vRetention: A User Viewing Dataset for Popular Video Streaming Services

Bo-Rong Chen, Jiayu Zhu, Yanxin Jiang, and Yih-Chun Hu
University of Illinois Urbana-Champaign

Abstract

Adaptive bitrate streaming (ABR) and quality of experience (QoE) metrics are proposed to enhance video streaming quality across various Internet connections. Traditional approaches to evaluating these metrics often ignore common user behaviors like seeking, jumping, or replaying video segments, leading to gaps in QoE understanding. Addressing this, we collected 229,178 audience retention curves from YouTube and Bilibili, offering a thorough view of viewer engagement and diverse watching styles. Our analysis reveals notable behavioral differences across countries, categories, and platforms. The YouTube data highlights varied content preferences, such as gaming and entertainment in some countries, and music, travel, and pets & animals in others. Additionally, Bilibili shows trends of early video abandonment, possibly influenced by platform-specific factors and shorter video formats. This enhanced grasp of user engagement aids in refining ABR and QoE metrics. We also highlight several potential applications of our dataset.

CCS Concepts • **Information systems** → **Multimedia streaming**;

ACM Reference Format:
Bo-Rong Chen, Jiayu Zhu, Yanxin Jiang, and Yih-Chun Hu. 2024. vRetention: A User Viewing Dataset for Popular Video Streaming Services. In *ACM Multimedia Systems Conference 2024 (MMSys '24), April 15–18, 2024, Bari, Italy*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3625468.3652175

1 Introduction

Video streaming represents 65% of global Internet traffic [19] and nearly 80% of mobile data traffic [4]. Platforms, like YouTube and Bilibili, and researchers focus on enhancing streaming quality over diverse Internet connections. They

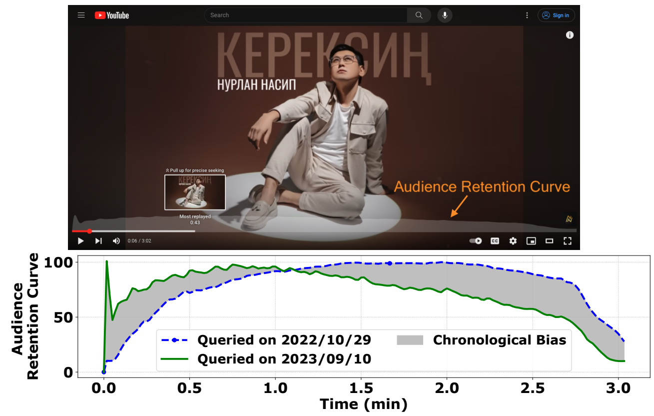

Figure 1. (top) Audience retention curve of video ('QBUXsvjZIWI') with over 10M views. (bottom) Chronological bias exists between two queries on different dates.

develop quality of experience (QoE) metrics [10] and adaptive bitrate streaming (ABR) techniques [5, 12, 13, 16, 21]. However, evaluations often assume users watch videos entirely, overlooking behaviors such as seeking, jumping to, or replaying parts of videos [6–8]. This leads to ABR evaluations following a full-viewing model, with current QoE metrics failing to differentiate between rebuffering caused by users skipping ahead or due to buffer exhaustion. The continuation of this evaluation method is largely due to a lack of extensive data on diverse user viewing behaviors.

As a result, the goal of our work is to develop the *largest open repository* of user viewing data. To this end, we collect 229,178 audience retention curves from YouTube and Bilibili to create our dataset, which we will make publicly available upon the publication of this paper. Each audience retention curve represents how many average users watch a particular video segment, reflecting the level of interest associated with each video moment, as shown in Figure 1(top), which are widely available on both YouTube and Bilibili websites. In our work, we use the audience retention curve to analyze the watching style and behaviors of users across different countries and video categories.

Contributions. We summarize our contributions as follows:

- Our dataset comprises 229,178 audience retention curves from both YouTube and Bilibili; our analysis (§ 2) shows that user behaviors vary dynamically across different countries, categories, and platforms.
- The analysis of YouTube data reveals that video popularity duration and audience retention vary significantly across

MMSys '24, April 15–18, 2024, Bari, Italy

ACM ISBN 979-8-4007-0412-3/24/04…$15.00
https://doi.org/10.1145/3625468.3652175

countries and categories, with gaming and entertainment being popular in some countries, and music, travel, and pets & animals in others, reflecting diverse viewing patterns and preferences.

