intelligent song picker

Counteracting decision fatigue with modern web technologies

Florian Kapaun
Interactive Media B. A.
University of Applied Science Augsburg
Design Faculty
Augsburg, Germany
hello@florian-kapaun.de

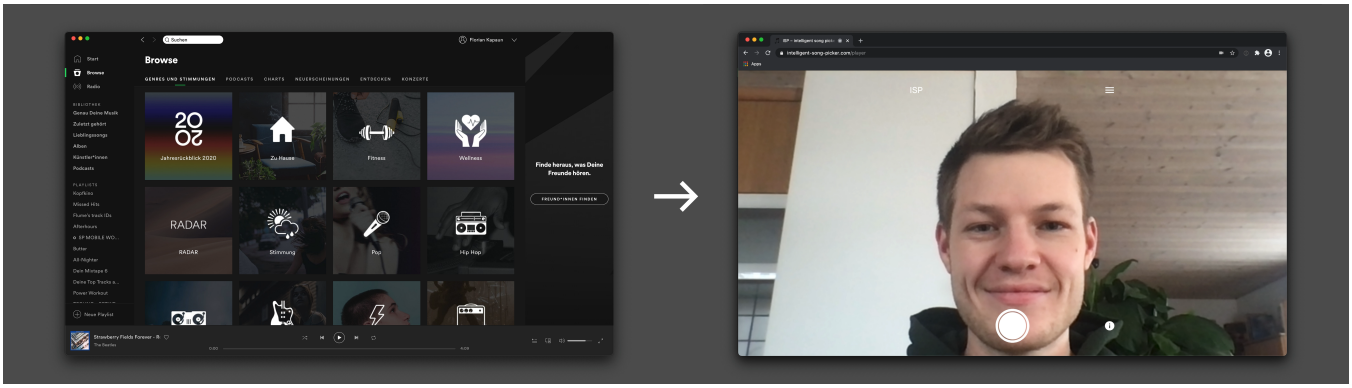

Figure 1: From decision heavy music choice to intelligent song picker

ABSTRACT

This work is an attempt to create a prototype that relieves users of the everyday decision of what music they want to listen to. I therefore used image analysis with machine learning and an intelligent song selection process. The result of my work is a fully functional and published prototype, which has shown promising results in initial tests.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; *Web-based interaction*; *HCI theory, concepts and models*; Interface design prototyping.

KEYWORDS

counteract decision fatigue, reduce cognitive load, new interaction technologies, image analysis, machine learning

Interaktion/Programmierung WS 20/21, October 01– January 31, 2021, Augsburg, Germany

ACM Reference Format:
Florian Kapaun. 2021. intelligent song picker: Counteracting decision fatigue with modern web technologies. In *Proceedings of Interaktion/Programmierung WS 20/21.* ACM, New York, NY, USA, 4 pages.

1 MOTIVATION

Streaming services have disrupted the music industry. Providers like Apple Music and Spotify allow their customers to access virtually any song, anywhere, anytime. But with these new opportunities come new challenges, such as *decision fatigue*.

Research showes that "there is a finite store of mental energy for exerting self-control"[1]. These and other observations indicate that "No matter how rational and high-minded you try to be, you can't make decision after decision without paying a biological price"[1]. which specifically means "that decision quality declines after an extensive session of decision-making"[2]. This is known as decision fatigue and causes us humans to make worse decisions over time, or even shy away from decisions in general.

One strategy against decision fatigue, made famous not least by Steve Jobs, Mark Zuckerberg and Barack Obama, is to wear the same outfit every day[3].

When it comes to streaming, many users are faced with the same, recurring decision: "What do I want to listen to?". Yet most people would probably describe this decision as unimportant. Therefore, in this project, I'll try to minimize the number of decisions when launching music on a streaming platform. For this, I will use existing technologies to take decision-making processes away from the user. The goal is to start a song that fits the user's mood and situation without the user having to make a decision. In the course of this

paper, I will explain conceptual choices, show which technologies are used, and what insights can be derived from the experiment.

