A Usability Study Comparing iTunes and Spotify

Juan S. Carrillo

September 26, 2014

1 Introduction

For this study our group wished to compare the usability of two music services, iTunes and Spotify. For the purpose of our study we only included the desktop version of Spotify, since iTunes does not have a mobile version. The focus of our study is to determine which of the two music services has better usability.

2 Background

2.1 Metrics

Our study focuses on the efficiency, rate of errors, and satisfaction of the user with the two music services. We study the *efficiency* of the music service by measuring the time is takes for the user to perform a given task. When we test for efficiency, we only consider users that are already familiar with the music service. In this way, we are sure that the user is not wasting time becoming familiar with the system itself; all the work on the part of the user is focused on completing the task at hand. When we measure the *rate of errors* while using the programs, we analyze how many errors are made on the part of the user, that is, when they perform an action that has unintentional results. We quantify these by simply counting how many errors were committed. In this study we found that people that took a long time to finish a task were hindered by one significant error rather than a series of minor errors. We also surveyed the *satisfaction* of the users. We simply asked the users to report on a numeric scale how satisfied they were with the ease of performing the given tasks on with the music service and we also asked how satisfied they were with the music service as a whole.

2.2 Tasks

For the study we wanted to come up with tasks that were complex enough that the user would have to perform two or more actions in order to accomplish the task but at the same

time were familiar enough to the user so that they understood the task. We choose the following tasks:

- **Creating an empty playlist**: The user was asked to create an empty playlist with the title "Hello World".

- **Creating a radio station based on a given artist**: The user was asked to create a radio station based on the artist Tiësto and play it.

- **Looping a playlist**: The user was asked to play a playlist and set it to loop indefinitely, i.e. the music service would play all the songs in the playlist and when it had gone through all the songs in the playlist it would iterate through the playlist again and continue playing the songs in the list. The user was allowed to use a playlist that they had already created beforehand or were given time to create a playlist in order to finish the task.

Each of these tasks are familiar to most users and required more than one action to complete, i.e. pressing a button and writing out the title of a playlist. These tasks are also possible in both systems. Since iTunes is a media player while Spotify is a music streaming service, we had to take care as to create tasks that were possible in both systems.

3 Procedure

We created a spreadsheet using Google Docs and created a form that would record all the data from the survey on the sheet. Each member of our group recorded the data on the form for an individual participant. The form contained the following fields:

- The service that the user was testing

- Whether or not the user had previously used the music service

- How long had the user been using the music service

- How the user rated themselves on how familiar they were with the music service

- How the user rated their proficiency with computers

- The time it took to the user to complete a given task

- How many errors, if any, were committed while completing each task

- A brief qualitative description of the errors, if any

- Any additional notes on any given task

- Satisfaction with the ease of the given task

- Overall satisfaction with the music service

For the question concerning how familiar they were with the music service, the user was asked to rank themselves from 1 to 10, 1 being very unfamiliar and 10 being very familiar with the service. Similarly, the user was asked to rate their proficiency with computers on a scale from 1 to 10, 1 being not very proficient and 10 being very proficient.

We allowed the users to use their personal computer for the survey in order to make sure the users did not take longer because they were using an operating system they are unfamiliar with. Also, there is no significant difference between the application windows across operating systems, as seen in Figures 1 and 2. Note that the application windows are stacked in the order of Mac and Windows.
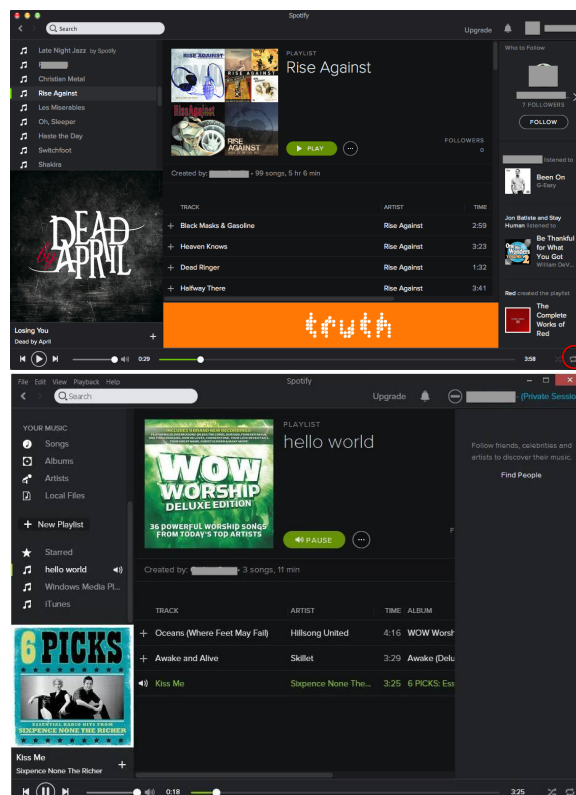