- Bilibili's audience retention curves indicate higher early abandonment compared to YouTube, influenced by platform incentives and shorter video lengths, with over 75% of videos being less than 10 minutes and over 35% showing less than 2 spikes in audience retention, suggesting less replayability.

2 Dataset Collection And Analysis

We collected and analyzed audience retention curves from both YouTube and Bilibili, the most popular video streaming services worldwide and in China, respectively.

2.1 YouTube Dataset Collection

Our YouTube dataset consists of two parts: (i) video metadata and (ii) audience retention curve. YouTube provides audience retention curves using a scalable vector graphic (SVG) with class `ytp-heat-map-path`. We found that the curve is usually released once a video has 50K+ views 5 days post-publication. To *quickly gather many audience retention curves*, we query the most popular videos using the YouTube Data API [24], focusing on the 20 countries with largest audience size [22], including IN, US, ID, BR, RU, JP, MX, DE, PK, VN, GB, TR, PH, FR, KR, EG, IT, TH, ES and BD country codes, totaling 1.96 billion users across the world. YouTube Data API returns the most popular videos (usually newly published) in each country on the query date (10/28/2022–12/07/2023), which results in 199,041 videos in our dataset. We use the list method of YouTube Video API to query the metadata of each video. Finally, selenium [1] is used to automate querying audience retention curve by launching the YouTube website and retrieving retention curve. The description of the data is listed in Table 1.

Reducing Chronological Bias. TVSum [15] states retention curves are subject to chronological bias; once retention curves are available, viewers often jump to the "Most Replayed" segments, which further increases the popularity of the segments. Figure 1(bottom) shows the difference between audience retention curves returned on different dates. We avoid chronological bias by querying only "fresh" data. Specifically, we run our query daily on newly released videos, and obtain their audience retention curve once available. As a result, our data are less likely to have chronological bias, returning data is drawn almost entirely from a single distribution.

Given our datasets, we now investigate:

1. What is the duration of popularity for each video category and country? (YouTube: § 2.2.1)
2. What is the distribution of video length for each video category (and country)? (YouTube: § 2.2.2; Bilibili: § 2.3.1)
3. For each video category (and country), what is the distribution of audience retention curve? (YouTube (non-YouTube Shorts): § 2.2.3; Bilibili: § 2.3.2)
4. How many spikes do audience retention curve have? (YouTube (non-YouTube Shorts): § 2.2.4; Bilibili: § 2.3.3)

Parameter	Description
id	*string*. The video ID.
publishedAt	*datetime*. The published date of the video.
title	*string*. Title of the video.
categoryId	*string*. The category of the video.
duration	*long*. The length of the video.
viewCount	*long*. The number of times the video has been viewed.
countryCode	*string*. The country of the video.
date	*datetime*. The query date of the video.
retentionCurve	*string*. SVG Path of audience retention curve.

Table 1. YouTube dataset parameters. 199,041 videos in total collected from 10/28/2022–12/07/2023. Dataset available at: https://github.com/flowtele/vRetention.

2.2 YouTube Dataset Analysis

2.2.1 Duration of Popularity

The breakdowns of YouTube dataset are listed in Table 2. We define the *duration of popularity* as the number of days a video remains in the popular list. Figure 2 shows the median popularity duration and the average view count for each video group. RU has the shortest popularity duration, since most videos are gaming (Ga) and entertainment (En), which have a median popularity duration of 2 days (in RU). We plot the top-3 video breakdowns by category in RU in Figure 3(left), which shows that the RU audience likes to watch Ga and En videos, which have short popularity duration, resulting in a large number of RU videos in our YouTube dataset. This corresponds to Table 2, where Ga/En and RU are in the top-2 categories and countries. We are unsure how representative our RU data may be, since RU is no longer in