2 RELATED WORK

There are three projects in particular that I will briefly review here to show what already exists in this experimental field.

In a blog article published in 2017 by Mario Noioso, he documents one of his projects whose use case he describes as follows: "choose a music track based on emotions detected on your face"[4]. He also goes into some detail about the technical implementation which makes clear that users have to upload a portrait photo, whose emotion is then classified. Based on the emotion found, a genre is then selected from which music is afterwards played.

A similar application was documented in 2020 by David Noah in the article "Microsoft Emotion and Spotify 'Mood Music' Project [API Smash]"[5]. This project is primarily about linking different API interfaces together. The result "identifies an emotion from an image [...], then recommends a musical playlist based on the emotion identified [...]"[5].

Kalyssa Owusu describes in her Medium article[6] the construction of an expression based playlist recommender. Her application "[...] takes in webcam input of the user's face, detects their facial expression"[6] and is then "Searching for a playlist with the emotion name in it (taking the first search result)"[6].

What these three projects have in common is that they exclusively use the emotions of the users to find suitable music. However, the photos provided by users contain other valuable information that I think should also be used. For example, one could make assumptions about the mood and nature of the environment in which the user is located.

I think that in all three projects the process of selecting appropriate music is superficial, which is not surprising considering that this selection was not the focus of the respective projects. Mario Noioso, for example, relies on genres when selecting songs to play. But within the genres there are serious differences in the sound and mood of the respective songs. In the project described by David Noah, predefined playlists are selected and Kalyssa Owusu uses the first playlist that appears when searching for the name of the emotion on Spotify. In these two cases,there is already more emphasis placed on the mood of the songs, as playlists curated according to the mood of the contained songs are used. But even these two approaches do not allow personalization of the played results to the users' preferences.

Another criticism from my project requirements perspective is the decision-producing process of selecting and uploading photos in Noioso and Noah's projects.

What all three already do very well is the linking of different existing services and APIs. Furthermore, the three projects show that it is possible to classify emotions in images reasonably well using modern web technology and that modern clients are capable to perform these operations.

Those three projects form the baseline for my experiment.

3 APPLICATION SCENARIO AND USE CASE

3.1 Web application

Development of a web application where users can take a photo of themselves and then have Spotify play a song that matches their mood and environment – an intelligent song picker, short *ISP*.

3.2 Definition of the use cases

The aim of the ISP is to relieve users of an everyday decision. They should not have to use up cognitive power by digging through song lists and deciding on a song. Instead, they take a photo of themselves and a song that matches their musical taste, mood and environment is started instantly.

Potential users are defined as people who mainly use a music streaming service and want to reduce everyday decisions to counteract decision fatigue.

To use this application, users must open the corresponding website (intelligent-song-picker.com), authenticate themselves with Spotify, give permission to use the camera (on the first time), and click on the photo trigger button to take a picture of themselves. Once users are logged in, this process is even faster. The next time they open the page, they are then taken directly to the camera function and only have to click on the photo trigger. A prerequisite for using the ISP is that users use a device with a built-in front-facing camera, which means they have to be able to take pictures of themselves. Users must also have a Spotify Premium subscription to connect the app to the streaming service. Furthermore, the user's entire environment and mood should represent the situation as realistically as possible, i.e. remain unaffected.

4 PROTOTYPE DESIGN

4.1 Technical approach

Machine learning models are used to analyze a photo of users in their web browser. Based on the image analysis, a song is then selected and played on a music streaming platform – in this case Spotify.

This process is based on the assumption that our mood is significantly influenced by our surroundings and that it can be read in our face[7].

