Waiting along the path: how browsing delays impact the QoE of music streaming applications

Anika Seufert, Ralf Schweifler, Fabian Poignee, Michael Seufert, Tobias Hoßfeld

Angaben zur Veröffentlichung / Publication details:

Seufert, Anika, Ralf Schweifler, Fabian Poignee, Michael Seufert, and Tobias Hoßfeld. 2022. "Waiting along the path: how browsing delays impact the QoE of music streaming applications." In 2022 14th International Conference on Quality of Multimedia Experience (QoMEX), 5-7 September 2022, Lippstadt, Germany, edited by Jan-Niklas Voigt-Antons, Oliver Hohlfeld, Florian Metzger, and Raimund Schatz, 1–6. Piscataway, NJ: IEEE.
https://doi.org/10.1109/qomex55416.2022.9900883.

Waiting along the Path: How Browsing Delays Impact the QoE of Music Streaming Applications

Anika Seufert, Ralf Schweifler, Fabian Poignée, Michael Seufert, Tobias Hoßfeld
University of Würzburg, Chair of Communication Networks,
Würzburg, Germany
{anika.seufert|ralf.schweifler|fabian.poignee|michael.seufert|tobias.hossfeld}@uni-wuerzburg.de

***Abstract*—Streaming has become the dominant source of media consumption, which not only applies to the widely researched field of video streaming, but also to music streaming. Here, previous studies so far have only researched the impact of streaming aspects, such as stalling events or initial loading times, on the QoE of music streaming. However, when using a music streaming application, users are already facing waiting times along the click path before they can start the actual streaming. These waiting times are caused by browsing delays, e.g., during searching for songs or scrolling through playlists, and can potentially deteriorate the QoE of the music streaming application. In this work, we conduct an online QoE study to quantify the impact of these browsing delays with the support of an emulated mobile music streaming web app. We found that browsing delays have no impact on the music streaming QoE, which shows that users are able to clearly distinguish between the two main functionalities of such apps, namely, browsing and streaming. However, browsing delays significantly reduce the QoE of the entire music streaming application, to a similar extent as if QoE degradations happen during the actual streaming. This shows that both browsing and streaming are equally important and have to be considered when designing music streaming applications.**

I. Introduction

Music is an important part of many people's everyday lives, and in today's digital age, music consumption is dominated by web-based applications, which stream music to their customers on demand over the public Internet. Apart from free basic services, which often include the presentation of advertisements to the users, music streaming providers typically offer paid, ad-free premium subscriptions ranging between \$5 and \$15 per month [1]. These music streaming apps are increasingly popular. For example, the music streaming provider Spotify reported 356 million monthly active users in the first quarter of 2021, which is an increase of 24% within a single year and includes 158 million premium subscribers [2]. Thus, music streaming is a competitive market, which had a value of \$12.8 billion in 2019, and is estimated to reach a total market share of \$24.7 billion until 2027 [3].

To ensure a high level of satisfaction, service providers are interested in identifying and mitigating factors that negatively influence the Quality of Experience (QoE) [4] of their customers while using their app and listening to music by implementing a QoE-aware design [5]. There are already a few studies on music streaming application QoE, e.g., [6,7], but these studies have mainly investigated the impact of streaming-related aspects, such as stalling events and initial delay. However, when using a music streaming application, users are already facing waiting times along the click path before they can start the actual streaming. These waiting times are caused by browsing delays within the application, which are the result of fluctuating and/or bad network conditions [8], and can potentially deteriorate the QoE of the music streaming application. Thus, to holistically assess the QoE of music streaming applications, both main functionalities of such apps have to be considered, namely, browsing and streaming. Since streaming applications are often used on mobile devices, their users are especially prone to changing network conditions.

In this work, we close the gap to holistically understanding the QoE of music streaming applications by investigating the impact of browsing delays along the click path of the users. In particular, these are a search delay when waiting for the result of a search bar query, a delay when opening a playlist, and a delay when loading additional titles while scrolling through a playlist. For this, we developed a web application for mobile music streaming, which emulates the look and feel of current music streaming mobile apps. Additionally, it acts as a QoE study framework which is able to insert those waiting time conditions and corresponding QoE questionnaires. We used our mobile application to conduct an online study on the QoE of music streaming applications, which aims to answer the following research questions:

RQ1: Do browsing delays impact the QoE of subsequent music streaming?
RQ2: Do browsing delays influence the overall music streaming app experience?

