context of large-scale digital platforms, where small proportional changes in engagement translate into substantial differences in absolute audience reach. These results underscore that relatively small enhancements in visual motion can compound to generate sizable aggregate effects across large populations of viewers.

A1.3. Decomposing Foreground and Background Visual Motion

A key limitation of standard optical-flow-based measures of visual motion is that they capture aggregate pixel-level movement without distinguishing whether motion originates from focal actors (e.g., presenters or characters) or from visually secondary elements (e.g., background). This distinction is theoretically important because foreground motion is often intertwined with semantic content, narrative action, and emotional expression, whereas background motion may not directly alter narration. As a result, conflating these sources risks wrongly attributing engagement effects.

To address this concern, we decompose visual motion into foreground and background components using person localization. Specifically, we identify foreground regions as the union of person bounding boxes detected using a pretrained YOLOv8 object detector (Varghese and Sambath 2024), and classify all remaining pixels as background. We acknowledge that persons are not always the foreground, and in many situations, objects could be a part of the foreground. However, this approach is well-suited to the structure of children's video content, in which human presenters or animated characters

Table A1 Summary Statistics and Description of Video Characteristics

Variable Description	N	Mean	SD	Min	P25	Median	P75	Max
View Count Total number of video views	33072	5605606	63371066	8	26026	226384	1279464	7996500680
Views (log) Natural logarithm of total video views	33072	12.155	2.862	2.079	10.167	12.330	14.062	22.802
Age_in_weeks Age of video in weeks since posting	33072	255.115	135.715	52.000	148.000	238.000	336.000	888.000
LK_Motion_log Log-transformed Lucas-Kanade optical flow measure of visual motion intensity	33023	7.745	0.947	0.082	7.424	7.895	8.309	9.775
FB_Motion_log Log-transformed Farneback optical flow measure of visual motion intensity	32897	15.519	0.926	0.000	15.045	15.654	16.131	20.838
Color_contrast Variation in color saturation throughout the video	33072	0.377	0.291	0.000	0.182	0.310	0.471	6.631
Pixel_m Video resolution in millions of pixels	33072	1.599	1.153	0.034	0.922	2.074	2.074	8.294
Has_text Binary indicator for text present in thumbnail (1=yes)	31655	0.441	0.497	0	0	0	1	1
Has_person Binary indicator for human figures in thumbnail (1=yes)	31655	0.531	0.499	0	0	1	1	1
Has_cartoon Binary indicator for cartoon characters in thumbnail (1=yes)	31655	0.172	0.377	0	0	0	0	1
Has_animal Binary indicator for animals in thumbnail (1=yes)	31655	0.119	0.324	0	0	0	0	1
Has_logo Binary indicator for logos in thumbnail (1=yes)	31655	0.215	0.411	0	0	0	0	1
Hue Average color tone in video (0-180 scale)	33072	61.852	25.170	0.003	44.296	60.466	78.525	176.369
Saturation Average color intensity in video (0-255 scale)	33072	103.593	40.752	0.000	74.837	101.836	131.013	246.983
Brightness Average luminance of video content (0-255 scale)	33072	141.020	42.075	0.416	111.579	141.762	172.523	250.109
Scene_cuts Number of scene transitions in the first 30 seconds	33072	6.566	4.839	0.000	3.000	6.000	9.000	98.000
Tempo Speed of musical beat in beats per minute (BPM)	33030	143.290	11.833	0.000	137.039	143.211	149.797	230.918
Pitch Average frequency of audio content	33030	25.014	13.030	0.000	16.209	22.846	31.449	174.178
Timbre Tonal character of sounds in audio	33030	0.055	0.034	0.000	0.028	0.048	0.075	0.297
Harmony Measure of tonal congruence and chord structure	33030	0.452	0.068	0.000	0.411	0.451	0.496	0.775
Dynamics Variation in audio intensity throughout the video	33030	0.091	0.057	0.000	0.047	0.080	0.125	0.485
Speech_rate Words per minute spoken in the video	31663	5.584	4.594	0.000	0.809	5.299	8.864	23.206
Emotions_happy_ratio Proportion of video showing happy expressions	14507	0.246	0.261	0.000	0.000	0.188	0.385	1.000
Emotions_sad_ratio Proportion of video showing sad expressions	14507	0.266	0.305	0.000	0.000	0.167	0.375	1.000
Emotions_angry_ratio Proportion of video showing angry expressions	14507	0.068	0.146	0.000	0.000	0.000	0.083	1.000
Emotions_surprise_ratio Proportion of video showing surprised expressions	14507	0.050	0.128	0.000	0.000	0.000	0.042	1.000
Emotions_disgust_ratio Proportion of video showing disgusted expressions	14507	0.001	0.016	0.000	0.000	0.000	0.000	1.000
Vader_sentiment_neg Proportion of negative sentiment in transcript	33072	0.035	0.045	0.000	0.000	0.021	0.055	1.000
Vader_sentiment_pos Proportion of positive sentiment in transcript	33072	0.148	0.119	0.000	0.050	0.154	0.212	1.000
Has_babies_tag Binary indicator for baby-related content (1=yes)	33072	0.220	0.414	0	0	0	0	1
Has_toddlers_tag Binary indicator for toddler-related content (1=yes)	33072	0.236	0.425	0	0	0	0	1
Has_kids_tag Binary indicator for kid-related content (1=yes)	33072	0.379	0.485	0	0	0	1	1
Has_family_tag Binary indicator for family-related content (1=yes)	33072	0.129	0.335	0	0	0	0	1
Has_educational_tag Binary indicator for educational content (1=yes)	33072	0.170	0.376	0	0	0	0	1
Has_cartoons_animation_tag Binary indicator for cartoon/animation content (1=yes)	33072	0.285	0.452	0	0	0	1	1
Has_music_songs_tag Binary indicator for music/song content (1=yes)	33072	0.207	0.405	0	0	0	0	1
Has_toys_play_tag Binary indicator for toys/play content (1=yes)	33072	0.192	0.394	0	0	0	0	1

Note: This table provides summary statistics for all variables used in the regression analysis. Binary variables (has_*) indicate the presence of specific features or content tags.

typically serve as the primary semantic carriers, while motion outside these regions is, generally, visually secondary (see Figure A2 for samples). Moreover, bounding-box-based person detection is