Category [1–8]	# Videos	Category [9–15]	# Videos
Entertainment (En)	57,143	Howto & Style (HS)	4,568
Gaming (Ga)	27,744	Autos & Vehicles (AV)	4,278
People & Blogs (PB)	25,320	Education (Ed)	2,781
Sports (Sp)	23,124	Science & Technology (ST)	2,856
Music (Mu)	19,558	Travel & Events (TE)	1,579
News & Politics (NP)	13,552	Pets & Animals (PA)	1,142
Comedy (Co)	8,649	Nonprofits & Activism (NA)	295
Film & Animation (FA)	6,451	Total	199,041
Country [1–10]	**# Videos**	**Country [11–20]**	**# Videos**
Russian Federation (RU)	27,876	Japan (JP)	7,819
India (IN)	23,102	Indonesia (ID)	6,591
Thailand (TH)	17,924	Korea (South) (KR)	6,485
Spain (ES)	12,908	Mexico (MX)	6,363
France (FR)	12,804	United Kingdom (GB)	6,087
Italy (IT)	11,754	Egypt (EG)	5,597
United States of America (US)	10,669	Philippines (PH)	4,473
Germany (DE)	10,491	Vietnam (VN)	4,160
Brazil (BR)	9,846	Pakistan (PK)	3,354
Turkey (TR)	8,970	Bangladesh (BD)	1,768

Table 2. The number of videos in each category (top) and country (bottom) (YouTube).

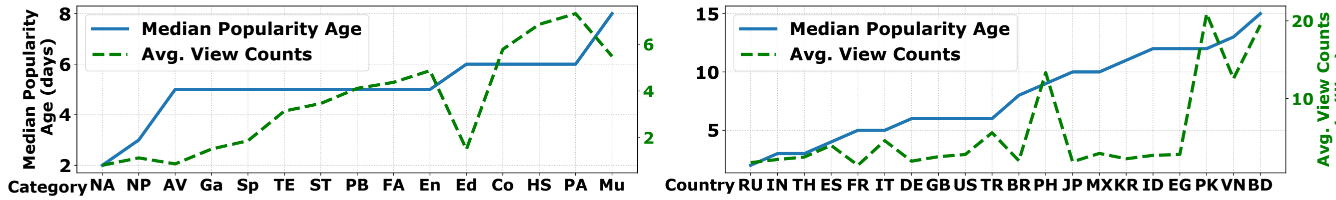

Figure 2. The median popularity duration and average view count by category (left) and country (right).

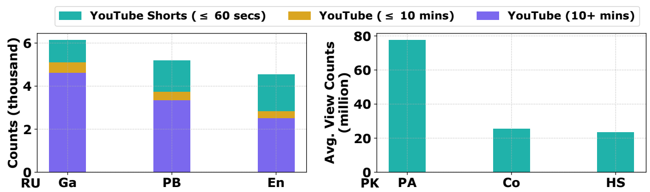

Figure 3. (left) Videos by category in RU and (right) the average view counts in PK by top-3 categories.

the list of countries with the most YouTube users, possibly because of increased political divisions between Russia and the West after February 2022. Moreover, IN and TH audiences tend to watch En videos, so they have the 2nd shortest median popularity ages. Second, music (Mu) and pets & animals (PA) not only have long popularity ages, but higher average view counts, indicating users tend to watch these videos repetitively. Users in PK and BD like to watch PA videos, so those regions have long popularity duration and top-2 average view counts, as shown in Figure 3(right). Finally, news & politics (NP) has very short popularity duration (2nd in Figure 2(left)), as previously found [7].

Takeaway. Users in different countries watch different types of videos. For example, users in RU/IN/TH like to watch gaming and entertainment videos, whereas users in PK/BD like pets & animals videos.

2.2.2 Video Length

Figure 4 shows the normalized counts of YouTube Shorts (≤ 60 seconds), YouTube (≤ 10 minutes), and YouTube (10+ minutes). First, over 80% music (Mu) videos are less than 10 minutes, since songs tend to be shorter than 10 minutes, whereas travel & events (TE), autos & vehicles (AV), and gaming (Ga) have the longest videos, with over 75% of videos over 10 minutes. Also, over 50% of PK videos are YouTube Shorts; users in PK like to watch YouTube Shorts and most of views are pets & animals (PA) (Figure 3(right)).