For this, the remainder of this work is structured as follows. In Section II, background information on QoE and methodology on how to conduct QoE studies is given. Furthermore, related work on music streaming and web browsing QoE is discussed. The design of our mobile music streaming app, the conducted QoE study, as well as the filtering applied to identify and remove unreliable participants are outlined in Section III. Section IV describes the collected data set and presents study results to answer the above stated research questions. Finally, Section V concludes this work and gives an outlook to future work.

II. Background and Related Work

Considering web-based multimedia services, the subjective experience of the end user is highly important for network and service providers, which are interested in quantifying their users' QoE [4]. To collect QoE ratings, subjective user studies are conducted for which many design and evaluation guidelines already exist, e.g., [9]–[12]. In addition to lab studies, crowdsourcing and remote user studies are increasingly popular to be used for scientific purposes as they allow for a large and diverse group of international participants [13]–[15]. The main pitfall for this type of studies is the unsupervised study execution of online and remote participants, which needs to be considered during study design as well as result evaluation [16,17]. In particular, this means that unreliable participants need to be identified and filtered from the study results. Conducting remote QoE studies on mobile devices, adds another aspect which has to be controlled, hence these studies are rarely performed.

With respect to QoE studies for music streaming applications, only few works so far have focused on this increasingly popular application type, and to the best of our knowledge, we are the first to conduct a QoE study for music streaming on mobile devices. Sackl et al. [6] conducted a desktop subjective user study to investigate the impact of initial delay and stalling on QoE in music and video streaming and found that users are generally much less tolerant to temporal impairments in the case of music streaming than in the case of video. Schwind et al. [18] investigated the behavior of Spotify under different network conditions and measured QoE-related streaming key performance indicators (KPIs) like stalling and initial delay. In [8], Schwind et al. examined the Spotify application more holistically. For this, they looked at the entire click path of users in the mobile app, from logging in via searching to listening to a song. However, instead of conducting a subjective QoE study, the process was emulated in an automated testbed and navigation times and streaming KPIs were measured under different network conditions. The streaming KPIs were then mapped to QoE scores using the QoE model obtained by Sackl et al. [6]. Nevertheless, as the model only considers the streaming aspects of music streaming applications, the impact of browsing delays is not reflected in their results.

Long browsing delays are a major annoyance on the Internet, and thus, have received a lot of attention by the QoE community. Starting from early results on response times [19], the page load time is considered the most important QoE factor, e.g., in [20]. The resulting web QoE models typically show a logarithmic or exponential relationship between QoE and waiting time [21]. In addition to time-based metrics, also integral metrics were proposed, such as Speed Index [22], which measures visual progress during the page loading.

Although browsing delays are common in many web-based applications including streaming applications, their impact on the QoE of streaming applications has received little attention yet. Closest to our work, Seufert et al. [23] investigated the impact of small page load times ($0.5\,\mathrm{s}$-$4\,\mathrm{s}$) on the QoE of subsequent video streams. They found that for such small delays, the page load time could not deteriorate the video streaming QoE compared to the conditions without browsing delay. As also equally short initial delays did not deteriorate the video streaming QoE, they concluded that users might be accustomed to small delays for web-based services, and that short waiting would not negatively impact the QoE of Internet services, especially if users anticipated the service. However, browsing delays were introduced only immediately before the actual service started, and users were only asked about the resulting streaming QoE. Thus, it remains unclear what impact browsing delays have on the overall application QoE, what impact these delays have when they occur on the click path during using the streaming application, and if these results could be transferred to music streaming. This work closes this gap by investigating the impact of browsing delays on the QoE of music streaming applications with a subjective QoE study.

