The impact of interactive features in music applications

Xiangyu Jiang
Clemson University
Clemson, SC, USA
xiangyj@clemson.edu

ABSTRACT

The interface design and functionality of music applications have a considerable impact on the user experience; consequently, it is of practical significance to study the attention distribution and preference responses of users when using the applications in order to optimize the application design. However, research on the attention distribution and actual responses of music application users is still limited. In this paper, we propose to investigate the attention distribution and preferred responses of music application users using eye-tracking techniques. By analyzing this data, we will explore the effect of different application features and elements, such as interactive features and UGC (user-generated content), on user attention and preference responses in order to optimize the user experience in the future design of music applications.

CCS CONCEPTS

• **Human-computer interaction** → **Eye tracking**.

KEYWORDS

Eye Tracking, Human-Computer Interaction, Attention

ACM Reference Format:
Xiangyu Jiang. 2018. The impact of interactive features in music applications. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Due to the rise in popularity of personal mobile devices, people spend significantly more time using electronic products and applications, while offline social opportunities have decreased, which may lead to an increase in mental stress. In conjunction with the effects of the epidemic and home office education, this pressure may increase. Although there have been numerous studies on the stress-relieving effects of music, little is known about the attention distribution and actual preference response of users when using music applications. Since a long time, music application developers have focused on the significance of song recommendation algorithms, whereas scientists have focused on the influence of different music frequency tones on individual emotions. Although different music can produce different effects, many studies on music application software emphasize user interface design strategies while ignoring the influence of social attributes and UGC. It is difficult to study and determine the role of interactive features in music software for stress reduction.

Therefore, this paper proposes to use eye tracking technology to investigate the attention distribution and user experience of music application users in different functional modules. We will compare and analyze the attention distribution and emotional response of users of three popular music apps (such as NetEase Cloud Music, Apple Music and Spotify). The to-be-analyzed functions include music playback and lyrics. The algorithm suggests similar musical characteristics and tastes. Users recommend music, like comments, and participate in other social activities. By analyzing the time and attention spent by users on different functional modules, we will investigate the impact of different application functions and elements on user preferences in order to optimize the user experience in the future design of music applications and to increase user retention and user stickiness.

1.1 Music recommendation

Music has been a source of entertainment for humans since ancient times, and existing research has shown that different types and frequencies of music can have different effects on people's emotional and psychological states. The ability of music of specific frequencies to relieve emotional stress[11] has been widely studied and applied. Currently, many music apps using recommendation algorithms based on the emotional value and contextual attributes of music have been shown to have good feedback[6]. However, not much research has been done on user-centered music recommendations.

1.2 User interface and software function

Research shows that the user interface is the most important component of an application and that interface design directly affects the user experience[1, 2]. Interface design has been extensively researched, and simple and convenient software operations can improve the user experience of an application[7]. In addition, a cleaner and clearer interface design can lead to a better user experience of the product. Recent studies have combined machine learning with existing interface designs[9], using algorithms to generate new interfaces. However, these studies focus more on the functional implementation and beautification of interfaces.

1.3 Online social and UGC

The prevalence of electronic devices and the impact of the epidemic have reduced offline social interactions, which can increase people's isolation[8]. Proper online social interaction can help alleviate stress[4]. Of the three apps we compared, NetEase Cloud Music has

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/XXXXXXX.XXXXXXX

commenting, liking and sharing features for songs, while Apple Music and Spotify do not. Customizing song playlists as user-generated content can increase user engagement and content sharing, which can boost the app and improve the user experience.

1.4 Research Goals and Questions

This research seeks to figure out which features of music applications users pay the most attention to, whether different features elicit different emotional responses from users, and whether users prefer system recommendations or recommendations from other users. There are three specific research targets:

- To examine the amount of time users spend on the various features of music applications, we have no hypothesis that social features will occupy more of the users' time.
- To compare user responses to system-recommended contextually similar songs with user-recommended songs.
- To gain insight into how users utilize the application and enhance it to meet their requirements.

2 METHODS

2.1 Participants

The study will recruit 6 grad students who have the ability to read and speak English. Participants should have different backgrounds, e.g., computer science, education, and mathematics. Participants are recruited voluntarily

2.2 Apparatus

Participants' eye movements will be recorded using a Gazepoint GP3 eye tracker device with a visual accuracy of 0.5-1.1 degrees and a sampling rate of 60 Hz (one sample every 16 ms). The device will have a total 30 cm range of depth movement. Stimulus presentation will be run by a music software program on a PC and can be freely controlled by the participant.