Takeaway. YouTube Shorts are very popular in PK, and over 80% of music videos are less than 10 minutes.

2.2.3 Area under Audience Retention Curve (AUC)

We exclude YouTube Shorts for the following analysis, since watching behaviors of YouTube Shorts are different due to swipe patterns [17]. To show the distribution of audience retention curve, we illustrate the distribution of audience retention curve using boxplots where the x-axis is scaled to percentage length. Figure 5(left) shows music (Mu) has a bell-shaped median, indicating that users tend to skip the intro and watch the verse, or replay the chorus. However, Figure 5(middle) and 5(right) show the median of AUC of entertainment (En) and gaming (Ga) are flat with a large number of outliers (like all other categories except music), meaning that unlike music, the highest retention can appear in any position in these videos, because the structure of such videos is not well-defined as music. We conclude that (i) users do not commonly watch a video from start to finish. Instead, users tend to seek, jump to interesting video segments, or replay the video. (ii) Each video segment has different audience retention, meaning some video segments are more popular than others.

Takeaway. Music videos reflect stable retention in each video segment, whereas in other video categories, the significant number of outliers suggests that the user watching behaviors are highly dynamic.

2.2.4 Spikes of Retention Curve

The spikes in audience retention curve indicate video segments that are more-watched. We use the peak detection algorithm of the wavelet transform [9], as implemented by `signal.find_peaks_cwt` in scipy [20], with widths $\in [10, 100]$s to detect spikes in audience retention curve. Figure 6(left) shows that over 50% of music (Mu) videos have ≤ 2 spikes; whereas over 80% of gaming (Ga), travel & events (TE), and autos & vehicles (AV) videos have 6+ spikes, showing that in these categories, users tend to seek, jump to, or rewatch particular interesting segments, as demonstrated in Figure 6(right), where 70% of top RU videos have 6+ spikes (due to the prevalence of gaming videos).

Takeaway. Music videos display less seek behavior, with over 50% of videos having ≤ 2 spikes; whereas over 80% of gaming, travel & events, and autos & vehicles videos have 6+ spikes, suggesting some video categories tend to see more seek and rewatch behavior.

Early Abandonment in Short Videos. In Figure 5, we show that the music category has more stable audience retention curve than other categories from YouTube videos (excluding YouTube Shorts). However, we find that the distribution of audience retention curve of YouTube Shorts indicates early abandonment, as shown in Figure 7, where all 3 categories (including music) have early abandonment.

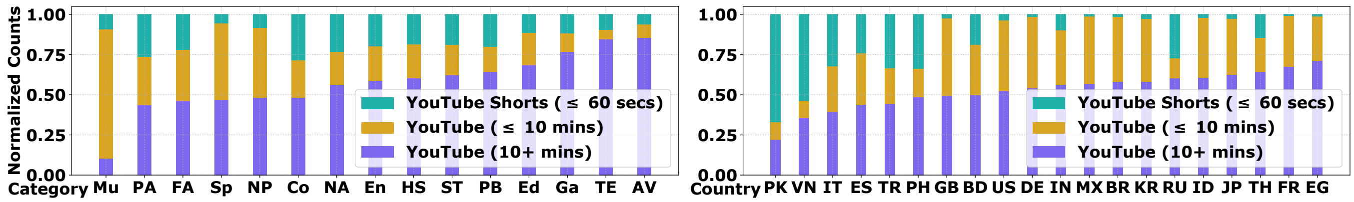

Figure 4. Length of videos from the YouTube dataset by category (left) and country (right).

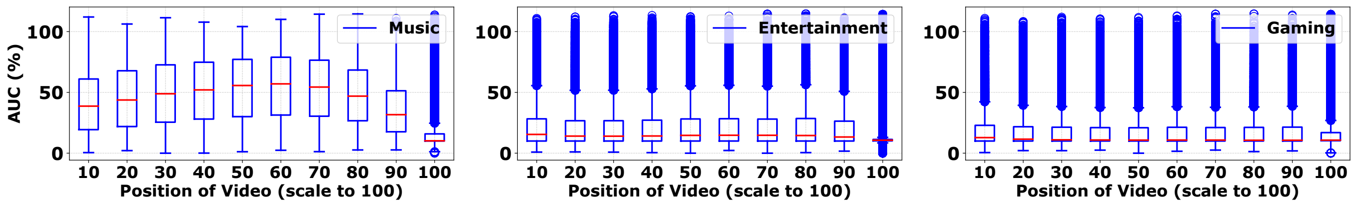

Figure 5. The AUC from the YouTube dataset by category: (left) music, (middle) entertainment, and (right) gaming.