Table A2 Effects of Video Characteristics on Video Views

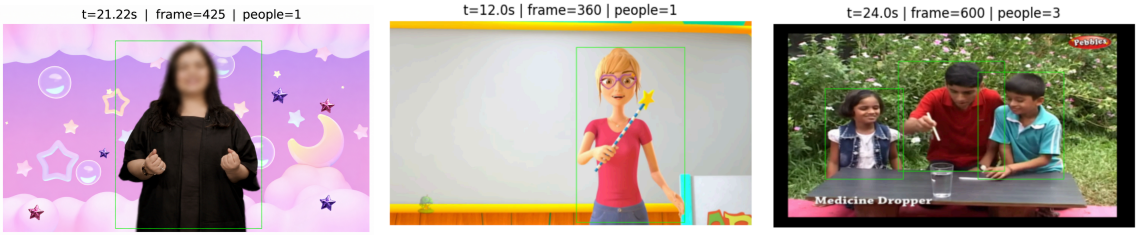
VARIABLES	(1) Views(log)	(2) Views(log)	(3) Views(log)	(4) Views(log)	(5) Views(log)	(6) Views(log)	(7) Views(log)	(8) Views(log)	(9) Views(log)	(10) Views(log)
LK_Motion (log)	0.095***	0.102***	0.128***	0.110***	0.102***					
	(0.024)	(0.025)	(0.037)	(0.025)	(0.031)					
FB_Motion (log)						0.086***	0.093***	0.137***	0.099***	0.106***
						(0.027)	(0.029)	(0.038)	(0.027)	(0.037)
Color_contrast	0.089				0.027	0.129				0.041
	(0.110)				(0.144)	(0.111)				(0.148)
Pixel_m	0.043				0.051	0.028				0.031
	(0.033)				(0.034)	(0.031)				(0.033)
Has_text	-0.124***				-0.068	-0.120***				-0.069
	(0.044)				(0.052)	(0.044)				(0.052)
Has_person	-0.019				-0.002	-0.021				-0.002
	(0.047)				(0.064)	(0.047)				(0.065)
Has_cartoon	-0.066				-0.082	-0.068				-0.081
	(0.052)				(0.064)	(0.053)				(0.065)
Has_animal	0.089				0.110	0.090				0.110
	(0.054)				(0.076)	(0.054)				(0.076)
Has_logo	0.018				0.054	0.013				0.046
	(0.053)				(0.066)	(0.052)				(0.065)
Hue	0.000				0.002*	0.000				0.002*
	(0.001)				(0.001)	(0.001)				(0.001)
Saturation	0.002**				0.003**	0.002**				0.003**
	(0.001)				(0.001)	(0.001)				(0.001)
Brightness	0.000				0.002	0.001				0.002
	(0.001)				(0.002)	(0.001)				(0.002)
Scene_cuts	0.008				0.013	0.009				0.014
	(0.009)				(0.011)	(0.009)				(0.011)
Tempo		0.002			0.002		0.002			0.001
		(0.001)			(0.002)		(0.001)			(0.002)
Pitch		0.001			-0.001		0.001			-0.001
		(0.002)			(0.002)		(0.002)			(0.002)
Volume		-0.006			-0.010		-0.004			-0.009
		(0.018)			(0.026)		(0.018)			(0.027)
Harmony		0.212			0.361		0.174			0.298
		(0.348)			(0.427)		(0.346)			(0.418)
Dynamics		2.082***			1.817***		2.056***			1.812***
		(0.494)			(0.578)		(0.502)			(0.594)
Emotions_happy_ratio			0.043		0.020			0.029		0.011
			(0.078)		(0.075)			(0.078)		(0.075)
Emotions_sad_ratio			-0.041		-0.042			-0.037		-0.040
			(0.067)		(0.061)			(0.066)		(0.061)
Emotions_angry_ratio			-0.244		-0.269*			-0.238		-0.264*
			(0.151)		(0.145)			(0.149)		(0.145)
Emotions_surprise_ratio			-0.165		-0.200*			-0.153		-0.191*
			(0.113)		(0.109)			(0.113)		(0.109)
Emotions_disgust_ratio			-0.999*		-1.087*			-1.013*		-1.108**
			(0.551)		(0.552)			(0.547)		(0.547)
Vader_sentiment_neg				2.437***	2.164**				2.386***	2.147**
				(0.697)	(1.012)				(0.701)	(1.021)
Vader_sentiment_pos				0.191	-0.374				0.169	-0.383
				(0.301)	(0.364)				(0.301)	(0.365)
Has_babies_tag				0.062	0.075				0.064	0.073
				(0.106)	(0.125)				(0.107)	(0.127)
Has_toddlers_tag				0.271	0.192				0.262	0.192
				(0.167)	(0.187)				(0.168)	(0.188)
Has_kids_tag				0.004	0.028				0.012	0.038
				(0.104)	(0.124)				(0.103)	(0.122)
Has_family_tag				-0.200*	-0.034				-0.196*	-0.031
				(0.113)	(0.109)				(0.113)	(0.110)
Has_educational_tag				-0.053	-0.061				-0.052	-0.062
				(0.131)	(0.108)				(0.131)	(0.110)
Has_cartoons_tag				-0.124	-0.187**				-0.100	-0.178*
				(0.128)	(0.091)				(0.129)	(0.090)
Has_music_songs_tag				0.019	-0.105				0.012	-0.110
				(0.128)	(0.207)				(0.126)	(0.204)
Has_toys_play_tag				-0.067	0.045				-0.054	0.056
				(0.169)	(0.219)				(0.168)	(0.219)
Speech_rate				-0.010	-0.014				-0.012	-0.016
				(0.013)	(0.016)				(0.013)	(0.016)
Observations	31,556	31,523	13,790	31,564	13,771	31,435	31,402	13,765	31,443	13,746
R-squared	0.749	0.748	0.798	0.749	0.801	0.749	0.749	0.798	0.750	0.801
Age (in Weeks) FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Channel FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Video Category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors clustered at the channel level in parentheses.
*** p<0.01, ** p<0.05, * p<0.10

computationally efficient and robust to framing variation, multi-person scenes, and heterogeneous production styles, making it appropriate for large-scale YouTube data used in Study 1.

For each video, we sample frames at a fixed rate (1 FPS) and compute optical flow between consecutive sampled frames using the Farnebäck estimator. At the frame level, person bounding boxes are used to construct a binary foreground mask, with pixels inside detected person regions

Figure A2 Examples of YOLOv8-based Person Bounding Boxes (Green Rectangles) for Foreground and Background Classification.



classified as foreground and all remaining pixels classified as background. Optical flow magnitudes are then aggregated separately over foreground and background regions within each frame, yielding frame-level measures of foreground and background motion intensity. These frame-level measures are subsequently aggregated to the video level across sampled frames to construct video-level measures of total motion, foreground motion, and background motion. All motion variables are log-transformed to account for skewness and are included alongside the same set of controls used in the main analysis.

Table 3 reports a sequence of regressions that decompose overall visual motion into background and foreground components using Farnebäck-based optical flow. Columns (1) to (4) present unconditional specifications that include all videos, regardless of whether people appear or move on screen. Foreground motion is inherently sparse, as many videos contain little or no sustained human presence or exhibit minimal bodily movement, attenuating their estimated effects. To address this issue, Columns (5) to (6) report conditional specifications that restrict the sample to videos with sustained human presence, defined as a person presence rate exceeding 0.5. This conditional specification allows for a more meaningful assessment of the role of foreground motion.

Column (1) presents a baseline specification using total visual motion. Total motion is positive and statistically significant (0.106, $p < 0.01$), confirming that aggregate visual dynamics are associated with higher video views. Columns (2) and (3) replace total motion with background motion. Background motion remains positive and highly significant, with coefficients of 0.087 and 0.094, respectively. The similarity of these magnitudes to the total motion coefficient indicates that background motion accounts for most of the predictive content of aggregate visual motion.

Columns (2) and (4) introduce foreground motion in the full, unconditional sample. Foreground motion is positive but not statistically significant when averaged across all videos, with coefficients of 0.030 and 0.039. In Column (2), which includes both background and foreground motion, the background motion coefficient remains stable and statistically significant, while the foreground coefficient remains smaller and statistically insignificant. This pattern indicates that variation in bodily motion does not systematically predict engagement when many videos contain little or no meaningful foreground activity.

Columns (5) and (6) report conditional specifications that restrict attention to videos with sustained human presence. In this subsample, foreground motion becomes positive and statistically significant, with coefficients of 0.095 ($p < 0.05$) and 0.080 ($p < 0.05$). At the same time, background motion remains positive and statistically significant (0.070, $p < 0.05$), though its magnitude is somewhat attenuated relative to the unconditional specifications. The person presence rate itself enters negatively and is not statistically significant, indicating that engagement is not driven by the mere presence of people, but rather by how dynamically they move when present.

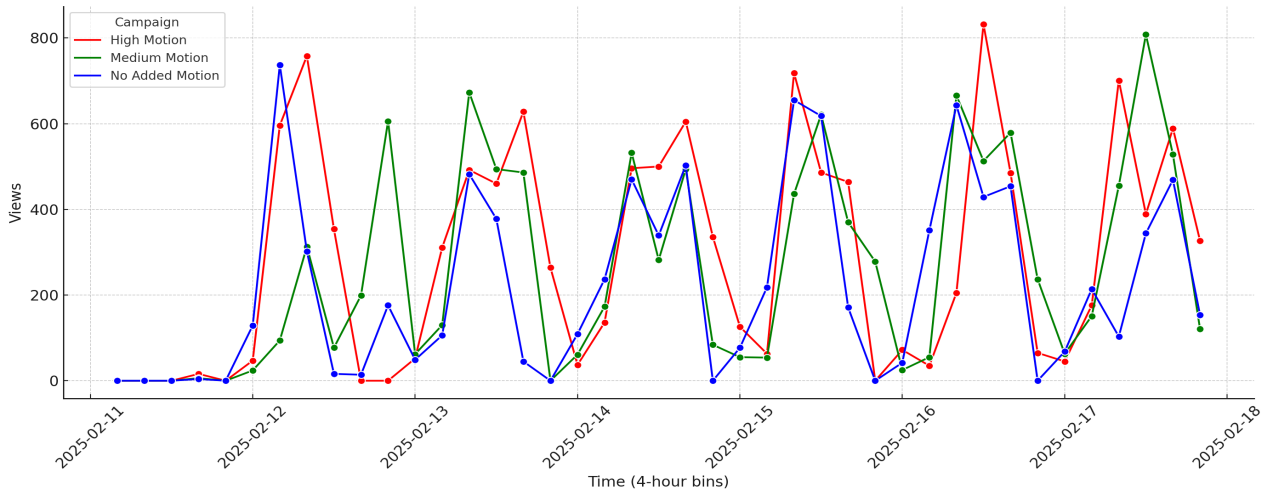
Taken together, the progression of coefficients across Columns (1)–(6) clarifies the distinct roles of background and foreground motion. Background motion is a robust and consistent predictor of video views across all specifications, capturing general scene-level visual dynamism. Foreground motion operates on an intensive margin, contributing to engagement only in videos where people are present for a substantial portion of the content.