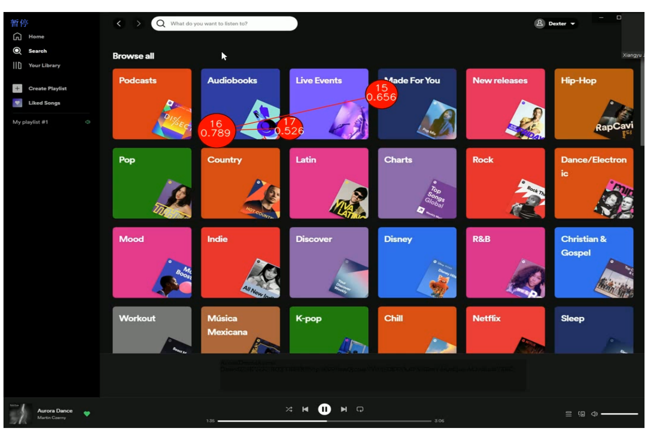

Figure 1: GazePoint3 eyetracking example

2.3 Yolo

This paper proposes an idea of applying the state-of-the-art YOLO (You Only Look Once) algorithm for automated detection of user eye tracking targets. YOLO, originally designed by Joseph Redmon [10], is an attractive CNN-based algorithm for object detection, classification, and localization in images and videos [3]. During the past years, YOLO kept improving with some new algorithms to optimize the computing speed and achieve better performance. The fifth version of YOLO (YOLO-v5) was introduced by Glenn Jocher in June 2020 [5]. This model significantly reduced the model size (YOLO-v4 on Darknet had 244MB size whereas YOLO-v5 smallest model is 27MB). YOLO-v5 also claimed a higher accuracy and more frames per second than all previous versions.

2.4 Research design

To compare the impact of system recommendation algorithms and social features on the user experience, we will use a two-factor design with song recommendation and interactive elements for the research design. The participants will be asked to use the three aforementioned music applications under two conditions: one in which they use only the system's song recommendation function, and the other in which they also use the social function. We will record the amount of time and frequency with which they use the applications, as well as collect their evaluation and feedback on the various functions suggested. We hope this design will shed light on the relative impact of the various features on the user experience, while preserving the naturalist's daily use. Our research will contribute to the enhancement of our application in order to better meet user requirements and enhance the user experience.

2.5 Measures

By calculating the amount of time participants spend in different areas, eye movement measurement will use dwell time as a measure of the participants' attention level for different recommended songs/interactive elements. We will create regions of interest using a semantic segmentation technique based on the YOLO model (AOI). AOI will be generated using a model of image semantic segmentation that has been trained using the training dataset. To ensure the accuracy and validity of eye movement measurements, we will manually label the segmentation of software interface functions to obtain the training dataset for the semantic segmentation model.

We will evaluate the impact of the system's recommendation algorithm and social features on user experience using these two metrics.

2.6 Procedure

Each participant will be evaluated independently. Participants will be given the following instructions: "Two of your favorite songs will now be played, and you will be asked to use the song recommendation function while these songs are playing; you can choose between the system recommendation or the user recommendation function, during which you can interact with the comments of other users," and they will sit in a comfortable position. Each participant will receive an eye-tracking calibration from Gazepoint GP3 eye tracker prior to viewing the video. After listening to each song, a 7-point rating preference task will appear in the headset, followed by additional questions at the conclusion of the viewing process. This will help participants remain focused while viewing. The total duration of image viewing will be approximately 10 minutes. After