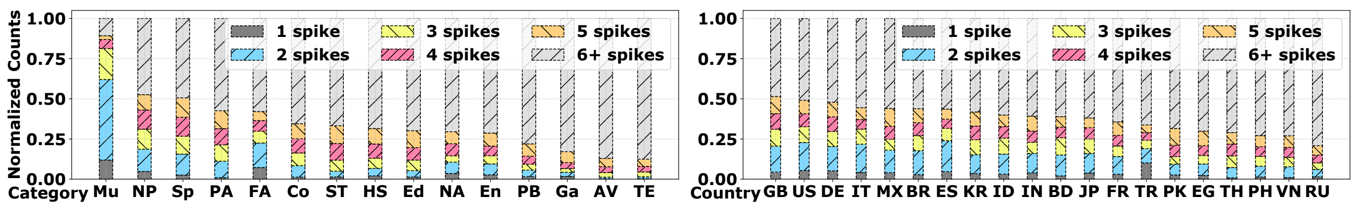

Figure 6. Number of spikes per video from the YouTube dataset by category (left) and country (right).

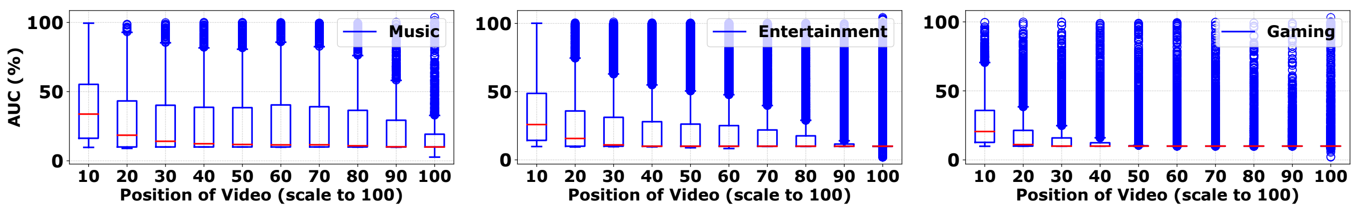

Figure 7. The AUC from YouTube Shorts by category: (left) music, (middle) entertainment, and (right) gaming.

Category [1–8]	# Videos	Category [9–16]	# Videos
Gaming (Ga)	2,772	Animal (An)	1,861
Life (Li)	2,526	Sports (Sp)	1,750
Food (Fo)	2,403	Entertainment (En)	1,736
Knowledge (Kn)	2,265	Car (Ca)	1,676
Douga (Do)	2,248	Fashion (Fa)	1,439
Cinephile (Ci)	2,210	Dance (Da)	1,317
Music (Mu)	2,136	Guochuang (Gu)	1,009
Technology (Te)	1,932	Kichiku (Ki)	857

Table 3. The distribution of 30,137 Bilibili by video category.

This suggests that ABR should be optimized for individual video types (*e.g.,* short videos), as proposed by Dashlet [17].

2.3 Bilibili Dataset Collection and Analysis

As YouTube does not operate in Mainland China, we evaluated Bilibili to represent the audience in China, where Bilibili has 332.6 million average monthly active users [23].

Dataset Collection. As in § 2.1, the Bilibili dataset consists of two parts: (i) video metadata, and (ii) audience retention curve. Bilibili provides audience retention curves using an SVG with class `bpx-player-pbp`. We query the daily popular videos as listed on the Bilibili website [3] across 16 categories: gaming (Ga), life (Li), food (Fo), knowledge (Kn), douga (animation and motion picture) (Do), cinephile (Ci), music (Mu), technology (Te), animal (An), entertainment (En), sports (Sp), car (Ca), fashion (Fa), dance (Da), guochuang (domestically-made animations) (Gu), and kichiku (remixed videos) (Ki); giving us a dataset of 30,137 videos (10/28/2022–05/22/2023, as shown in Table 3). However, since the daily popular videos are not always newly published, we do not analyze the duration of popularity for Bilibili.