A2. Study 2 - YouTube Ads Experiments - Supporting Information and Extension

A2.1. Viewership Trend of Campaign Videos

Figure A3 represents 4-hour interval viewership patterns over the first seven days of the campaign. All three conditions show pronounced diurnal cycles, with engagement peaking during late morning and early evening hours. The High Motion and Medium Motion campaigns consistently outperform the No Added Motion baseline across most intervals, suggesting that increased visual motion enhances viewer retention throughout the day, not just in aggregate daily totals.

Figure A3 Comparing viewership trend for different campaigns



A2.2. Study 2 Extension 1: Changing Optical Flow (Visual Motion) by Changing Video Speed

This study extends the YouTube ads study 2, using another set of videos that had a floating bubble from left to right, at different speeds (for different study arms). Like Study 2, we target the viewers of 'made-for-kids' content, for which children are the primary audience. We needed to ensure that the videos we use in our ads are of interest to the target audience. To identify the right content, we checked prior research to identify that young children commonly watch videos featuring toys, food products, and bubbles (Grotewell and Burton (2008), page 95). Toys have brand associations that can be problematic in this context, as awareness of some brands can affect viewership. Also, food product viewership can be based on local preferences. Therefore, we selected a video containing floating bubbles of different colors with pleasing background music with a Creative Commons license allowing video reuse without permission. The video contains floating bubbles that start from the left of the screen and move to the right, and the process repeats. Because the video is repetitive, even after changing the video's speed (e.g., more frames per second), overall, the same content is shown to the audience.

One can change OF in three different ways; one, by increasing the number of moving objects, two, by changing the size of objects; and three, by changing the video speed. Increasing the number

of moving objects in our context, i.e., increasing the number of bubbles in the videos, would also affect other visual characteristics like color contrast. Similarly, changing the size of bubbles used in the videos would also change the OF; however, that again would change the overall color contrast. Therefore, we find that changing the speed of videos is more appropriate in our setting, where we desire to change the OF without affecting other video characteristics. The selection of a repeating video with floating bubbles allows us to easily manipulate the OF by changing the speed of the videos, keeping most other aspects of the video the same.

We create four videos by changing the video speed {Slow (0.5x), Normal (1x), Fast (1.5x), Super-fast (2x)} that change their OFs. The increase in speed increases the OF {Low:9.71, Regular:18.77, High:29.93, Very-high:37.75 } as there is more visual stimulus per unit of time because of the faster movement of the bubbles. This setting also preserves the audio and video synchronicity while still observing the individual effects through econometric analysis. Please find the anonymized links to videos used in the study in Table A3.

Table A3 Video Descriptions and Anonymized URLs (First Submission)

S.No	Video Description	Video Motion Type	Anonymized URL
1	Very_high_OF_super_fast_30.mp4	Very High	https://figshare.com/s/05ae0c9dfccd2578c570
2	High_OF_fast_30.mp4	High	https://figshare.com/s/f51f77bd8059b78437e0
3	Regular_OF_normal_30.mp4	Regular	https://figshare.com/s/dc72ff6fd04d31c690ac
4	Low_OF_slow_30.mp4	Low	https://figshare.com/s/06b34986ee83e074ad8a

We use Adobe Pro software to adjust the speed of the video {50% (Slow), 100% (Normal), 150% (Fast), 200% (Super-fast)}, which alters the OF. Moreover, we use optical flow-based time interpolation, present in Adobe Pro, to preserve the smoothness of the videos[11]. We also cut the video to the same length of 31 seconds, so that video duration does not affect viewership. Moreover, as mentioned earlier, YouTube video view count only depends on the first 30 seconds of the view, so video duration as long it is above 30 seconds should not affect the view count. We show the summary stats of the features for the four videos in Table A4. As a validity check, we can see that videos with higher speeds also have higher OF.

[11] https://helpx.adobe.com/in/premiere-pro/using/duration-speed.html

Table A4 YouTube Ads Videos - Features Summary Statistics

	Low (0.5x)	Regular (1x)	High (1.5x)	Very High (2x)
Kanade_Optical_flow	9.71	18.77	29.93	37.75
Farneback_Optical_flow	1.08	3.20	5.15	5.80
Audio_tempo	128.17	130.47	140.26	156.92
Audio_pitch	7.74	12.48	9.80	9.06

Experiment Setup The Google experimental setup requires us to first create video campaigns, which are then run using the setup to evaluate the effectiveness of the campaigns. We, therefore, create four video ad campaigns for each of the four videos using the standard YouTube ad campaign settings. For creating campaigns on the Google ads page, we select 'Video reach campaign' using 'Efficient reach (Bumper, skippable in-stream, or a mix)', which allows viewers to skip ads. For such campaigns, the only bid strategy available is 'Target CPM (cost per thousand impressions)', a way to bid wherein an advertiser sets the average amount willing to pay every thousand times the ad is shown. Using the 'Target CPM' amount, the Google ads system helps to get as many impressions as possible[12].

We select the campaign total to be 1000 for each campaign in local currency. We then choose 'YouTube videos' excluding 'YouTube search results' and 'Video partners on the Display Network' as ad networks to study the effects specific to the YouTube platform. As mentioned earlier, the campaigns are placed on the top 100 children's channels. The video ad campaigns are then run using Google's experimental setup to split the traffic among all four arms from July 6, 2023, to July 21, 2023.

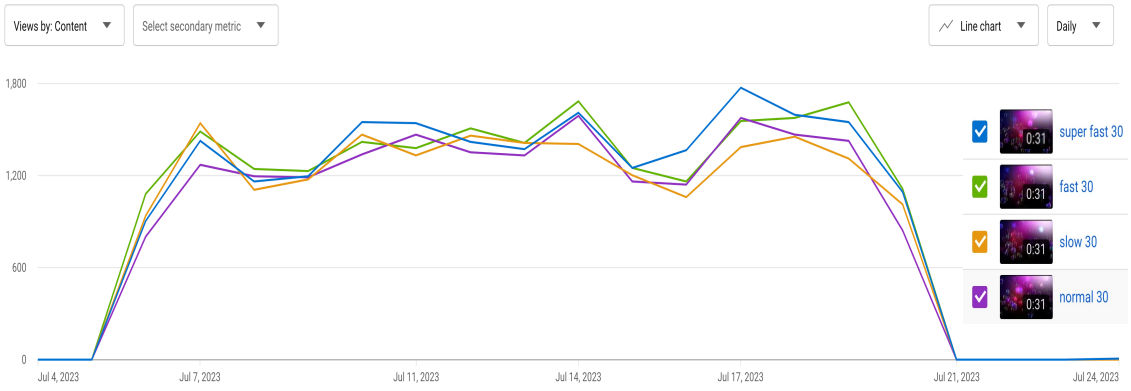
Data and Analysis: We summarize the video features in Table A4. We estimate $Kanade\_optical\_flow$ and $Farneback\_optical\_flow$ and other video features by employing the methodology delineated in Study 1. Unlike in the previous study, we do not apply logarithmic transformations to the view counts and optical flow metrics, as their distribution in the current data set is not skewed. For in-stream video ads, a view is counted when the video is watched to completion or

[12] https://support.google.com/displayvideo/answer/9173684?hl=en

has 30 seconds of view time[13], and an impression is counted every time the ad is served as a pre-roll. YouTube ads allow viewers to skip ads after 5 seconds, and for a video to get an ad view, the viewer should not use the 'Skip Ad' button till 30 seconds of the video is watched.