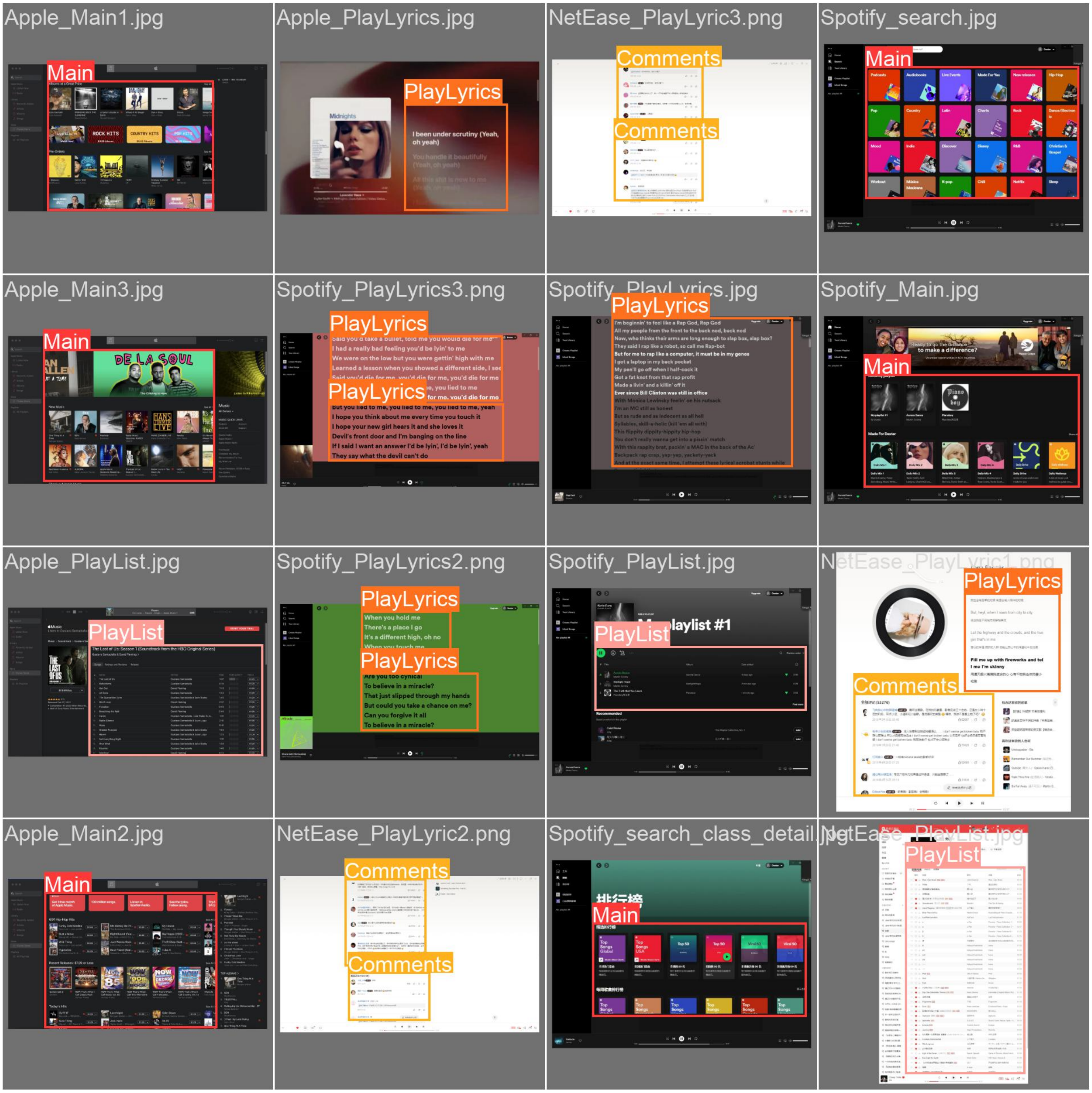

Figure 2: Yolo labeled data

viewing all videos, participants will be asked to complete a qualitative questionnaire regarding their preferences for the recommended songs and their reactions to the commenting feature.

3 EXPERIMENT, RESULTS AND DISCUSSION

3.1 Experiment Process

The experimental design of this study consisted of three main steps: YOLO object detection, Gazepoint eye tracking, and data analysis. In the first step, we used the YOLO algorithm to perform object detection on the input images and output the results in txt format.

Figure 3: Yolo predict result

For the second step, Gazepoint eye tracking was utilized to obtain user annotation data, which allowed us to collect accurate gaze direction and fixation data during the experiment. In the final step of the experiment, we integrated and analyzed the collected data from both YOLO object detection and Gazepoint eye tracking to draw meaningful conclusions.By combining the information from these two sources, we were able to gain a more comprehensive understanding of how users interacted with the objects in the images.