4.2 Conceptual design of the prototype

The ISP consists of four core components.:

(1) User authentication with Spotify to use the streaming service and access personalized information
(2) Camera application, embedded in the website to take photos of the user and his environment
(3) Image analysis, whose individual steps extract as relevant information as possible from the user's photo
(4) Song search to find the right title on Spotify
(5) Music player that maps the relevant functions like play/pause or mark songs as favorites

5 IMPLEMENTATION

5.1 Hardware and software environment

This project is using web technologies only. On the server-side there is a lightweight *Node.js* application that is responsible for authentication with Spotify and provides the frontend. The frontend on the client side is implemented in *Vue.js* a modern Single Page Application (SPA) Framework and compiled by *Webpack* – a powerful module bundler. The computationally intensive image analysis, was in part implemented with the machine learning framework *Tensorflow.js*. This as well as the communication with Spotify is outsourced to respective *web workers* that allow computation on multiple threads.

Everything else was implemented using native JavaScript and modern browser interfaces like the *MediaDevices API*[8]. In the GitHub repository[9] I created a deployment pipeline that allows me to automatically publish new project states on my server. Using different branches, I was able to test different versions and features and then discard or integrate them into the application which helped me a lot during prototyping.

5.2 Core of the implementation

At the heart of the implementation are the image analysis and song selection process.

During image analysis, parameters such as contrast, colorfulness and brightness of the photo are analyzed in addition to *face recognition* with the Blazeface Model[10] from Tensorflow.js and *emotion classification* with a CNN Model trained on the FER-2013 dataset[11].

Based on the image analysis resulting data the perfect audio parameters for the users are defined. This abstraction into categories such as *danceability, energy, mode, speechiness, accousticness, instrumentalness, liveness and valence* helps to better understand the effects of the image data on the music and thus enables a good mapping.

Once the ideal audio parameters have been defined, a multi-step process follows to find the perfect song for the user. First, the song whose audio parameters are closest to those calculated as optimal is selected from the user's 100 favorite songs. Based on this favorite song, 100 recommendations are then requested from Spotify, which Spotify assumes the user will also like. From these 100 songs, the one that best matches the optimal audio parameters is filtered out. This song is then played.

This process ensures that the selected song matches the user's preferences and mood, but is most likely not yet known to them.

5.3 Prototype scope

Server and client application for the prototype were fully implemented, including image analysis, song search and authentication with Spotify. In addition, I was even able to develop a design concept for the website, implement some SEO measures and tracking via Matomo to be able to publish the project and get a better insight into the usability and attractiveness of the solution. Through the integrated tracking, I can now evaluate my application permanently and after every adjustment. The project is publicly available at intelligent-song-picker.com.

In addition to what has been achieved, I still see potential for expansion in the image analysis. The emotion recognition model used should be viewed critically, as it was trained on the FER-2013 dataset which provides neither a representation of the faces our society offers, nor perfect labeling[12][13]. Furthermore, the algorithms for determining the image features could still be extended and thus the relevance of the results increased.

6 EVALUATION

6.1 Objectives of the evaluation and methodology

The important thing now is to find out if the ISP offers a useful alternative to the otherwise decision heavy selection of a song.

In addition to functional questions such as: "Do users understand the purpose of the application?" and "Do they understand what they have to do?", the main question of this evaluation is: "Does image analysis provide reliable data, or is the information that can be drawn from a photo insufficient to make an accurate statement about the mood of the users and the device they are interested in?"

6.2 Implementation of the evaluation

So, to find out if users like the suggested songs, if they fit the current mood, or if users would rather listen to something else, I integrated several controls whose usage I track via *Matomo* for evaluation. These control element are:

(1) Take new photo to get new song suggestion
(2) Pause song
(3) Save song to favorites on Spotify

These three interaction classes and their number of relative occurence give me an indication of how good the ISP's recommendations are.

So far, there is not enough reliable data from the website tracking to make a statement about the quality of the prototype. However, a few people could be asked directly about the project. In almost all cases I got the feedback that the suggested songs were a very good fit. Some songs were even saved for later. One person criticized during the questioning that he had not understood what actually happened, i.e. how the ISP proceeded. The usability of the ISP was generally described as very simple and intuitive.

6.3 Interpretation of the evaluation results

Spotify knows pretty well what music its users like. And whether a song really fits the current situation leaves a lot of room for interpretation. In my opinion, most songs cannot be clearly classified, so at least the basic mood usually fits. This also coincides with the consistently positive feedback from the test persons. In hindsight, the music seems to me to be quite a grateful subject for this project.