For model-free evidence, we show the views trend for the four videos using YouTube Analytics, allowing us to visualize the views per day for the four video campaigns in Fig. A4. The visual is taken from the YouTube Analytics platform, and the legends for different campaigns are shown on the right side of the figure. If we observe the trends carefully, we find that in the initial few days of posting the ads, no clear pattern is evident. However, in the last few days, the 'Fast' and the 'Super-fast' videos received more views.

Figure A4 **A screenshot from YouTube Analytics comparing video views for the two weeks when the video campaigns were run using YouTube Ads experiments. Best seen on a computer screen after zooming in.**



Besides visual analysis, we also used statistical models to estimate the effect of OF on ad views. We retrieve ad campaign results from the platform, which contains dates, hours, costs, views, and impressions for each video. We create an hour-level panel and conduct regression analysis to relate dependent variables (ads views) to different visual characteristics, as discussed next.

$$Views_{ij} = \alpha + \boldsymbol{\beta X_{ij}} + T_j + \epsilon_{ij} \tag{6}$$

[13] https://support.google.com/youtube/answer/2991785?hl=en

In the above equation, subscript i denotes the video OF index, j denotes the date-hour index, and $Views_{ij}$ represents the dependent variable of interest of campaign $video_i$ at $date-hour_j$. The independent variables are indicated by the set $\boldsymbol{X}$, and $\boldsymbol{\beta}$ represents their regression coefficients. We account for date-fixed effects using dummy variables represented by T_j. Note that these videos have no facial emotions (as no faces are present), and video contrast values and pixel count are approximately the same across videos.

Table A5 Ads Outcome for Different Campaigns

Arm (OF Type)	Traffic split	Cost	Impr.	Views
Low	25%	985.71	55,510	19,167
Regular	25%	985.51	54,371	19,099
High	25%	985.91	57,447	20,724
Very high	25%	986.34	58,514	20,746

We present the ads metrics in Table A5. We find that the advertisement cost is almost the same for the four arms. This is as expected as the ad campaign settings ensure that. However, we see a substantial difference in the views and the impressions. Video campaigns with higher OF ('High' and 'Very high') get more views and impressions. The 'Very high' OF campaign receives the highest number of impressions, and the 'High' OF campaign receives the highest number of views. We find that by increasing the OF (Regular to Very high), the video views increase by 8.6% (from 19,099 for Regular OF vs. 20,746 for Very high OF) without much change in the cost (985.5 to 986.3 in local currency).

Results and Discussion: We use a fixed effect model to estimate the regression coefficients and present the results in Table A6. We use video views count ('Views') as dependent variables and different types of OFs as independent variables. We also show the estimates controlling for ad impressions (similar to many prior works, e.g., (Bleier and Eisenbeiss 2015)) and audio characteristics. We find that video views have a positive and statistically significant association with OFs. We find similar positive effects for both 'Farneback' and 'Lucas-kanade' based measures of OFs for the videos used in the ads. We also find positive coefficients for 'Impressions' (0.33, $p<0.05$), which is expected as

Table A6 Effects of Optical-Flow and Audio Characteristics on Ads Outcome

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Views	Views	Views	Views	Views	Views
Kanade_Optical_flow	0.193**	0.072**	0.255*			
	(0.096)	(0.032)	(0.137)			
Farneback_Optical_flow				1.094**	0.450**	1.044*
				(0.556)	(0.183)	(0.560)
Impressions		0.333***	0.333***		0.333***	0.333***
		(0.004)	(0.004)		(0.004)	(0.004)
Audio_tempo			-0.182			-0.104
			(0.132)			(0.091)
Audio_pitch			-0.215			-0.232
			(0.275)			(0.281)
Observations	1,411	1,411	1,411	1,411	1,411	1,411
R-squared	0.093	0.897	0.897	0.092	0.897	0.897
Date FE	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

the impression is a precursor to video views, and therefore videos with higher impressions get higher views. However, even if we control for the number of impressions, we still see a positive coefficient for video views for both types of OFs, indicating video ads with higher OF get more views on the top 100 children's channels.

We don't observe any significant effects of audio features on video views. Including audio features reduces the significance of OFs because the variables ('Optical_flow' and 'Audio_tempo') are correlated. We still keep columns 3 and 6 results for completeness, allowing comparison to the results in Study 1.

A3. Study 3 - Supporting Information

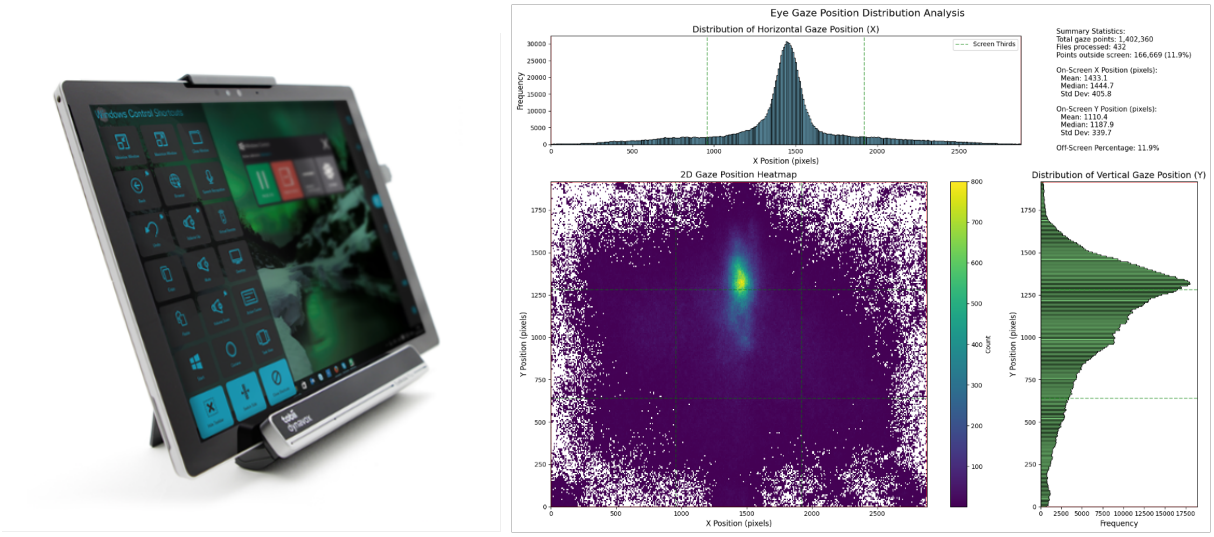
We provide additional information on 'Lab Setup and Study Procedure', 'Participant Selection' and 'Variables Description'.

A3.1. Lab Setup and Study Procedure

We conduct the experiment in a child-friendly laboratory space within a university library. Each participant uses a Microsoft Surface device fitted with a non-invasive eye tracker to capture real-time

gaze data (Figure A5). We chose this device because it closely resembles the typical screen environment children use at home, offering greater ecological validity than traditional desktop monitors. The lab environment minimizes distractions and provides comfortable seating for children, along with adjacent observation areas for parents to monitor the session unobtrusively.

Figure A5 **Illustration of a Microsoft Surface device with an eye tracker attached (left) and visualization of sample eye gaze data (right)**



A3.2. Variables Definition

We define the main variables for this study next:

A3.2.1. Screen Time (%): Screen Time (%) represents the percentage of a video viewing session during which a participant's eye gaze was detected within the physical boundaries of the tablet screen. This variable is calculated by first determining the total number of eye-tracking data frames recorded during a user-video session, then counting how many of those frames contained valid gaze coordinates that fell within the screen dimensions (0-2880 pixels horizontally, 0-1920 pixels vertically). The on-screen proportion is computed as (total frames - off-screen frames) / total frames, then multiplied by 100 to convert to a percentage. Values range from 0% (participant never looked at the screen) to 100% (participant maintained visual attention on the screen throughout the entire video). This measure captures basic visual engagement with the display device, regardless of whether

the participant was looking at educationally relevant content (AOI) or other areas of the screen, and excludes periods when gaze was undetected, outside screen boundaries, or when the participant looked away from the device entirely.