III. Study Description

To answer the research questions, we conducted a user study on mobile devices. Here, the best practices for crowdsourcing-based QoE studies [15]–[17] were considered. The study was conducted remotely and hosted as web application with optimized UI for mobile devices, accessible via a browser. As Figure 1 shows, the study consisted of three parts:

1. Study Preparation and Demographic Questionnaire: First, before the start of the study, all media files (including all songs and images) were preloaded to the local browser cache to ensure that no unplanned delays occur during the online and remote execution of the experiment, which could distort the measurement results. To ensure usage of a mobile device, the touchscreen property of the participants' devices was queried together with the screen size. If the study framework determined that the study was not opened on a mobile device, a QR code containing the URL link to the study was displayed with the instruction to scan it with a mobile device in order to participate. After the participants passed this check, they were provided a textual introduction including information about the study and the rules of the experiment. To check that all participants turned on the volume, an audio CAPTCHA [24] was included. Here, the participants had to listen to a read out number and enter this number in a text field. Only users who entered the read-out number correctly were allowed to continue in the study. Moreover, the users had to provide basic demographic information, namely, gender, age, and country of residence. In addition, they were asked how often they used music streaming services and if they had a premium subscription for a music streaming service.

2. Study Tasks and QoE Questionnaire: The main part of the experiment included three explicit study tasks, in which participants were asked to test and rate a music streaming application. Screenshots of the application are shown in Figure 2. At the start of each task, participants were given a set

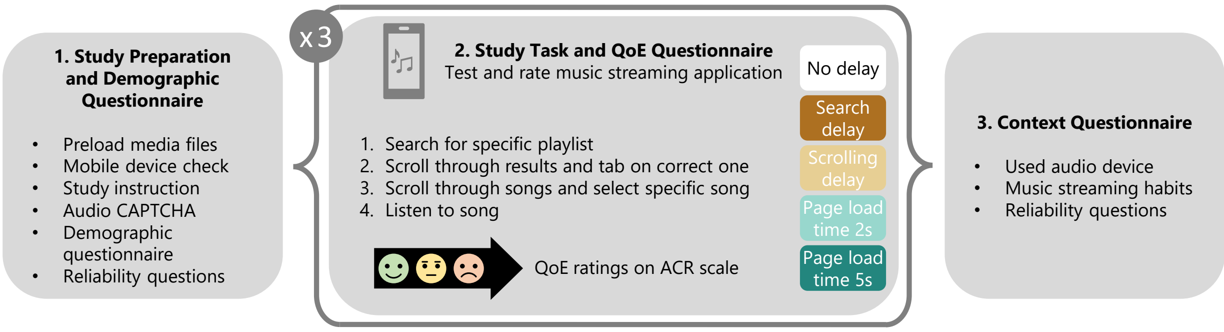
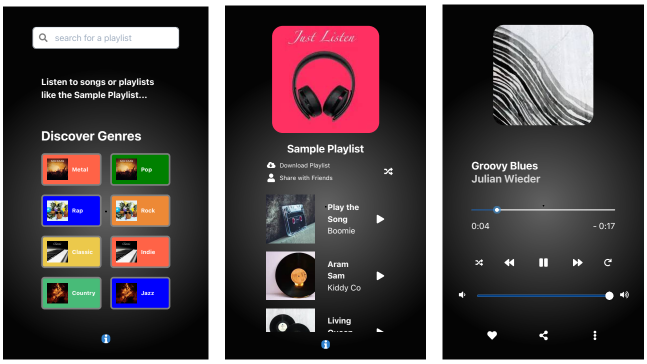


Fig. 1: Structure of the user study.



(a) Searching screen (b) Playlist screen (c) Player screen

Fig. 2: Screenshots of the music streaming web application.

of four basic instructions. The first instruction was to tap the search bar and search for a specific playlist (see Figure 2a). The search bar has a dynamic search functionality, but will only include the correct result after the participant enters at least six letters. Second, they had to find the correct playlist in the results and tap on it to open the playlist. Third, after loading the playlist, they had to search for a specific song by scrolling through the playlist and tap on the title, the cover, or the play button to start the playback (see Figure 2b). Finally, the participants had to listen to the selected song (see Figure 2c) for about 20 s. Afterwards, they were asked to rate their experience with the application, as well as with the streaming of the music, by answering typical QoE questions with 5-point Absolute Category Rating [25], e.g., "*How would you rate your experience using the app?*" and binary questions, such as "*Did you notice any delays?*". If at any point during the task, the participants had trouble to remember the next instruction, there was an information icon at the bottom of the screen with information about what to do next.