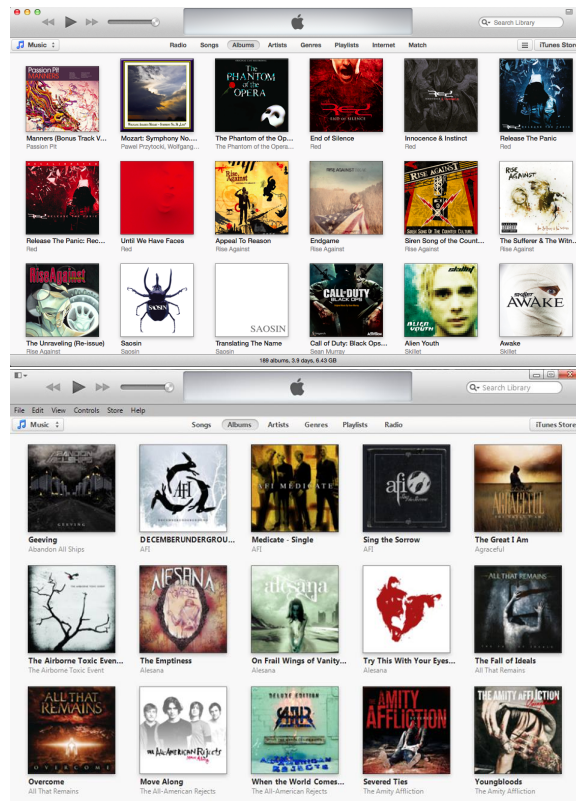Figure 1: Spotify Desktop Application View

Figure 2: iTunes Application View

At the beginning of the survey we collected information for the first five fields. After collecting the preliminary data, we explained the task to the participant and then timed how long it took for each participant to complete the given task. We also made note of any errors committed by the user. In between each task we asked the participant to report their satisfaction with each task on a scale of 1 to 10, 1 being very unsatisfied and 10 being very satisfied. We also recorded any additional comments provided by the user, if relevant. At the end of the survey we asked the user to rate their overall satisfaction with the music service on a scale of 1 to 10, 1 being very unsatisfied and 10 being very satisfied.

4   Results

When looking at the results, I placed most of the importance on rate of errors and learnability than satisfaction. I do this because I believe that satisfaction is a direct result of the other two metrics; if the system is very learnable o if the rate of errors is low then the satisfaction of the users with the system will be high. Thus, I base my assessment of the

system more on the metrics of learnability and rate of errors, while using satisfaction to confirm the results from the other metrics.

4.1 Participants

We had 15 examinations, 8 for iTunes and 7 for Spotify. Some of the users participated in the survey twice, testing one service and then the other. All participants that tested iTunes were already familiar with iTunes. Of the 7 participants who tested Spotify only 4 were prior experience with Spotify. The average self-assessed proficiency of the participants was about 7.46, with a standard deviation of 1.73, meaning that users considered themselves to be proficient at using a computer.

4.2 Experience and Satisfaction

iTunes Satisfaction	Spotify Satisfaction
10	8
8	10
9	8
6	10
8	8
10	9
3	10
8	-
Average: 7.75	Average: 9
Standard Deviation: 2,31	Standard Deviation: 1

Years using Spotify	Years using iTunes
0	8
0	8.5
0	5
0	18
0.5	4
0.16	8
1.5	10
-	7
Average: 0.542	Average: 7.31
Standard Deviation: 0.56	Standard Deviation: 1.94

Figure 3: User Satisfaction and Experience

All the people who tested iTunes were already familiar with the service. Looking at Figure 3, we see that on average, those participants that had used iTunes had been using it for about 7 years. In contrast, three out of the seven participants for Spotify had no previous experience with Spotify. As a result of this and the small sample size, the average participant had only 6 months of experience with Spotify. Also, no user had more than 18 months of experience with Spotify. The average is not The average satisfaction with the system was 7.75, with a standard deviation of 2.31, which means that users were overall satisfied with the system. It can be noted that one user in particular rated their satisfaction with iTunes at a 3, and the next lowest score was a 6. Due to the small sample size, the user's score greatly lowered the average satisfaction for iTunes. It can also be noted that the user took 1 minute and 30 seconds on one of the tasks, and their overall satisfaction with iTunes may be attributed to the relatively long time it took the participant to complete that task. On the other hand, those who participated with Spotify reported on average a satisfaction of 9. The data we collected suggests that overall users have greater satisfaction with Spotify.