A3.2.2. Total Off-Screen Breaks: Total Off-Screen Breaks represents the number of discrete episodes during which a participant's gaze moved away from the screen or became undetectable by the eye-tracking system. This variable is calculated by identifying continuous sequences of frames where is_gaze_outside_screen == 1 (indicating gaze coordinates were missing, invalid, or fell outside the 2880×1920 pixel screen boundaries). The algorithm iterates through all eye-tracking frames chronologically, counting each time the participant's gaze transitions from on-screen to off-screen and back, treating each uninterrupted off-screen period as a single "break" regardless of its duration. This measure captures both brief interruptions like blinks or momentary glances away ($\leq$6 frames, $\approx$200ms) and longer attention breaks where participants look away from the screen for extended periods (>6 frames). The count includes all types of visual disengagement events, providing a comprehensive measure of how frequently participants' attention shifted away from the educational content display during video viewing sessions.

A3.2.3. Long Off-Screen Breaks: Long Off-Screen Breaks represents the number of extended episodes during which participants looked away from the screen for a sustained period, indicating deliberate attention shifts rather than brief interruptions. This variable is calculated by first identifying all continuous sequences where is_gaze_outside_screen == 1 (off-screen frames), then filtering to include only those sequences that exceed 6 frames in duration. The 6-frame threshold ($\approx$200ms at 30fps sampling rate) is based on eye-tracking literature that distinguishes between brief physiological events like blinks or micro-saccades ($\leq$6 frames) and intentional attentional disengagement (>6 frames). Each qualifying sequence represents a discrete instance where the participant consciously directed their visual attention away from the educational content, whether to look around the room, at other people, or engage in off-task behavior. This measure provides insight into sustained attentional lapses and voluntary disengagement from the learning material, as opposed to brief technical interruptions captured by short off-screen breaks.

A3.2.4. Short Off-Screen Breaks: Short Off-Screen Breaks represents the number of brief episodes during which participants' gaze was temporarily undetected or moved outside screen boundaries for very short durations, primarily capturing physiological events rather than deliberate attention shifts. This variable is calculated by identifying all continuous sequences where is_gaze_outside_screen == 1 (off-screen frames), then filtering to include only those sequences that last 6 frames or fewer. The 6-frame threshold (≈200 ms at 30 fps sampling rate) is based on eye-tracking literature distinguishing between involuntary physiological events and intentional attentional behavior. These short breaks typically represent natural eye blinks (100-400ms duration), micro-saccades, brief tracking losses due to head movements, or technical artifacts in the eye-tracking system rather than conscious decisions to look away from educational content. This measure helps researchers separate noise and unavoidable physiological interruptions from meaningful attentional disengagement, providing a cleaner assessment of deliberate attention patterns when analyzed alongside long off-screen breaks.

A3.2.5. Video Rating: Video Rating is the rating a participant provides for each video segment on a scale of 1 to 7.

A3.2.6. AOI Fixation Count For fixation-based analysis, we need an area of interest (AOI). In our educational videos, the foreground AOI- the instructor- is largely static and centrally positioned, whereas the experimental manipulation affects only the side background. Because the AOI location is stable across all frames, we were able to construct a conservative AOI that encompasses the instructor's head and upper torso (Figure 1) and to compute attention separately for AOI and non-AOI regions.

AOI Fixation Count represents the number of sustained visual fixations that were directed toward AOI region during video viewing. This variable is calculated through a two-step process: first, fixations are detected using velocity-based criteria (gaze speed <31 pixels/frame for ≥3 consecutive frames at 30fps), then each fixation is classified based on its spatial location. For each detected fixation, the algorithm counts how many individual frames within that fixation period had gaze coordinates falling

within the Area of Interest (AOI), which is defined by a pixel-level mask identifying the educational video content region on the 2880×1920 screen.

A fixation is classified as "AOI fixation" if 50% or more of its constituent frames were located within the AOI region. This measure captures sustained visual attention specifically directed toward pedagogically relevant material, as opposed to fixations on peripheral screen areas. The count excludes brief glances and requires sustained attention (minimum 3 frames ≈ 100ms) within the AOI region.

A3.2.7. Non-AOI Fixation Count: Non-AOI Fixation Count represents the number of sustained visual fixations that were primarily directed toward the non-AOI region of the screen during video viewing sessions. This variable is calculated using the same velocity-based fixation detection algorithm as AOI fixations (gaze speed <31 pixels/frame for ≥3 consecutive frames at 30fps), but applies the inverse spatial classification criteria. For each detected fixation, the algorithm counts how many individual frames within that fixation period had gaze coordinates falling within the Area of Interest (educational content region defined by pixel-level mask).

A fixation is classified as "non-AOI fixation" if fewer than 50% of its constituent frames were located within the AOI region, meaning the participant's sustained attention was primarily directed toward background regions visible on the 2880×1920 display. This measure captures sustained visual attention directed away from pedagogically relevant material while still remaining on-screen, providing insight into divided attention patterns and potential distraction by non-educational visual elements during learning sessions.

A3.2.8. AOI Fixation Duration (log): AOI Fixation Duration (log) represents the natural logarithm-transformed average length of sustained visual fixations directed toward educationally relevant content. This variable is calculated through a multi-step process: first, fixations are detected using velocity-based criteria (gaze speed <31 pixels/frame for ≥3 consecutive frames); second, fixations are classified as "AOI fixations" if 50% or more of their constituent frames had gaze coordinates within the educational content region defined by the pixel-level AOI mask; third, the mean duration

of these AOI fixations is computed in frames (at 30fps sampling rate); finally, this mean duration is log-transformed using the formula log(1 + mean_duration_frames) to normalize the distribution and handle cases where participants had no AOI fixations (avoiding log(0)). The log transformation is applied because fixation durations typically follow a right-skewed distribution, and the transformation helps meet statistical assumptions for regression analyses. This measure captures the typical amount of sustained visual processing time participants devoted to the AOI region during each fixation episode, with higher values indicating longer, more thorough visual engagement with the AOI.

A3.2.9. Non-AOI Fixation Duration (log): Non-AOI Fixation Duration (log) represents the natural logarithm-transformed average length of sustained visual fixations directed toward non-AOI areas of the screen. This variable is calculated through a multi-step process: first, fixations are detected using velocity-based criteria (gaze speed <31 pixels/frame for ≥3 consecutive frames); second, fixations are classified as "non-AOI fixations" if fewer than 50% of their constituent frames had gaze coordinates within the educational content region defined by the pixel-level AOI mask; third, the mean duration of these non-AOI fixations is computed in frames (at 30fps sampling rate); finally, this mean duration is log-transformed using the formula log(1 + mean_duration_frames) to normalize the distribution and handle cases where participants had no non-AOI fixations. This measure captures the typical amount of sustained visual processing time participants devoted to the non-AOI screen area during each fixation episode.

A3.3. Additional Eye-Tracking Analyses: Re-engagement vs. Seductive Details (Study 3)

This section provides detailed methodology and complete results for the four additional eye-tracking analyses reported in Study 3. These analyses use the frame-level gaze data (approximately 1.6 million rows across 49 children and 322 sessions) to distinguish between a re-engagement mechanism and a seductive details interpretation of the observed increase in non-AOI fixations reported in Table 8.

A3.3.1. Data and Definitions All analyses use the same frame-level eye-tracking data as the primary Study 3 analyses. Each frame (∼33ms at 30 Hz) is classified into one of three gaze regions: *area of interest* (AOI), *other screen area* (non-AOI), or *outside screen* (off-screen). We restrict the

sample to educational video sessions and use the same thresholds as the primary analysis pipeline: velocity threshold of 31.0 pixels/frame for fixation detection, minimum fixation duration of 3 frames (∼100ms), and a re-fixation distance threshold of 50 pixels. Off-screen sequences of 6 or fewer frames (∼200ms) are classified as blinks and excluded from the off-screen episode analyses (Analyses C and D). Stimulation conditions are coded as no motion (static control), medium motion, and high motion.

All comparisons across motion conditions use session-level means aggregated by child and condition. Statistical inference uses Kruskal-Wallis tests, which are appropriate given the non-normal distributions of many gaze metrics.

A3.3.2. Analysis A: Scan-Path Transition Rates We compute transition rates between all pairs of gaze regions (AOI, non-AOI, off-screen) for each session. Transition rates are expressed per 1,000 frames spent in the origin state, controlling for differences in time spent in each region across conditions. For example, the off-screen → AOI transition rate for a given session equals (number of off-screen → AOI transitions) / (total frames in off-screen state) × 1,000.