In each of the three study tasks, exactly one of the following five delay conditions occurred: No delay, a search delay while waiting for the result of the search query (5 s), a page load time when tapping on the playlist to open it (2 s or 5 s), or a scrolling delay while scrolling through songs in the playlist and waiting for additional titles to show up (5 s). The length of the delays was chosen with respect to the time it takes a person to notice a delay (approx. 1 s) and the time that is the limit for someone to keep focused (approx. 10 s) [19]. Note that the audio playback was always uninterrupted such that no additional QoE degradation is introduced during streaming.

3. Context Questionnaire: Finally, the participants had to answer several context questions. This questionnaire included questions with respect to the used audio device (speakers or headphones) and the participants' habits regarding the usage of music streaming. Additionally, some questions were repeated from the initial questionnaire as reliability checks.

Filtering of unreliable participants: In order to detect participants, who did not complete the task properly, and to exclude their results before the evaluation [16], we processed the study logs and assessed the reliability of all participants.

First, users were filtered if they failed to consistently answer closely related personal questions. For example, on the demographic questionnaire in the first part of the study, we asked the participants on which continent they lived and how often they used music-streaming services. In the context questionnaire at the end of the study, the participants had again to state how often they used music-streaming services and, additionally, in which country they lived. We filtered out participants who could not consistently answer both sets of questions.

In addition to these personal data checks, we also included study-related reliability checks to verify that the participants listened to the music carefully. Here, the participants had to answer questions about the audio content, for example, whether the streamed song contained vocals. Again, we filtered out participants who failed to answer such questions correctly.

Finally, we monitored the time from beginning to completion of the study to ensure that participants were not able to bypass any study task and skip to the end or interrupted the study to spend time on any other task. Here again, we excluded participants with a significantly too low or too high study execution time.

IV. Results

In the following, we explore the collected data to answer the research questions. First, the data set is described. Afterwards, **RQ1** is examined by investigating the influence of browsing delays on the perceived music streaming QoE. Finally, to address **RQ2**, the influence of browsing delays on the overall app experience is evaluated including the impact on the acceptance of the app under the given conditions.

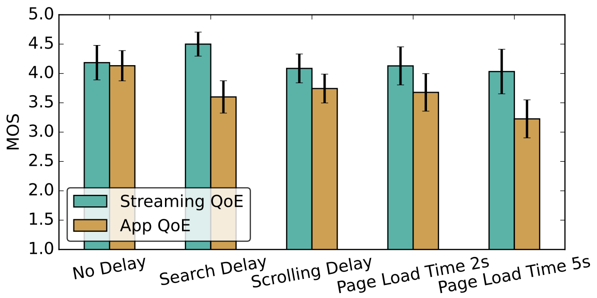


Fig. 3: Comparison of the application's MOS and the stream's MOS for each scenario.

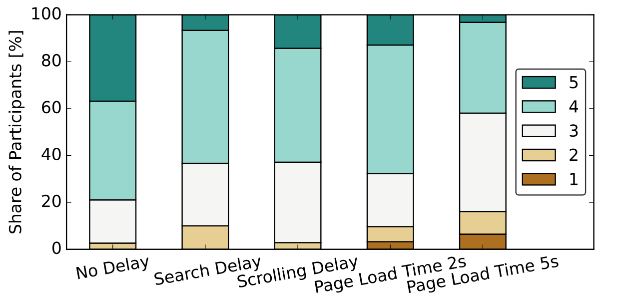


Fig. 4: Distribution of the overall app QoE ratings per scenario.

TABLE I: Distribution of participants per setting.