2.3.1 Video Length

Figure 9 shows the normalized counts of Bilibili (≤ 60 seconds), Bilibili (≤ 10 minutes), Bilibili (10+ minutes). First, over 75% of Bilibili videos are less than 10 minutes. For video remixes (Ki), very few videos are longer than 10 minutes. The distribution of video length on Bilibili differs greatly from YouTube, where over 45% of videos on YouTube are greater than 10 minutes except music (Figure 4). The distributions of music video length on both YouTube and Bilibili are very similar. We conjecture that YouTube videos tend

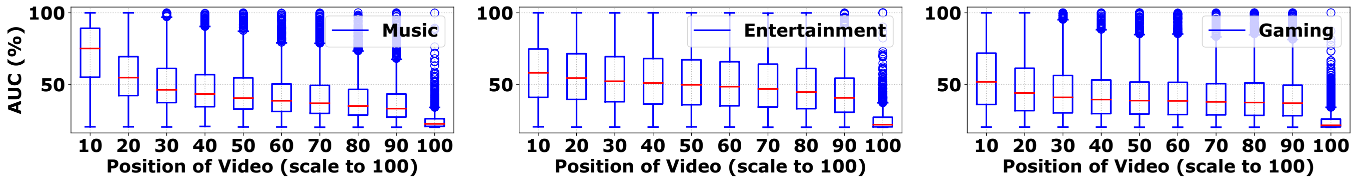

Figure 8. The AUC from the Bilibili dataset by category: (left) music, (middle) entertainment, and (right) gaming.

to be longer because YouTube's ad placement and revenue sharing policies.

Takeaway. Over 75% of Bilibili popular videos are less than 10 minutes, whereas on YouTube, over 45% of videos on YouTube are greater than 10 minutes, possibly due to YouTube policies favoring longer videos.

2.3.2 AUC

As in § 2.2.3, we find that audience retention curves of videos in some categories (*e.g.*, music) are different from YouTube dataset. In particular, Figure 8 shows Bilibili has much higher AUC in the first half, suggesting that the Bilibili viewers tend to abandon the videos earlier than YouTube viewers. We believe that these differences are driven in large part by the incentives created by the platforms. Since YouTube requires videos to be a certain length before allowing mid-roll advertising, creators tend to create longer videos. Second, since Bilibili does not require viewers to watch advertisements at the beginning of each video, the time cost of switching between videos is significantly lower than YouTube. Finally, Bilibili creators derive most of their revenue from sponsorship from the platform based on user preference [2], which does not have specific retention requirements. User preference is mostly based on "I like this" ratings rather than total watch time. Thus, some makers tend to put the best part at the beginning to directly earn user preference.

Takeaway. The Bilibili dataset differs significantly from the YouTube dataset and shows increased early abandonment across all video categories, including music.

2.3.3 Spikes of Retention Curve

Unlike YouTube videos, over 35% of Bilibili videos have less than 2 spikes, as shown in Figure 10, possibly due to early abandonment. Also, this indicates that the Chinese audience tends to replay videos less often in categories such as video remixes (Ki) and music (Mu). We believe that the shorter video lengths of Bilibili result in reduced replayability. In

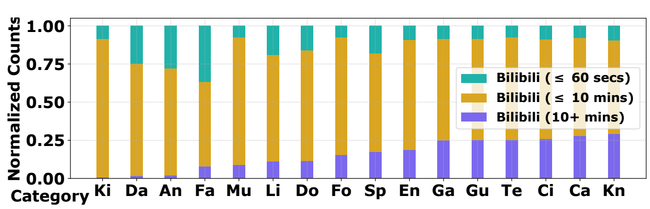

Figure 9. Length of videos (Bilibili dataset) by category.