4.3   Creating the Playlist

Errors for Creating Playlist	
iTunes	Spotify
1	0
2	0
0	2
0	0
0	1
0	0
0	1
0	-
Average: 0.375	Average: 0.571

Times for Creating Playlist in Seconds	
iTunes	Spotify
16	19
14.84	6
20	15.69
26	7
10.84	9.50
10.07	13.03
18	8
15	-
Average: 16.34	Average: 11.17
Standard Deviation: 5.11	Standard Deviation: 4.87

Satisfaction for Creating Playlist	
iTunes	Spotify
10	9
9	10
8	8
8	10
10	9
10	10
9	10
8	-
Average: 9	Average: 9.42
Standard Deviation: 0.92	Standard Deviation: 0.79

Figure 4: Times, Errors, and Satisfaction Creating a Playlist

Users were significantly faster using Spotify when creating the playlist, despite also committing more errors while using Spotify. Most users noted that it was easy to create a playlist in Spotify because there is a widget on the left-hand side with a plus sign labeled New Playlist. Most errors involved misspelling the title of the playlist, "Hello World". Although those using iTunes varied in their approaches to creating the playlist, i.e using keyboard shortcuts or clicking on buttons, they were on average much slower than when the participants used Spotify. It is to be noted that one of the users complained that iTunes is constantly changing its layout, making the users remember new ways to create a playlist. I speculate that the change of the layout causes users to waste time trying to use their usual approach and realizing that the change of the layout changes the way they must create a playlist. We can see that Spotify had a higher satisfaction for this task than iTunes.

4.4 Creating the Radio Station

Errors for Creating Radio Station	
iTunes	Spotify
0	2
0	0
2	0
0	0
2	0
2	0
0	0
2	-
Average: 1	Average: 0.285

Times for Creating Radio Station in Seconds	
iTunes	Spotify
9	78
21.4	14
121	15.82
8	18
24.83	10.5
28.28	18.04
90	8
30	-
Average: 41.56	Average: 23.19
Standard Deviation: 41.12	Standard Deviation: 24.45

Satisfaction for Creating Radio Station	
iTunes	Spotify
10	1
9	10
5	8
9	10
7	9
9	9
8	9
6	-
Average: 7.875	Average: 8
Standard Deviation: 1.73	Standard Deviation: 3.16

Figure 5: Times, Errors, and Satisfaction for Creating Radio Station

In both music services there were users that took a very long time to figure out how to create a radio station based on an artist. However, Spotify again proved to be faster as users took on average 23.19 seconds complete the task, while completing the same task on iTunes took on average 41.56 seconds. Since the standard deviation for Spotify is much smaller than the one for iTunes, it also demonstrates that the distribution of times is also much tighter than that of iTunes. It is important to note that there were many errors when using iTunes. All of them involved users being unable to find the method to create a radio station based on an artist. However, on Spotify users usually looked up the artist and then looked for ways to create a radio station from there. I speculate that the reason that users were much faster on Spotify is that there were two ways of creating a radio station: finding the artist and creating a station, or going to the radio button and then clicking "Create New Station", which prompted the user to type in the name of the artist. However on iTunes the option of looking up the artist and then creating a radio station was not available. Here, I believe the deciding factor was that there was more than one way to do the same thing in Spotify. Again, we see that Spotify had a higher satisfaction than iTunes.

4.5 Looping the Playlist

Errors for Looping Playlists		Times for Looping Playlist in Seconds	
iTunes	Spotify	iTunes	Spotify
2	2	82	77
2	0	26.13	7
0	0	7	12.92
1	0	63	7
0	0	12.09	5
0	0	9.97	11.52
1	2	61	20
1		7	-
Average: 0.875	Average: 0.571	Average: 33.52	Average: 20.06
		Standard Deviation: 30.35	Standard Deviation: 25.60

Satisfaction for Looping Playlist	
iTunes	Spotify
9	6
8	10
10	8
9	10
10	10
10	10
9	7
10	-
Average: 9.375	Average: 8.71
Standard Deviation: 0.74	Standard Deviation: 1.70