	No Delay	Search Delay	Scrolling Delay	Page Load Time		$\sum$
	5s	5s	5s	2s	5s	
#Participants	38	30	35	31	31	165

A. Data Set Description

The study was carried out with the help of volunteers who were approached through students at our university and their personal contacts. The time necessary to complete the study was estimated at five to ten minutes. Overall, a total of 72 participants took part in this study. After filtering out incomplete and unreliable participation by processing the reliability checks, in the end, 55 participants remained. Out of these, 27 were female, 27 were male, and one person identified themselves as other. Since the participants were mainly recruited among students of the university, the majority (90.91 %) were between 20 and 29 years old. Looking at music streaming habits, 85.45 % of the participants stated that they use music streaming apps at least twice a week while only 0.04% reported that they never use such apps. When asked about the audio device they used for the study, 41 participants indicated that they used a speaker while 14 participants used headphones. Although especially when listening to music, the output device can have an influence on the sensation, pretests showed that the used audio device type had no significant influence on the perceived QoE. As described in Section III, each participant repeated the main study three times with a different delay in each run. The study conditions were assigned uniformly random to the participants. The resulting distribution per delay type after filtering the participants can be seen in Table I. In total, we can evaluate ratings of 165 study tasks with at least 30 participants per delay type.

B. Influence of Browsing Delays on Music Streaming QoE

Before we can investigate the influence of browsing delays on music streaming QoE, we first have to check whether participants can distinguish between streaming QoE and application QoE at all. Figure 3 shows the MOS values and the corresponding 95% confidence intervals for all delay conditions, which are listed on the x-axis. For each condition, the MOS for streaming QoE is depicted in teal green and the corresponding MOS for the overall app QoE in brown. While the MOS values for streaming and app experience without delay do not differ much, larger differences can be seen for the other delay types. However, it is unclear if any of these differences is significant, or if the ratings of streaming QoE and app QoE are generally similar. Bringing this to a statistical context, this means, if we cannot reject the null hypothesis that both concepts were rated similarly, we cannot assume that participants could distinguish both concepts. Following [12], we test the QoE differences between the paired streaming QoE and app QoE ratings of all participants by conducting Friedman tests for all delay conditions. Since we compare only two QoE distributions at a time, streaming QoE and app QoE, this is equivalent to conducting the sign test. To account for the multiple comparisons problem, we apply the Šidák correction to the resulting p-values. Focusing on the scenario with no delay, we obtain a high p-value (> 0.05), and thus, no significant difference between the rating of streaming QoE and app QoE is visible. However, the situation is different for the scenarios with delays. Except for the condition with scrolling delay, for all other conditions, small p-values below 0.05 result, which shows that there are indeed significant differences between the rated experience for streaming and the overall app QoE. Thus, we conclude that the participants generally understood the study task and were able to distinguish correctly between the satisfaction with the streaming alone and the general satisfaction with the app.

Next, we can investigate the actual influence of browsing delays on music streaming QoE. For this purpose, we take another look at Figure 3 and focus on the streaming QoE in teal green. Here, we see that the MOS in each delay condition is above 4. Moreover, it can be seen that the confidence intervals of the scenarios with delays are overlapping with the confidence interval of the no delay scenario. Since, we compare different delay conditions, which were rated by different subsets of participants, we test the streaming QoE differences between the delay conditions and the no delay condition using pairwise Mann-Whitney U tests, again including Šidák correction. As all $p > 0.05$, we find no significant differences between the perceived streaming QoE of study tasks which experienced browsing delay and those which experienced no delay. Thus, **RQ1** can be negated as browsing delays do not impact the QoE of subsequent music streaming.

Summarizing, no difference can be seen between the brows-

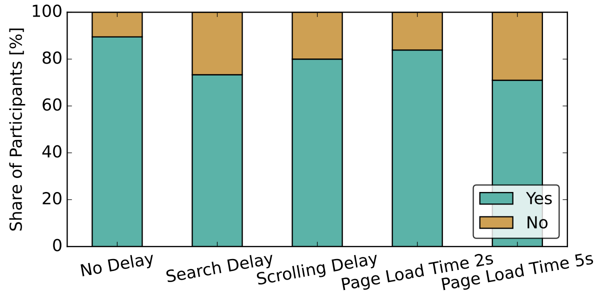


Fig. 5: Percentage of participants who would accept the app's performance for each scenario.