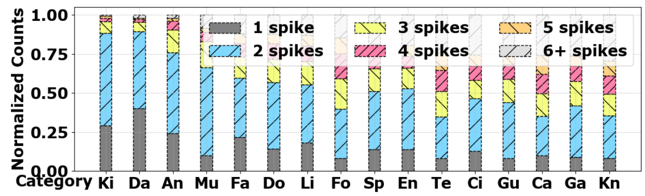

Figure 10. Number of spikes per video from the Bilibili dataset by category.

longer videos, interesting bits can be scattered throughout, resulting in more spikes than shorter videos. The lower production quality on Bilibili may also be a factor that reduces replayability.

Takeaway. Over 35% of Bilibili videos have less than 2 spikes, whereas less than 20% of YouTube videos have less than 2 spikes.

Same Videos on YouTube and Bilibili. To show the impact of different platforms on the same video, we selected 217 videos that are on both YouTube and Bilibili, and show the AUC in Figure 11. It is clearly that even for the same videos, early abandonment is much more prevalent on the Bilibili platform.

3 Discussion And Application

Skipping Boring Moments. To explore why users skip certain video segments, we identified the "Most Boring" segment, which has the same length as the "Most Replayed"

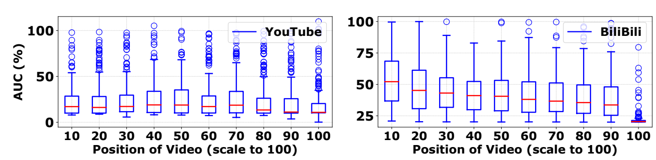

Figure 11. The AUC from the same videos by platform: (left) YouTube, and (right) Bilibili.

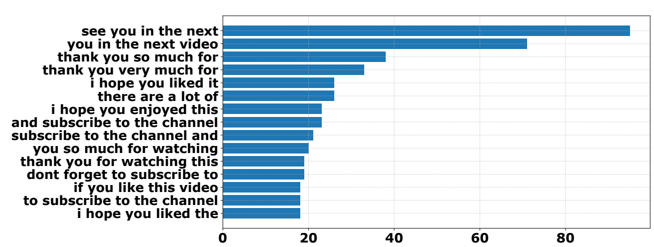

Figure 12. Top-15 occurrence of 5-grams phrases extracted from transcripts during the "Most Boring" moment.

segment but the lowest audience retention curve area. We analyzed 39,008 randomly downloaded YouTube videos and their transcripts, using Whisper [18] for transcription if unavailable. Our analysis using 5-grams phrases, depicted in Figure 12, reveals that users often skip segments containing closing remarks ("See you in the next"), expressions of gratitude and audience engagement ("thank you so much for"), and subscription prompts ("subscribe to the channel and"). These segments are typically skipped because they are considered less engaging or informative by the audience.

Video Highlight Detection. Utilizing audience retention curves for video highlight detection is more effective than human-labeled scores [11, 14], as they authentically reflect diverse audience interests across different video moments. Our dataset, encompassing various video categories, presents new challenges for accurate highlight detection.

4 Related Work

Predicting YouTube video popularity presents intricate challenges. Yu et al. [7] utilized power-law models for their lifecycle analysis of 172, 841 videos, identifying music as the most popular genre. They noted that the top 5% popular videos typically experience several phases, in contrast to news videos that often show extended power-law decays. Differing from their approach, we focus on audience retention curves to assess per-segment popularity, providing deeper insights into user engagement than simple view count analysis. We contend that high view counts do not always correlate with longer watch times, making audience retention curves a more precise tool for gauging popularity. YouSlow [6] explored the impact of rebuffering on user engagement, primarily from a QoE perspective. Similarly, Park et al. [8] investigated the correlation between user engagement, defined as total watch time, and video performance metrics such as likes per view ratio, negative comment sentiment, and overall view counts. In contrast, our research delves into analyzing video popularity through detailed audience retention curves, offering a comprehensive perspective on user viewing behaviors and preferences.

5 Conclusion

In our work, utilizing 229,178 audience retention curves from YouTube and Bilibili, reveals diverse user behaviors and viewing preferences across countries and platforms. Key findings include the popularity of gaming and entertainment in specific countries and Bilibili's trend towards shorter videos with early abandonment. This dataset, showcasing varied user engagement, not only aids in refining QoE metrics and ABR techniques but also has potential applications in video understanding and analysis.