3.2 Data Pre-Processing

In order to address potential issues with inaccurate sampling and exceeding critical values in the raw data, we preprocessed our data before starting the experiment. Specifically, for the eye tracking data, we used a one-dimensional mean filter to preprocess the data and reduce errors caused by unstable sampling. In order to better present the smoothed results, we used a mean filter with a size of 50 to process the data, as shown in Fig. 5., making the data smoother and easier for subsequent calculations.

3.3 Data Analysis

In the data analysis section, we first read and load the YOLO output txt files containing detection results for each frame into a DataFrame. Since there may be multiple targets in each frame, we read all the detection results for each target in the file and store them together in the DataFrame. Then, we read the eye tracking data and calculate the frames per second (fps) based on the total time of eye tracking and the number of frames in the video. We use the fps and the current time to calculate the corresponding frame number for the current time point. We then compare the position data of the target detection in the current frame with the user's gaze data. If the user's gaze point is within the range of the target detection, the gaze point for the current time is recorded as the category of the target detection. Finally, we comprehensively analyze the classification data to calculate the total gaze time for each target category.

3.4 Result Analysis

Based on the analysis of experimental data, we found that users spend more time using applications with comment and user-generated content features than those without such features (about 60 seconds vs. 40 seconds, respectively). After removing the time spent on failed classifications and background fixation, we found that users' attention on comment functions accounted for approximately one-third of their total usage time. From these results, we can conclude that user-generated content and social features play a significant role in the time users spend on music applications. Adding user-generated content and social features to music applications can increase usage time, improve the effectiveness of the software, and enhance the user experience.

4 CONCLUSION

The results of our data analysis revealed key insights into user behavior, which can be used to inform future research and development in this field. Overall, the integration of YOLO object detection and Gazepoint eye tracking provided a powerful tool for investigating human-object interactions and yielded valuable findings that have important implications for a range of applications, from computer vision to human-computer interaction.

5 FUTURE WORK

Based on the current work, there are several potential future directions for further research:

Testing on a larger and more diverse dataset: While the current dataset used in this work is sufficient for initial analysis, expanding the dataset to include more diverse scenarios and populations would provide a more comprehensive understanding of the system's performance.

Incorporating machine learning algorithms: Machine learning algorithms, such as deep neural networks, could be integrated into the system to improve accuracy and efficiency in detecting and analyzing eye movement data.

Integration with other physiological measures: Combining eye movement data with other physiological measures, such as heart rate or skin conductance, could provide additional insights into cognitive and emotional states of the user.

Application in real-world settings: The current work was conducted in a controlled laboratory setting. Future research could explore the application of the eye tracking system in real-world settings, such as driving or sports performance, to evaluate its effectiveness and practicality.

Comparison with other eye tracking systems: Comparing the performance of the developed eye tracking system with other existing eye tracking systems could provide a better understanding of its strengths and weaknesses and inform further improvements.

REFERENCES

[1] Murilo C. Camargo, Rodolfo M. Barros, and Vanessa T. O. Barros. 2018. Visual Design Checklist for Graphical User Interface (GUI) Evaluation. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (Pau, France) *(SAC '18)*. Association for Computing Machinery, New York, NY, USA, 670–672. https://doi.org/10.1145/3167132.3167391

[2] Chunyang Chen, Sidong Feng, Zhengyang Liu, Zhenchang Xing, and Shengdong Zhao. 2020. From Lost to Found: Discover Missing UI Design Semantics through Recovering Missing Tags. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 123 (oct 2020), 22 pages. https://doi.org/10.1145/3415194

[3] Kamya Desai, Siddhanth Parikh, Kundan Patel, Pramod Bide, and Sunil Ghane. 2020. Survey of object detection algorithms and techniques. In *Cybernetics, Cognition and Machine Learning Applications: Proceedings of ICCCMLA 2019*. Springer, 247–257.

[4] John Gleeson, Reeva Lederman, Peter Koval, Greg Wadley, Sarah Bendall, Sue Cotton, Helen Herrman, Kingsley Crisp, and Mario Alvarez-Jimenez. 2017. Moderated Online Social Therapy: A Model for Reducing Stress in Carers of Young People Diagnosed with Mental Health Disorders. *Frontiers in Psychology* 8 (2017). https://doi.org/10.3389/fpsyg.2017.00485

[5] Glenn Jocher, Alex Stoken, Jirka Borovec, STAN Christopher, and Liu Changyu Laughing. 2021. ultralytics/yolov5: v4. 0-nn. SiLU () activations, Weights & Biases logging, PyTorch Hub integration. *Zenodo* (2021).