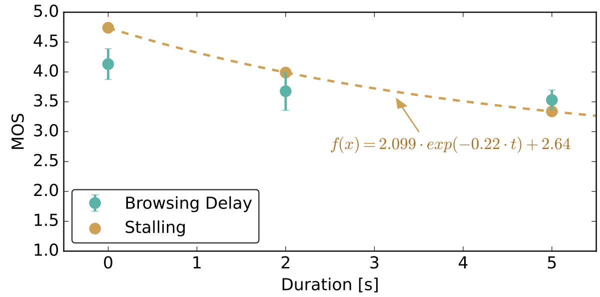


Fig. 6: Comparison of browsing delay QoE degradation to stalling QoE degradation using the model $f(x)$ from [6].

ing delay scenarios with respect to the streaming QoE, whereas some browsing delays do have an impact on the general satisfaction with the app.

C. Influence of Browsing Delays on Overall App Experience

To examine the influence of browsing delays on the overall app experience in more detail, Figure 4 shows the distribution of QoE ratings regarding the overall app experience per experienced delay type. Here, in alignment with the ACR rating scale [25], rating of category 1 corresponds to bad experience, whereas category 5 represents excellent experience. While the QoE distribution for no delays shows mainly ratings between 3 and 5, i.e., between fair and excellent experience, with a good or better ratio (GoB) of 78.95%, the distributions differ for the scenarios which include delays. Testing the QoE differences of the overall app QoE ratings, again using pairwise Mann-Whitney U tests for the different delay types and Šidák correction, we find that only search delay and page load time of 5 s differ significantly to the no delay condition, showing $p < 0.05$. This can be especially seen for a page load time of 5 s, for which the GoB ratio decreases down to only 41%. However, also for the other delay types, there are differences in the distributions visible. In general, considerably fewer excellent ratings have been given, from 36.84% for no delay down to, for example, 14.28% for the scrolling delay and 14.90% for a page load time of 2 s. This shows that, although the overall QoE rating distribution did not significantly change in these conditions, even these delays can be perceived and negatively experienced by some users. Thus, **RQ2** can be answered in the affirmative as we see that browsing delays show an influence on the overall music streaming app experience.

D. Implications for Music Streaming Providers

Finally, we take a look at the implications for music streaming providers. Together with the QoE ratings, the study participants indicated whether they would still use the application despite the experienced delays. The resulting acceptance ratios for the app's performance in each delay condition are shown in Figure 5. Here, yes (teal green) means that participants would use the music streaming app under the given conditions, no (brown) means that they would not accept the performance and would therefore not use the app. Having a look at no delay, 89.47% would accept the performance of the app. Although we did not artificially deteriorate the app performance in this scenario, this is a reasonable baseline result, as participants might have still disliked the implemented music streaming application for other reasons. However, it can be seen that the acceptance ratio decreases considerably for the other conditions with additional delay. In the best case, which is the experience of a page load time of 2 s, the acceptance ratio decreases down to 83.87%. In the worst case, when experiencing a page load time of 5 s, the acceptance ratio drops by 18.50% to only 70.97%. Considering that music streaming is a competitive market, this substantially increased share of unsatisfied users can lead to a massive reduction of revenues for music streaming applications providers. Thus, improving only streaming QoE is not sufficient for reaching high overall app QoE, but also browsing experience has to be considered.