Table A7 Scan-Path Transition Rates per 1,000 Frames in Origin State, by Motion Condition

	No Motion		Medium Motion		High Motion			
Transition	Mean	SE	Mean	SE	Mean	SE	χ^2	p
AOI → non-AOI	6.24	0.56	8.69	0.56	7.96	1.20	16.56	<0.001
non-AOI → AOI	99.26	10.70	60.78	4.05	67.41	3.27	21.01	<0.001
Off-screen → AOI	120.78	14.97	130.21	11.88	140.58	12.12	2.41	0.299
Off-screen → non-AOI	22.90	1.78	33.98	3.32	38.24	4.29	4.35	0.114
AOI → Off-screen	23.48	4.84	16.73	1.46	16.77	1.94	1.79	0.409
non-AOI → Off-screen	57.00	6.10	29.80	2.12	37.10	5.66	21.93	<0.001

Session-level means. Standard errors clustered at the participant level. Kruskal-Wallis χ^2 test across three conditions.

The key finding is that the off-screen → AOI transition rate increases monotonically with motion (120.8, 130.2, 140.6), though this trend does not reach conventional significance ($p = 0.30$). At the same time, the non-AOI → off-screen rate decreases sharply with motion ($p < 0.001$), indicating that motion keeps children who are gazing at peripheral regions from leaving the screen entirely.

A3.3.3. Analysis B: Re-fixation Classification by Region We classify each re-fixation (a return of gaze to within 50 pixels of a previously fixated location) by whether it lands in an AOI or non-AOI region. This allows us to assess whether motion disproportionately increases non-AOI re-fixations, as a seductive details account would predict, or maintains the AOI share of re-fixations, consistent with re-engagement.

Table A8 Re-fixation Counts and AOI Share, by Motion Condition

	No Motion		Medium Motion		High Motion			
Metric	Mean	SE	Mean	SE	Mean	SE	χ^2	p
Total re-fixations	94.2	6.8	106.2	5.5	106.9	5.3	2.48	0.290
AOI re-fixations	91.1	6.7	99.7	5.5	102.7	5.2	2.06	0.356
Non-AOI re-fixations	3.1	0.5	6.5	0.7	4.3	0.4	26.49	0.000
AOI share (%)	94.7	1.4	92.2	0.8	95.6	0.4	20.41	0.000

Session-level means. Standard errors clustered at the participant level. Kruskal-Wallis χ^2 test across three conditions.

The AOI share of re-fixations remains high across all conditions (92%–96%) and, if anything, increases for high motion. The pattern in non-AOI re-fixations (highest for medium motion at 6.5, lowest for no motion at 3.1) is inconsistent with a seductive details account, which would predict monotonic increases.

A3.3.4. Analysis C: Re-engagement Latency We define re-engagement latency as the time (in milliseconds) from gaze leaving the screen to the first frame of on-screen gaze return. We compute this for each off-screen episode, excluding blinks (≤200ms off-screen sequences). We report mean session-level latencies stratified by return destination (AOI vs. non-AOI) and by break type (all episodes vs. long breaks only, where long breaks are defined as off-screen episodes exceeding 6 frames).

Re-engagement latency to AOI decreases monotonically with motion for both all episodes (2,545 → 1,147 → 811ms) and long breaks (4,538 → 2,469 → 1,887ms). Although these trends do not reach conventional significance, the monotonic pattern is consistent with the theorized mechanism of motion-driven feedforward re-orientation. Critically, the percentage of returns directed to AOI is stable across all conditions (∼74–77%), indicating that motion does not redirect post-break attention away from instructional content.

Table A9 Re-engagement Latency (ms) and Return-to-AOI Rate, by Motion Condition

	No Motion		Medium Motion		High Motion			
Metric	Mean	SE	Mean	SE	Mean	SE	χ^2	p
All off-screen episodes								
Latency, returns to AOI (ms)	2,545	1,472	1,147	158	811	82	1.62	0.445
Latency, returns to non-AOI (ms)	1,048	126	813	94	957	128	2.58	0.275
% returns to AOI	76.6	1.8	75.9	1.4	76.3	1.2	0.35	0.838
Long off-screen breaks only								
Latency, returns to AOI (ms)	4,538	1,849	2,469	270	1,887	123	0.92	0.632
Latency, returns to non-AOI (ms)	1,961	202	1,758	137	2,370	498	0.09	0.956
% returns to AOI	73.8	1.9	74.4	1.7	73.6	1.7	0.41	0.814

Session-level means. $N = 79$ (no motion), 121 (medium), 122 (high) sessions.

Kruskal-Wallis χ^2 test across three conditions. Long breaks > 6 frames (∼200ms).

A3.3.5. Analysis D: Return Destination Distribution For each off-screen episode, we classify the return destination as AOI or non-AOI based on the gaze region of the first on-screen frame following the break. We compare session-level return-to-AOI proportions across motion conditions using Kruskal-Wallis tests.

Table A10 Return Destination Following Off-Screen Episodes (%), by Motion Condition

	All Episodes		Long Breaks Only	
Condition	% to AOI	% to non-AOI	% to AOI	% to non-AOI
No Motion	76.6	23.4	73.8	26.2
Medium Motion	75.9	24.1	74.4	25.6
High Motion	76.3	23.7	73.6	26.4
χ^2 p-value	0.84		0.81	

Row percentages. Kruskal-Wallis test on session-level return-to-AOI proportions.

Return-to-AOI rates are virtually identical across conditions for both all episodes (75.9–76.6%) and long breaks (73.6–74.4%), with no statistically significant differences. This directly contradicts the seductive details interpretation, which would predict that motion diverts post-break gaze away from instructional content.

A3.3.6. Summary The four analyses are collectively more consistent with a screen-level re-engagement interpretation than with a pure seductive-details account. Motion appears to bring children back to the video (decreasing off-screen episode duration and latency), and once on-screen, children continue to return to the instructor at similar rates across motion conditions. The increased non-AOI fixations reported in the main analysis (Table 8) arise from extended on-screen exploration of the motion-enriched background during continuous viewing, not from a redirection of post-break returns away from the area of interest.

A4. Study 4 - YouTube Ads Experiments: Comparing Kids and Grownups Video Viewership

It is important to elicit the differences in the preferences of children and grownups to understand the factors that only influence the kids. Therefore, this study aims to determine whether the increase in optical flow (OF) affects kids vs. grownups differently. As in the last study, we use kids to refer to the viewers of 'made-for-kids' videos that, as per YouTube, are aimed at children below 13 years. We use grownups to refer to those who do not primarily consume 'made-for-kids' content and are likely to be of higher age.

While age is the intended differentiator, our experiment setup precludes us from retrieving age. Instead, we use the target channels for ad placement as a proxy. For kids, as in the previous study, we target the top 100 channels that produce children's content[14], also known as made-for-kids channels. For grown-ups, we use the top 100 channels categorized as 'people' on social blade[15]. These channels are from influencers who create their own videos and put them on YouTube, and for these channels, children are not the intended audience. We remove any channels in the 'people' category that also created 'made-for-kids' videos to keep the sets exclusive.

Experimental Setup

We use a similar Google ads setup as used in the previous study. However, in this study, we manipulate the OF {Regular, Very-high} and the target channels {Children x Grownups}. We create four

[14] https://socialblade.com/youtube/top/category/made-for-kids

[15] https://socialblade.com/youtube/top/category/people

Table A11 Ad metrics for video ads targeting kids and grownups.

Arm	Traffic split	Cost (local currency)	Impr.	Views
Children				
Regular_OF	50%	1,983.40	88,287	29,572
Very high_OF	50%	1,980.88	93,345	31,711
Grownups				
Regular_OF	50%	1,980.68	30,931	8,315
Very high_OF	50%	1,979.57	32,367	8,211

YouTube-based video campaigns (two using videos with 'Regular' OF, and two using videos with 'Very-high' OF), with the same settings (except placements) as described in the last study. As the video remains the same, the summary statistics of the video features are still the same (available in Table A4). For picking the ad placement, as mentioned earlier, our selection criteria included channels from the top 100 children's channels (kids_campaign) and the top 100 people's channels (grownup_campaign).

As we have different targets (100 made-for-kids channels and 100 people's channels), Google Ads experimental setup does not allow running experiments with such video campaigns. Therefore, instead of running one experiment with four video campaigns, we used two experiments with two video campaigns each (the first video with 'Regular' OF and the second video with 'Very high' OF). We run the campaigns from 16^{th} of May, 2023 to 11^{th} of June, 2023. Except for the target channels, the two experiments used the same parameters, including budget (2000 in local currency), and start and end date. Google's experimental setup ensures that users are not shared across different experiment arms. As the target channels are also two exclusive sets, we do not expect any viewer to be shared across all experimental arms, and, therefore, the experimental results should represent the differences in the four targeted groups.