Figure 6: Times, Errors, and Satisfaction for Looping the Playlist

This task was unusual in the sense that the music service that had the least amount of errors and took the least amount to complete had the lower satisfaction. One average it took 20 seconds to loop a list in Spotify, while it took 33 seconds to perform the same task on iTunes. Also, there were less errors on Spotify than on iTunes. However, there were two users who gave Spotify a satisfaction of 6 and 7, which is lower than the any satisfaction score given to iTunes. Both of these users were unable to get the playlists to loop right away; one of the users couldn't find the icon to loop and the other did not to play the playlist first, which is necessary to activate the loop button. I believe that the reason for these low scores is that the users were disappointed at not being able to find the icons right away. Seeing as how this case goes against the pattern established by the previous two tests, where the music service that required the least amount of time and had less errors had higher satisfaction, it prompts the author to want to gather a larger sample size in order to draw a good conclusion.

5 Analysis

5.1 Efficiency

It is difficult to measure the efficiency for all the tasks evenly. We certainly cannot compare the efficiency of one system to the next because all participants were already familiar with iTunes, while not all participants were previously acquainted with Spotify. The sample size for Spotify is too small to consider the efficiency of Spotify.

5.2 Learnability

I believe that Spotify had better learnability than iTunes. This becomes obvious in the task that asked users to create a radio station based on an artist. I attribute Spotify's success because they had two different methods to do the same thing, and both seem intuitive given the model of Spotify. As a streaming service, Spotify gives users to music they haven't purchased, unlike iTunes which serves as a media library to organize the music files they had saved on their computer. Since Spotify allows users to music from almost any artist, it is natural for a user to search an artist on Spotify's search bar. From the artist's page the user has the option to create a radio station based on the artist by clicking a button with ellipsis and then selecting "Start Artist Radio". Spotify also had a button on the side bar labeled Radio. If this button is clicked the view changes to the most recent radio station. This view has a button on the upper right hand corner labeled "Start New Station". After clicking this button the user can input an artist name into the dialog box and immediately start a new radio station based on that artist. So while this task was the one that most users were unfamiliar with, and most, if not all, users had to figure out how to complete this task for the first time regardless of the platform. It is difficult to measure the learnability as easily with the other two tasks. Most users are already familiar with the action of creating a new playlist; most of the errors in this task involved mistyping the name of the playlist.