The importance of considering browsing delays when using music streaming applications is also evident in Figure 6. Here, browsing delay induced QoE degradation is compared to QoE degradation caused by stalling during the actual music streaming. The x-axis shows the overall duration of browsing delay or stalling, while the y-axis depicts the resulting MOS value. To obtain MOS values depending on the browsing delay duration, we combined the results of the study of the paper. The figure shows MOS and 95% confidence intervals in teal green over the no delay condition with a browsing delay duration of 0 s (MOS: 4.13), the page load time condition with 2 s (MOS: 3.68), and all other conditions with 5 s (MOS: 3.53). We obtain MOS values for stalling at these durations using the model $f(x) = 2.099 \cdot exp(-0.22 \cdot t_{stall}) + 2.64$ from [6], which is depicted as a brown dashed line. This means, when experiencing no stalling, 2 s, or 5 s of stalling and otherwise perfect app experience, according to the model in [6], an average user gives a QoE rating of 4.74, 3.99, or 3.34, respectively. It can be seen that, for durations of 2 s and 5 s, both stalling and browsing delay result in similar MOS degradation. Thus, we can see here that both factors are equally important, and therefore, when considering a music streaming app, both components should be taken into account for a holistic QoE assessment.

V. Conclusion

Since music streaming is becoming more and more popular, especially on mobile devices, a high Quality of Experience is very important to providers of music streaming applications. While streaming-related QoE factors, such as stalling events, have already been studied in related work, the impact of waiting times during using and browsing a music streaming app have not been investigated yet.

Aiming towards obtaining a more holistic understanding of the QoE of music streaming application, this work investigated the impact of browsing delays along the click path of the users. For this, we developed a custom web application for music streaming, which mimics the look and feel of current mobile music streaming apps. Our custom app is designed for mobile phones and allows to artificially insert various browsing delays along the click path and to gather corresponding QoE ratings. Using this mobile app, we conducted an online QoE study and investigated the influence of different browsing delays on the perceived streaming QoE and overall application QoE.

The results showed that browsing delays along the click path have no direct impact on the QoE of subsequent music streaming. This indicates that study participants were able to clearly separate both main functionalities of music streaming apps, and did not allow negative experiences with browsing to deteriorate the streaming experience, which is in line with the results of [23] for video streaming. Nevertheless, we found that certain browsing delays have a significant influence on the overall app experience. This, in particular, included delays when waiting for the result of a search bar query and long page load delays when loading a playlist. In addition, our results showed that browsing delays can reduce the participants' acceptance of the music streaming app performance by up to 18.50%. This could lead to a substantial reduction of revenues for music streaming providers in this competitive market.

Thus, our results show that it is not sufficient for music streaming service providers to measure and improve only streaming QoE. In contrast, the overall QoE with a music streaming app is also negatively affected by browsing delays, which, according to the results of [6], can result in a QoE degradation similar to stalling during the streaming. What makes browsing QoE along the click path even more important is that customers may leave the app before the actual streaming starts when they are dissatisfied with their browsing experience. Thus, browsing needs to be taken into account for the QoE-aware design [5] of music streaming applications.

In future work, we will extend our mobile music streaming app to include longer and other types of delays along the click path, a combination of different browsing delays, or a combination of browsing delays and streaming impairments to investigate more complex QoE degradation patterns that can occur when using music streaming apps. This will allow to come closer towards fully and holistically understanding the QoE of music streaming applications.

References

[1] T. Germain, "Best Music Streaming Services," 2021, accessed: 2022-03-31. [Online]. Available: https://www.consumerreports.org/streaming-music-services/best-music-streaming-service-for-you/

[2] Spotify, "Number of Spotify Premium Subscribers Worldwide from 1st Quarter 2015 to 1st Quarter 2021," 2021, accessed: 2022-03-31. [Online]. Available: https://s22.q4cdn.com/540910603/files/doc_financials/2021/q1/Shareholder-Letter-Q1-2021_FINAL.pdf

[3] K. Aniket and R. Deshmukh, "Online Music Streaming Market," 2021, accessed: 2022-03-31. [Online]. Available: https://www.alliedmarketresearch.com/online-music-streaming-market-A11156

[4] *Qualinet White Paper on Definitions of Quality of Experience*. Lausanne, Switzerland: European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), 2012.

[5] A. Seufert, S. Schröder, and M. Seufert, "Delivering User Experience over Networks: Towards a Quality of Experience Centered Design Cycle for Improved Design of Networked Applications," *SN Computer Science*, vol. 2, no. 6, 2021.