[6] Dongmoon Kim, Kun-su Kim, Kyo-Hyun Park, Jee-Hyong Lee, and Keon Myung Lee. 2007. A music recommendation system with a dynamic k-means clustering algorithm. In *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*. 399–403. https://doi.org/10.1109/ICMLA.2007.97

[7] Kati Kuusinen and Tommi Mikkonen. 2013. Designing User Experience for Mobile Apps: Long-Term Product Owner Perspective. In *2013 20th Asia-Pacific Software Engineering Conference (APSEC)*, Vol. 1. 535–540. https://doi.org/10.1109/APSEC.2013.77

[8] Edward C. Malthouse, Bobby J. Calder, Su Jung Kim, and Mark Vandenbosch. 2016. Evidence that user-generated content that produces engagement increases purchase behaviours. *Journal of Marketing Management* 32, 5-6 (2016), 427–444. https://doi.org/10.1080/0267257X.2016.1148066

[9] Tam The Nguyen, Phong Minh Vu, Hung Viet Pham, and Tung Thanh Nguyen. 2018. Deep Learning UI Design Patterns of Mobile Apps. In *Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results* (Gothenburg, Sweden) *(ICSE-NIER '18)*. Association for Computing Machinery, New York, NY, USA, 65–68. https://doi.org/10.1145/3183399.3183422

[10] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

[11] Nechama Yehuda. 2011. Music and stress. *Journal of Adult Development* 18 (2011), 85–94.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