References

[1] Baiju Muthukadan. 2011. Selenium with Python. https://selenium-python.readthedocs.io. (2011). Accessed April 14, 2023.
[2] Bibibili Content Creator Center. 2018. Bilibili Creation Incentive Program Rules. https://www.bilibili.com/read/cv173108/?from=search&spm_id_from=333.337.0.0. (2018). Accessed Sepetember 9, 2023.
[3] Bilibili Inc. 2023. Bilibili popular video rank. https://www.bilibili.com/v/popular/rank/all. (2023). Accessed April 25th, 2022.
[4] Cisco Networks. 2019. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022. http://media.mediapost.com/uploads/CiscoForecast.pdf. (2019). Accessed April 23, 2023.
[5] Hongzi Mao et al. 2017. Neural adaptive video streaming with pensieve. In *Proceedings of the ACM SIGCOMM'17*. 197–210.
[6] Hyunwoo Nam et al. 2016. QoE matters more than QoS: Why people stop watching cat videos. In *Proceedings of the IEEE INFOCOM'16*. IEEE, 1–9.
[7] Honglin Yu et al. 2015. The lifecyle of a youtube video: Phases, content and popularity. In *Proceedings of the AAAI ICWSM'15*, Vol. 9. 533–542.
[8] Minsu Park et al. 2016. A data-driven study of view duration on youtube. In *Proceedings of the AAAI ICWSM'16*, Vol. 10. 651–654.
[9] Pan Du et al. 2006. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *bioinformatics* 22, 17 (2006), 2059–2065.
[10] S Shunmuga Krishnan et al. 2012. Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. In *Proceedings of the ACM IMC'12*. 211–224.
[11] Taivanbat Badamdorj et al. 2021. Joint visual and audio learning for video highlight detection. In *Proceedings of the IEEE ICCV'21*. 8127–8137.
[12] Te-Yuan Huang et al. 2014. A buffer-based approach to rate adaptation: Evidence from a large video streaming service. In *Proceedings of the ACM SIGCOMM'14*.
[13] Xiaoqi Yin et al. 2015. A control-theoretic approach for dynamic adaptive video streaming over HTTP. In *Proceedings of the ACM SIGCOMM'15*. 325–338.
[14] Ye Liu et al. 2022. Umt: Unified multi-modal transformers for joint video moment retrieval and highlight detection. In *Proceedings of the IEEE CVPR'22*. 3042–3051.
[15] Yale Song et al. 2015. Tvsum: Summarizing web videos using titles. In *Proceedings of the IEEE CVPR'15*. 5179–5187.
[16] Yi Sun et al. 2016. CS2P: Improving video bitrate selection and adaptation with data-driven throughput prediction. In *Proceedings of the ACM SIGCOMM'16*. 272–285.
[17] Zhuqi Li et al. 2023. Dashlet: Taming Swipe Uncertainty for Robust Short Video Streaming. In *USENIX Symposium on NSDI'23*. 1583–1599.
[18] OpenAI. 2023. OpenAI Whisper. https://github.com/openai/whisper. (2023). Accessed April 15, 2023.
[19] Sandvine. 2023. Global Internet Phenomena Report. https://www.prnewswire.com/news-releases/sandvines-2023-global-internet-phenomena-report-shows-24-jump-in-video-traffic-with-netflix-volume-overtaking-youtube-301723445.html. (2023). Accessed April 23, 2023.
[20] Scipy. 2023. scipy.signal.find_peaks. https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html. (2023). Accessed April 15, 2023.
[21] Iraj Sodagar. 2011. The mpeg-dash standard for multimedia streaming over the internet. *IEEE multimedia* 18, 4 (2011), 62–67.
[22] Statista. 2022. Leading countries based on YouTube audience size as of April 2022. https://www.statista.com/statistics/280685/number-of-monthly-unique-youtube-users/. (2022). Accessed December 7, 2022.
[23] Wikipedia contributors. 2023. Bilibili/User Characteristics/Users. https://en.wikipedia.org/wiki/Bilibili. (2023). Accessed April 24, 2023.
[24] YouTube. 2022. YouTube Data API. https://developers.google.com/youtube/v3/docs. (2022). Accessed April 7, 2023.