[6] A. Sackl, S. Egger, and R. Schatz, "Where's the Music? Comparing the QoE Impact of Temporal Impairments Between Music and Video Streaming," in *5th International Workshop on Quality of Multimedia Experience (QoMEX)*, 2013.

[7] A. Schwind, C. Moldovan, T. Janiak, N. D. Dworschak, and T. Hoßfeld, "Don't Stop the Music: Crowdsourced QoE Assessment of Music Streaming with Stalling," in *12th Intl. Conf. on Quality of Multimedia Experience (QoMEX)*, 2020.

[8] A. Schwind, L. Haberzettl, F. Wamser, and T. Hoßfeld, "QoE Analysis of Spotify Audio Streaming and App Browsing," 2019.

[9] A. Raake, *Speech Quality of VoIP: Assessment and Prediction*. Wiley Online Library, 2006.

[10] International Telecommunication Union, "ITU-T Recommendation G.1011: Reference Guide to Quality of Experience Assessment Methodologies," 2015.

[11] R. C. Streijl, S. Winkler, and D. S. Hands, "Mean Opinion Score (MOS) Revisited: Methods and Applications, Limitations and Alternatives," *Multimedia Systems*, vol. 22, no. 2, 2016.

[12] M. Seufert, "Statistical Methods and Models based on Quality of Experience Distributions," *QUEX*, vol. 6, no. 1, 2021.

[13] M. Hirth, T. Hoßfeld, and P. Tran-Gia, "Anatomy of a Crowdsourcing Platform - Using the Example of Microworkers.com," 2011.

[14] B. Gardlo, S. Egger, M. Seufert, and R. Schatz, "Crowdsourcing 2.0: Enhancing Execution Speed and Reliability of Web-based QoE Testing," in *Intl. Conf. on Communications (ICC)*, 2014.

[15] S. Egger-Lampl, J. Redi, T. Hoßfeld, M. Hirth, S. Möller, B. Naderi, C. Keimel, and D. Saupe, "Crowdsourcing Quality of Experience Experiments," in *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments*. Springer, 2017.

[16] T. Hoßfeld, M. Hirth, J. Redi, F. Mazza, P. Korshunov, B. Naderi, M. Seufert, B. Gardlo, S. Egger, and C. Keimel, "Best Practices and Recommendations for Crowdsourced QoE - Lessons learned from the Qualinet Task Force Crowdsourcing," 2014.

[17] F. Daniel, P. Kucherbaev, C. Cappiello, B. Benatallah, and M. Allahbakhsh, "Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions," *ACM Computing Surveys*, vol. 51, no. 1, 2018.

[18] A. Schwind, F. Wamser, T. Gensler, P. Tran-Gia, M. Seufert, and P. Casas, "Streaming Characteristics of Spotify Sessions," in *10th Intl. Conf. on Quality of Multimedia Experience (QoMEX)*, 2018.

[19] J. Nielsen, *Usability Engineering*. Morgan Kaufmann, 1993.

[20] S. Egger, T. Hossfeld, R. Schatz, and M. Fiedler, "Waiting Times in Quality of Experience for Web based Services," in *4th International Workshop on Quality of Multimedia Experience (QoMEX)*, 2012.

[21] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A Generic Quantitative Relationship between Quality of Experience and Quality of Service," *IEEE Network*, vol. 24, no. 2, 2010.

[22] WebPageTest.org, "Speed Index," 2012, accessed: 2022-03-31. [Online]. Available: https://docs.webpagetest.org/metrics/speedindex/

[23] M. Seufert, O. Zach, M. Slanina, and P. Tran-Gia, "Unperturbed Video Streaming QoE Under Web Page Related Context Factors," in *9th Intl. Conf. on Quality of Multimedia Experience (QoMEX)*, 2017.

[24] L. v. Ahn, M. Blum, N. J. Hopper, and J. Langford, "CAPTCHA: Using Hard AI Problems for Security," in *22nd Intl. Conf. on the Theory and Applications of Cryptographic Techniques (Eurocrypt)*, 2003.

[25] International Telecommunication Union, "ITU-T Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications," 2008.