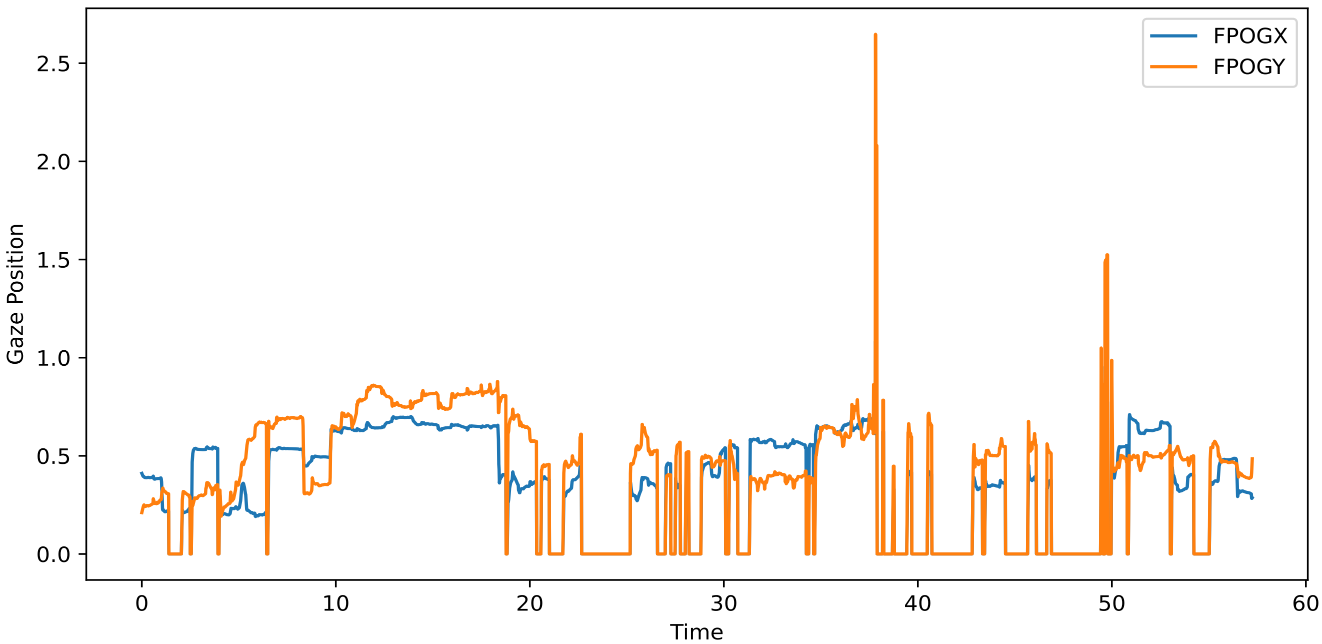

Figure 4: Video0 raw eye tracking data in 1D

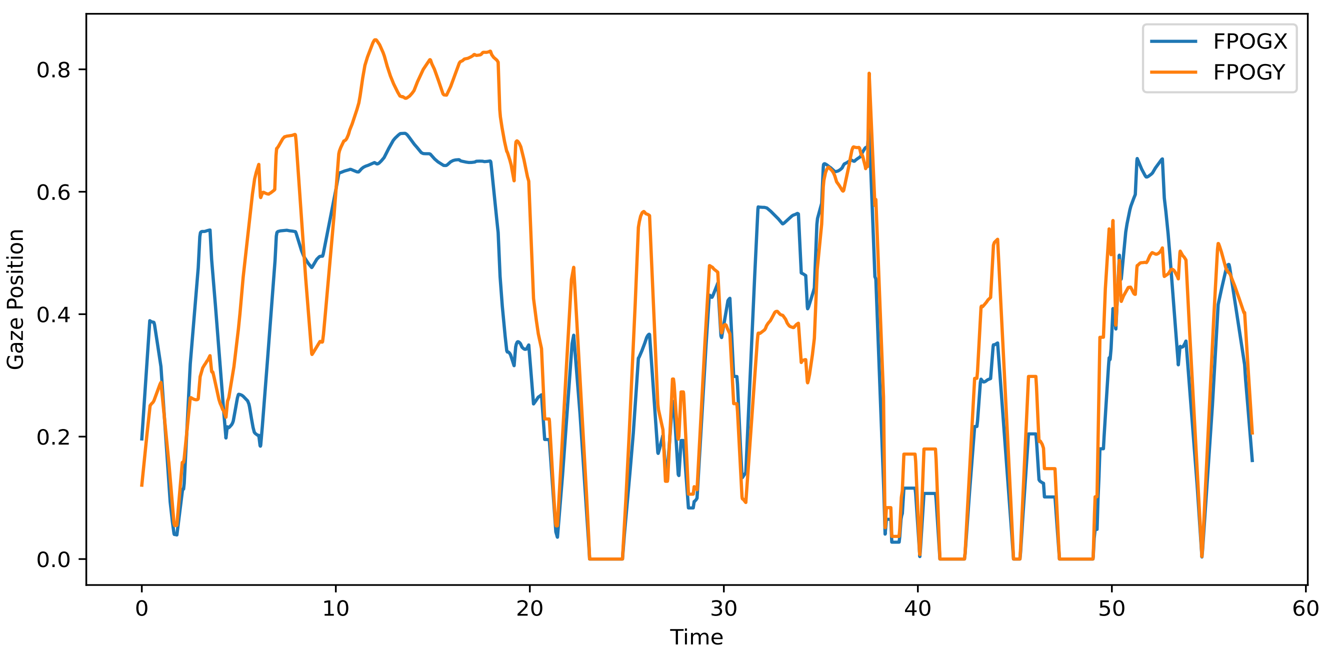

Figure 5: Video0 smoothed eye tracking data in 1D