Results and Discussion

The results from the ads campaign, detailed in Table A11, reveal distinct patterns. The campaign targeting children characterized by video with 'Very high' OF secures more views (Very high=31,711 versus Regular=29,572). In contrast, the grownup segment shows a negligible difference in views

(8,315 versus 8,211), with the 'Very high' OF campaign getting slightly fewer views. Notably, campaigns using videos with 'Very-high' OF garner more impressions for both children and grownups, suggesting that YouTube's algorithm might leverage visual features to prioritize video ads for increased circulation.

We created a panel data comprising views, impressions, video type, campaign type, and date-hour for analysis. We then estimate the effects of OFs and campaign-related independent variables on video views. To better understand the effects of differences in OF types and video campaigns, we interact OF types with differences in campaign placements (Grownups vs kids) in the regressions. In the below equation (7), subscript i denotes the video OF type index, j denotes the date-hour index, and $Views_{ij}$ represents the dependent variable of interest of $video_i$ on $date-hour_j$. Additionally, we control for impressions (γ) and account for date-fixed effects using dummy variables represented by T_j.

$$Views_{ij} = \alpha + \boldsymbol{\beta}[\text{video\_type}_i \times \text{campaign\_type}_i] + \gamma \text{Impressions\_num}_{ij} + T_j + \epsilon_{ij} \tag{7}$$

Table A12 Interactive effects of videos with different video motions on the top children's channels and the top channels for grownups.

	(1)	(2)
VARIABLES	Views	Views
Regular_OF#kids_campaign	27.809***	4.731***
	(0.884)	(0.358)
Very high_OF#grownup_campaign	-0.023	-0.561**
	(0.325)	(0.230)
Very high_OF#kids_campaign	30.250***	5.074***
	(0.971)	(0.349)
Impressions		0.304***
		(0.004)
Observations	2,592	2,592
R-squared	0.448	0.917
Date FE	Yes	Yes
Robust standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.10		

In Table A12, we show the effects of two types of OFs ('Regular', 'Very high') when the video campaigns were run on the top children's channels (kids_campaign) and the top channels for grownups (grownup_campaign). In the first column, we show the regression coefficients without 'Impressions', and in the second column, we add 'Impressions' as a control variable.

The regression coefficients have the same directionality with and without 'impressions' as the control. The regression coefficients are positive for both ('Regular OF' $(4.73, p < 0.01)$ and 'Very high OF' $(5.07, p < 0.01)$) OF types when interacted with kids_campaign. Using 'Regular_OF#grownup_campaign' as the base, we find video campaigns targeting children's channels receive more views. Moreover, for kids_campaign, the video campaign with a higher OF, labeled 'Very high OF', receives even more views $(5.07, p < 0.01)$ than the video with a lower one ('Regular OF'). We also find that video with a higher OF when interacted with 'grownup_campaign' has a negative association, indicating that the ads with the higher OF on the channels viewed by grownups receive fewer views.

The empirical evidence suggests distinct viewership patterns for videos with varying OF levels between audiences of leading channels for children and grownups. Specifically, for the experimental videos, we find a higher viewership for campaigns with higher OF when the video campaigns are placed on the top children's channels. The result holds even after controlling for impressions. This implies that children, the primary viewers on these channels, conditioned on getting an impression, are more likely to keep watching the video campaigns for at least 30 seconds, resulting in more views. In contrast, the views of the video campaigns with higher OF are lower when the campaigns are run on the top channels for grownups. The data thus reveal a unique pattern in how visual stimuli in videos affect the viewers of children's channels differently from that of grown-ups.

One limitation of our experiment is that it does not allow us to directly isolate the age of the viewers who were exposed to the ad campaigns. Prior work in developmental psychology highlights established attentional differences between children and adults. Children's attention is likely more sensitive to low-level perceptual cues like motion and visual salience, whereas adults rely more heavily on semantic

and narrative structures. Our objective is not to re-establish these known developmental distinctions, but rather to test whether dynamic visual motion—as operationalized by optical flow—is a stronger predictor of engagement in content settings primarily designed for young viewers. Nonetheless, given that the data were collected in a heterogeneous setting, we advise interpreting the results with caution.

A5. YouTube Videos Dataset

As mentioned in the paper, the list of top children's channels is obtained from the Social Blade website, which tracks the top players on many social media platforms, including YouTube. In particular, for the children's list, we used the list of top channels that produce made-for-kids content[16]. Sample stats for some of the children's channels are shown in Table A13. The channel IDs are listed in Table A14, and Table A15.

References

Bleier A, Eisenbeiss M (2015) Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science* 34(5):669–688.

Choi S, Go YM, Park K (2025) Capturing children's attention: AI-driven analysis of thumbnail design in YouTube educational content. *Proceedings of the Pacific Asia Conference on Information Systems (PACIS) 2025*, URL `https://aisel.aisnet.org/pacis2025/is_education/is_education/9`.

Deng J, Guo J, Ververas E, Kotsia I, Zafeiriou S (2020) RetinaFace: Single-shot multi-level face localisation in the wild. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5203–5212.

Farnebäck G (2003) Two-frame motion estimation based on polynomial expansion. *Proceedings of the 13th Scandinavian Conference on Image Analysis (SCIA 2003)*, 363–370 (Springer).

Gibson JJ (1950) *The perception of the visual world* (Boston: Houghton Mifflin).

Grotewell PG, Burton YR (2008) *Early childhood education: Issues and developments* (Nova Publishers).

Hutto C, Gilbert E (2014) VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the international AAAI conference on web and social media*, volume 8, 216–225.

Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)*, 674–679 (Vancouver).

Serengil SI, Ozpinar A (2020) Lightface: A hybrid deep face recognition framework. *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 23–27 (IEEE), URL `http://dx.doi.org/10.1109/ASYU50717.2020.9259802`.

Sivasankar M, Bauer JJ, Babu T, Larson CR (2005) Voice responses to changes in pitch of voice or tone auditory feedback. *The Journal of the Acoustical Society of America* 117(2):850–857.

Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: Closing the gap to human-level performance in face verification. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1701–1708.

Varghese R, Sambath M (2024) YOLOv8: A novel object detection algorithm with enhanced performance and robustness. *2024 International conference on advances in data engineering and intelligent computing systems (ADICS)*, 1–6 (IEEE).

[16] https://socialblade.com/youtube/top/category/made-for-kids

Table A13 Sample list of YouTube channels considered. 'B' implies billions, 'M' implies millions, 'K' is thousands. Channel statistics reflect data as on 20 Jan 2022.

Channel Name	Channel Views	Channel Subscribers
Cocomelon - Nursery Rhymes	152B	155M
Pinkfong Baby Shark - Kids' Songs & Stories	36B	65M
Ryan's World	54B	34M
BabyBus - Kids Songs and Cartoons	25B	31M
Sesame Street	21B	23M
PowerKids TV	17B	33M
Kids TV - Nursery Rhymes And Baby Songs	12B	21M
HobbyFamilyTV	8B	4M
Tayo the Little Bus	7B	9M
Little Treehouse Nursery Rhymes and Kids Songs	5B	10M
Jugnu Kids - Nursery Rhymes and Best Baby Songs	4B	10M
Kids Channel - Cartoon Videos for Kids	3B	6M
Shemaroo Kids	3B	8M
King Games	2B	3M
WildBrain Kids	2B	6M
KidsBabyBus HD	1B	2M
AllToyCollector	1B	1M
Cartooning Club How to Draw	992M	4M
Polly Pocket	813M	1M