Since all the iTunes users have had years of experience with the media player, it is safe to consider that they were all familiar with the task of creating a playlist. On the other hand, almost half of the users for Spotify had never used Spotify before. For Spotify, then, creating a new playlist was more of a learnability issue while for iTunes is was more of an efficiency issue. As previously stated, we have a rather small sample size, but it is interesting to note that beginner Spotify users take about the same time that an experienced Spotify user would to create a new playlist. We can attribute this to Nielsen's principle of "Speaking the user's language." Spotify has a button on the side of left hand side labeled "+New Playlist". When a user wants to create a new playlist, they can easily recognize the button. iTunes is a little less straightforward in comparison, as it has a simple, small button labelled + sign.
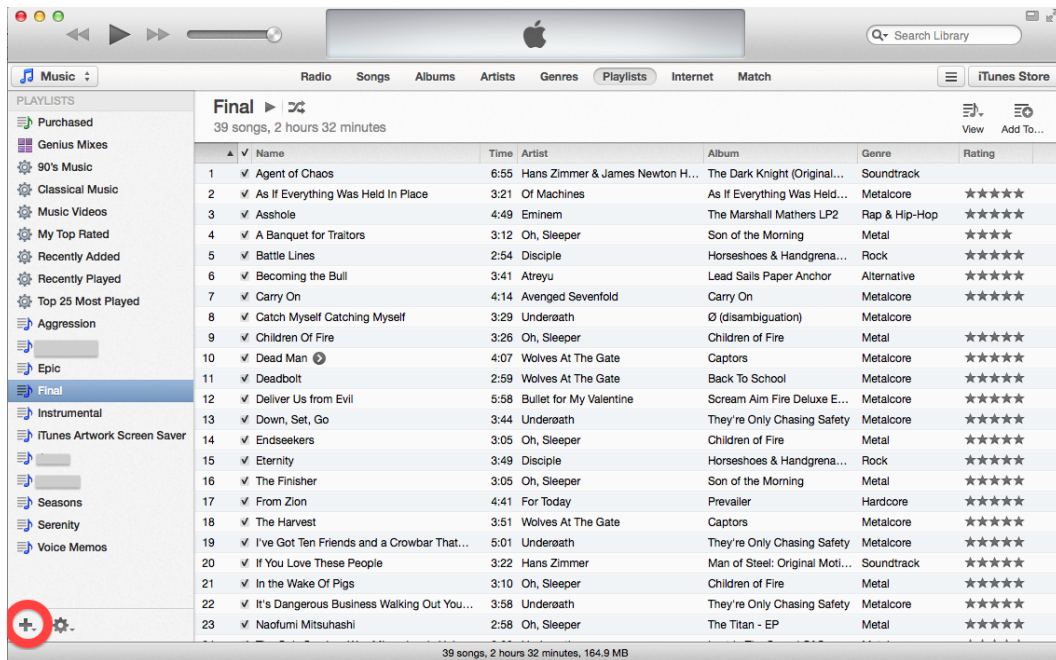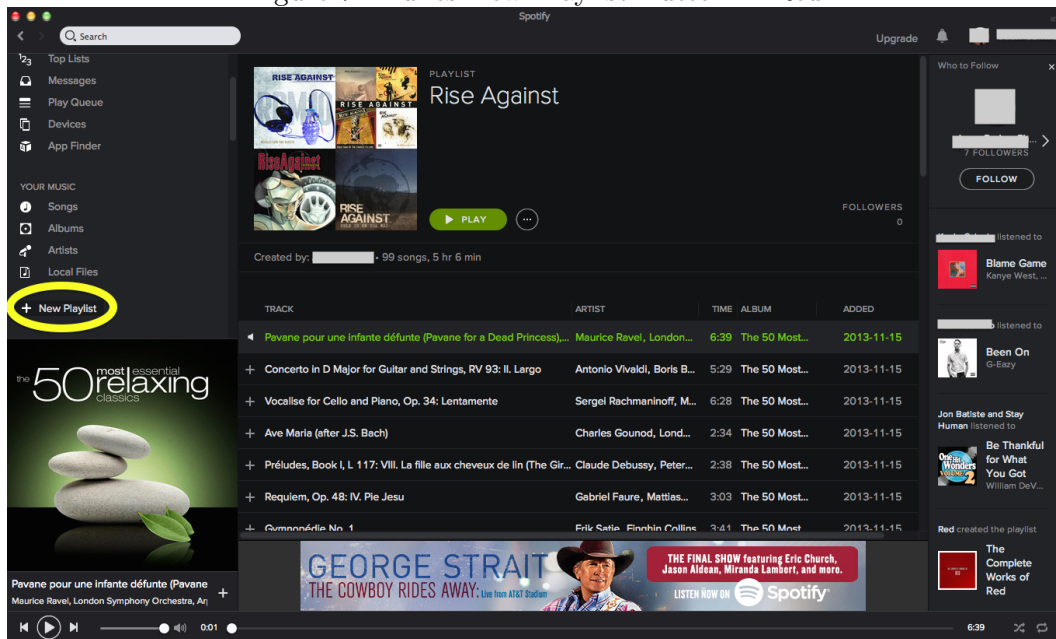
Figure 7: iTunes New Playlist Button in Red

Figure 8: Spotify New Playlist Button in Yellow

5.3 Rate of Errors

By analyzing the data, I see that Spotify has a lower rate of errors than iTunes. As previously stated, some of the users of iTunes noted that the media player has gone through several layout changes, which require iTunes users to relearn several tasks that they were familiar with. This contributes to the rate of errors because iTunes users might attempt to navigate to the command to create a new playlist and find it missing. During the release of iTunes 11**??**, the layout of iTunes changed significantly, and so users had to get used to the new release. It is interesting to note that iTunes, as an established music service, did not do much better than Spotify, a much newer piece of music software. Considering that all participants of the iTunes surveys had years of experience on iTunes, they should have had a clear advantage over mundane tasks such as creating a playlist or looping a playlist. However, I believe that the recent changes in the iTunes interface cause iTunes to lose the advantage of having a familiar interface. This is in clear violation of one of the principles shared by several interaction design thinkers such as Tognazzini, Nielsen, and Shneiderman: consistency. The lack of consistency on iTunes caused it to be just as difficult to relearn an task in an old interface as it is to learn the same task in an unfamiliar interface, in this case Spotify.

5.4 Satisfaction

Satisfaction was closely related to the previous two metrics. The higher satisfaction with the ease of a specific task usually went to the system that had lower rate of errors and higher efficiency or learnability. Even though this was not the case for the task of looping a playlist, Spotify was still able to come on top in terms of satisfaction, which is not surprising considering it did better in the other two metrics.

6   Conclusion

7   References