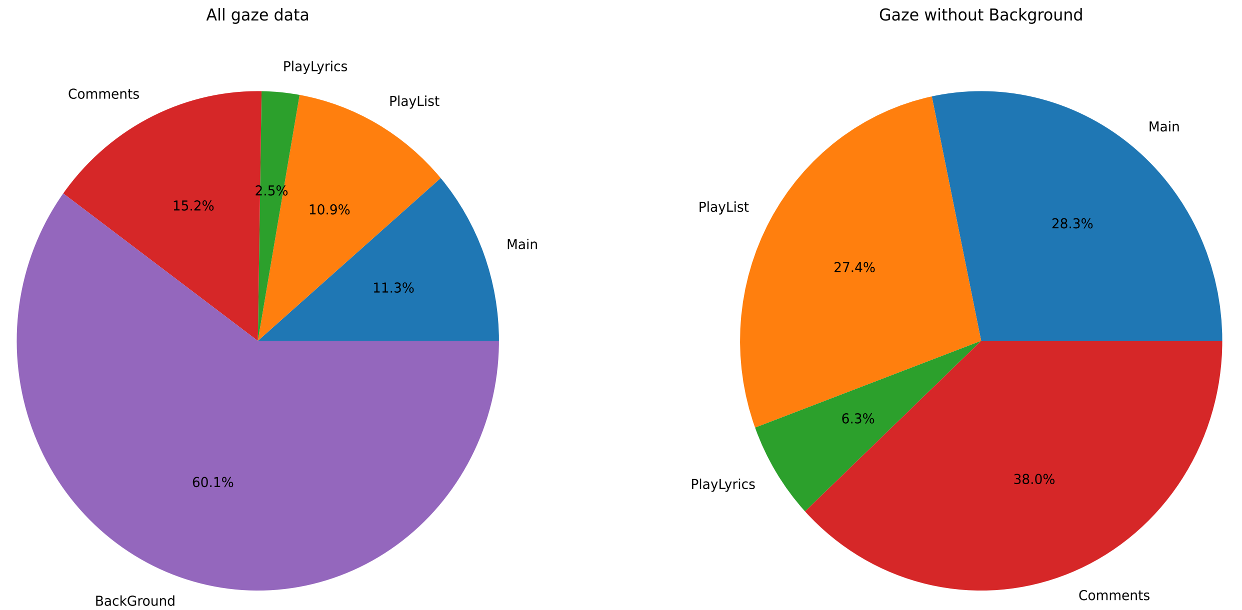

Figure 6: Video0 Gaze Data Pie Chart

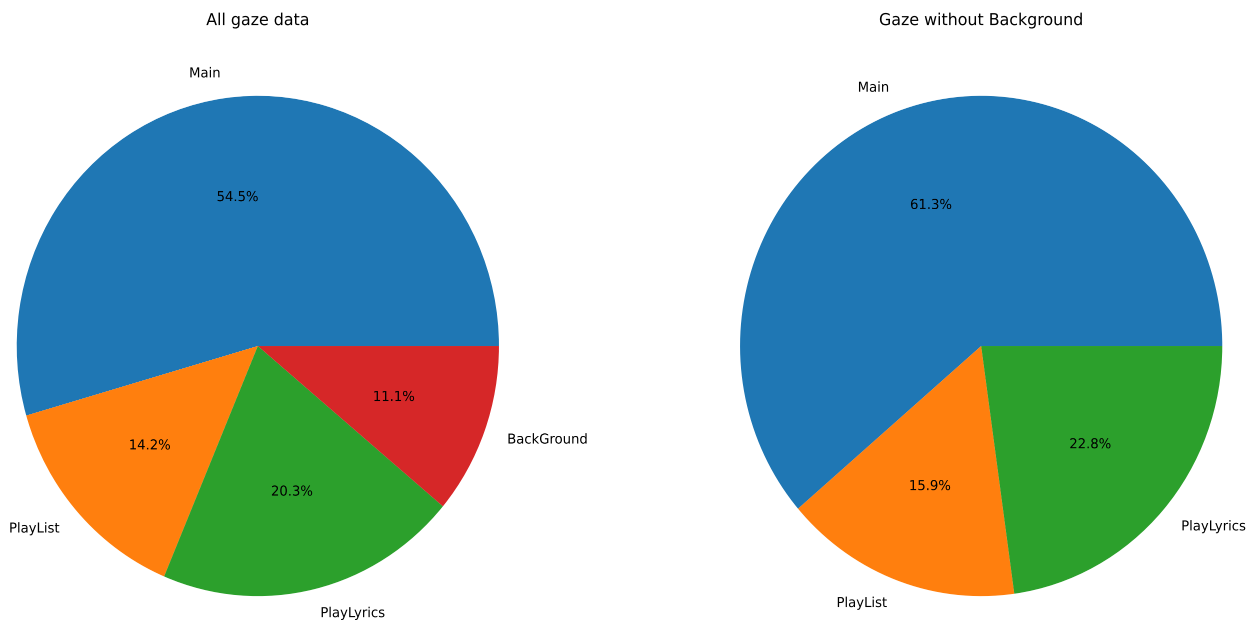

Figure 7: Video1 Gaze Data Pie Chart