Table A14 List of the Top (1-50) Children's Channels - 1

www.youtube.com/channel/UCBnZ16ahKA2DZ_T5W0FPUXg
www.youtube.com/channel/UChGJGhZ9SOOHvBB0Y4DOO_w
www.youtube.com/channel/UCwjoPtSoNLAoX2sLBaKLYng
www.youtube.com/channel/UCKjjmMqvuxNYw8c6knceexQ
www.youtube.com/channel/UCLsooMJoIpl_7ux2jvdPB-Q
www.youtube.com/channel/UC7Pq3Ko42YpkCB_Q4E981jw
www.youtube.com/channel/UCoookXUzPciGrEZEXmh4Jjg
www.youtube.com/channel/UCRx3mKNUdl8QE06nEug7p6Q
www.youtube.com/channel/UCKcQ7Jo2VAGHiPMfDwzeRUw
www.youtube.com/channel/UCc-2P5tCezbxegb7gxp6EXg
www.youtube.com/channel/UCbCmjCuTUZos6Inko4u57UQ
www.youtube.com/channel/UCcdwLMPsaU2ezNSJU1nFoBQ
www.youtube.com/channel/UC4NALVCmcmL5ntpV0thoH6w
www.youtube.com/channel/UCpYye8D5fFMUPf9nSfgd4bA
www.youtube.com/channel/UCj-SWZSE0AmotGSQ3apROHw
www.youtube.com/channel/UCsCe7SNQckiRJ6y563SIupg
www.youtube.com/channel/UCpzYfBXbEHHQHU2e89jM9Tg
www.youtube.com/channel/UC3KknIJZXRygH2pZ6MDtGbg
www.youtube.com/channel/UCI8V0wOaigGdBpv5uk7DxxQ
www.youtube.com/channel/UCej8z9NGaA9Hdd8k5hueXKw
www.youtube.com/channel/UC9iBBFfq7L3ipvodSLrU8gQ
www.youtube.com/channel/UCCc1DMB-AcOssKJ7KweLXBg
www.youtube.com/channel/UCMfZ_z0LUm805JOZLktl2QQ
www.youtube.com/channel/UC2VERcmMqi3vdtoQQ0ZP2sA
www.youtube.com/channel/UC8MR0wSTbzs5Yo7DgP04P-w
www.youtube.com/channel/UCbt63GNsB5wet6NO3dmhssA
www.youtube.com/channel/UCbmWOFTbZoDLWqG1rvFnJ0g
www.youtube.com/channel/UC6LKuH7RPkvRmzS9-8URtqA
www.youtube.com/channel/UC5mMf2NE17_MiTL94ihRN2w
www.youtube.com/channel/UC5mMf2NE17_MiTL94ihRN2w
www.youtube.com/channel/UChz5aEi3dfrDVC8-YJsMUDA
www.youtube.com/channel/UC-biucJWhM8HwjsQ96uoIUw
www.youtube.com/channel/UCUJz3Kx_UkivcPzdAmcFAPg
www.youtube.com/channel/UCTRNV3t2jxbkZgBjhyDv8UQ
www.youtube.com/channel/UC2VjS4B4YBTH00TE-SMD_aw
www.youtube.com/channel/UCVHO-W63u8RrXa6BAHxfLtQ
www.youtube.com/channel/UCHSnzMeCGtVOjenB1Grnn5Q
www.youtube.com/channel/UCMyyXCurk6jXXOWt2eKyCgg
www.youtube.com/channel/UCmKy_mpl8mCDDTTg96d7mhw
www.youtube.com/channel/UCAwIKJ84zBsiuuy9DVR13mA
www.youtube.com/channel/UCfFzNEYxqsSGoiw1ojgEhjw
www.youtube.com/channel/UCGVrxaJHUQHbPSUVDZ8sUVQ
www.youtube.com/channel/UCpQfF_qehNal4D7KpakObuA
www.youtube.com/channel/UCxezak0GpjlCenFGbJ2mpog
www.youtube.com/channel/UCZA2NJWgiWtLQZPfMhhEZUQ
www.youtube.com/channel/UC7EFWpvc1wYuUwrtZ_BLi9A
www.youtube.com/channel/UCyXWYhbJomJcTUg98MR5PFA
www.youtube.com/channel/UCtBOJ9bEAoMOY9keGK6p2hQ
www.youtube.com/channel/UCotX63w9fF1eTCjda7Ux3Rw
www.youtube.com/channel/UCsiqlKIUDHZJVG4smMvhF4w
www.youtube.com/channel/UC5uIZ2KOZZeQDQo_Gsi_qbQ

Table A15 List of the Top (50 - 100) Children's Channels - 2

www.youtube.com/channel/UC5wtOwEHCG__JAIpDdkiaZ6A
www.youtube.com/channel/UCwynFfAvCF2kiRSMWiozQHg
www.youtube.com/channel/UC0l7SdYHH6WAmXECumlKSkg
www.youtube.com/channel/UCmKsULSe7oGtdGcv3bz-__bw
www.youtube.com/channel/UCkJYEAR9ZM7hvAM-SkniJ0Q
www.youtube.com/channel/UCGI6tdFKWbSUDSw0ceI4mLQ
www.youtube.com/channel/UCcUqbcfVrRSImkF0vDdgc9w
www.youtube.com/channel/UCEJfVga185COFm1MsN4KQUA
www.youtube.com/channel/UC5pJi9mLIy38m2e__u3sboKQ
www.youtube.com/channel/UC3PRgO4VJJxZ6nTVHazUVKA
www.youtube.com/channel/UCbVGLnrXJzqwvxS9hyTsWDw
www.youtube.com/channel/UCzGoN1TCO2uiAQWyiNrGhuw
www.youtube.com/channel/UCvuXJ7jxbGakmLXgwRn88gQ
www.youtube.com/channel/UCTlThMcv2OFj7KfbOFhLrvg
www.youtube.com/channel/UCmumhdE__anzs3__jVgZHWndw
www.youtube.com/channel/UCGdQ8OUs2qM0gwkaa__MtFNg
www.youtube.com/channel/UCWCOoCApFUtccYENOYeL1uw
www.youtube.com/channel/UCKyywickFyBGRRxbxpn7CGA
www.youtube.com/channel/UC1THeiEwEW3N02ZmqmDKquQ
www.youtube.com/channel/UCqKNzUHmCrvZsFDzpAR3GjA
www.youtube.com/channel/UCtvYrG5IeaYNeHDKjKJJNRA
www.youtube.com/channel/UC8CssPJ0drSvFQ2nAq1TKSQ
www.youtube.com/channel/UC8Ae-2QbWgFjgAfFMHn02Zw
www.youtube.com/channel/UCyxe3g0khLIXRxRd3G9qSrw
www.youtube.com/channel/UCihTA5crIIdkStBXZNxmI-A
www.youtube.com/channel/UCrZ9CtlECK__Jl92di5fCfSg
www.youtube.com/channel/UCn3dC5Pe__PLudpWIEWL1v__w
www.youtube.com/channel/UCXXunTfMOcUlUeHlSDxmsCQ
www.youtube.com/channel/UC68XdlxjeqSsjoAdR3nw9Vg
www.youtube.com/channel/UCuiTofxf5rE__oZoE7LVQj-g
www.youtube.com/channel/UCE8YdasdvUDlAyFdFcI-AHg
www.youtube.com/channel/UCHyIeqvhdd85GxtfWBkbEgw
www.youtube.com/channel/UCLJ5EBmiuSLBcQR-R6rGVHA
www.youtube.com/channel/UC-VrU4EkjIMkIik0-7La5OA
www.youtube.com/channel/UCJcc5qvHowxRshW9P6YZwKA
www.youtube.com/channel/UCxidbvHajzXm-9wBK-w8IZA
www.youtube.com/channel/UCJxd2QGa3__-j1dbGcynjAlw
www.youtube.com/channel/UC__nVcTv8k1SETJ1QFrwz7tA
www.youtube.com/channel/UCaKssrpIyGDuKsuOjZhoWfg
www.youtube.com/channel/UCPWPAf6QR__oAcd__nfsLwzsg
www.youtube.com/channel/UC157gj–r7hkPTqNu9bS72Q
www.youtube.com/channel/UC0KchmlIoXBZSvumBr97PCQ
www.youtube.com/channel/UC__M0w5yZ5J7BxpC51NPszoQ
www.youtube.com/channel/UCxs5IqCTpCRYd39tbfiYbKg
www.youtube.com/channel/UCGu9-9FqQtXAitxkOnHzz6Q
www.youtube.com/channel/UCbmsBnSEOyX7lI9MACnJMLw
www.youtube.com/channel/UCfxN7O1NzZqoGBCJi6V6wpA
www.youtube.com/channel/UC2RNg__QGZriSGQo6enPLpeQ
www.youtube.com/channel/UCvlE5gTbOvjiolFlEm-c__Ow
www.youtube.com/channel/UC5PYHgAzJ1wLEidB58SK6Xw