I had already expected the feedback regarding not understanding the functionality of the application, since the application is built as minimalistic as possible to live up to its claim of minimal cognitive effort. However, I added some more detailed information and explanatory texts about the project for users interested in them on an extra page.

7 SUMMARY

7.1 Achieved results

This paper addressed the question of whether image analysis using machine learning can be used to automate song selection and thus reduce decisions in the music selection process. To this end, I first developed a concept of how the technology can be used to set up a process with as few decisions as possible. I then created and published a technical prototype that minimizes the cognitive effort required for music selection. The ISP prototype consists of the image analysis and song selection process, which were created from assembled modules and my own concept and development.

Then the prototype was made available to individual users for testing and feedback was gathered from them. From that feedback I could already draw a fairly positive conclusion.

Subsequently, a larger evaluation was prepared by means of website tracking, which is intended to provide quantitative information about the quality of the song suggestions in order to determine whether the technology really is a good fit for this purpose.

7.2 Next steps

The next step is to improve the image analysis module. There is still a lot of potential hidden here and there are still many approaches as to how further information could be obtained from images. In addition, the models used so far should be revised so that they work equally well for all people and their use in a real product is ethically justifiable.

REFERENCES

[1] John Tierney. Do you suffer from decision fatigue? *The New York Times*, August 2011. URL https://chrissisdogtraining.com/wp-content/uploads/2014/12/decision-fatigue.pdf.

[2] David Hirshleifer, Yaron Levi, Ben Lourie, and Siew Hong Teoh. Decision fatigue and heuristic analyst forecasts. *Journal of Financial Economics*, 133(1):83–98, 2019. URL https://doi.org/10.1016/j.jfineco.2019.01.005.

[3] Drake Baer. The scientific reason why barack obama and mark zuckerberg wear the same outfit every day. *Business Insider*, 28, 2015. URL https://www.businessinsider.com/barack-obama-mark-zuckerberg-wear-the-same-outfit-2015-4?r=DE&IR=T.

[4] Mario Noioso. Mood music station based on emotions and spotify. 2017. URL https://marionoioso.com/2017/01/11/mood-music-station-based-on-emotions-and-spotify/.

[5] David Noah. Microsoft emotion and spotify "mood music" project [api smash]. July 2020. URL https://rapidapi.com/blog/microsoft-emotion-and-spotify-mood-music-project-api-smash/.

[6] Kalyssa A. Owusu. Building a playlist recommender with vue.js & face-api.js. *Medium*, June 2020. URL https://medium.com/@K_Lysssa/how-i-built-a-playlist-recommending-web-app-with-vue-face-api-js-d573102c890.

[7] Devi Arumugam and S Purushothaman. Emotion classification using facial expression. *International Journal of Advanced Computer Science and Applications*, 2(7), 2011.

[8] Mozilla and individual contributors. Mediadevices, December 2020. URL https://developer.mozilla.org/en-US/docs/Web/API/MediaDevices.

[9] Florian Kapaun. Github repository isp - intelligent song picker, January 2021. URL https://github.com/floriankapaun/intelligent-song-picker.

[10] Valentin Bazarevsky, Yury Kartynnik, Andrey Vakunov, Karthik Raveendran, and Matthias Grundmann. Blazeface: Sub-millisecond neural face detection on mobile gpus. *arXiv preprint arXiv:1907.05047*, 2019.

[11] Manas Sambare. Fer-2013, August 2020. URL https://www.kaggle.com/msambare/fer2013.

[12] Panagiotis Giannopoulos, Isidoros Perikos, and Ioannis Hatzilygeroudis. Deep learning approaches for facial emotion recognition: A case study on fer-2013. In *Advances in hybridization of intelligent methods*, pages 1–16. Springer, 2018.

[13] Emad Barsoum, Cha Zhang, Cristian Canton Ferrer, and Zhengyou Zhang. Training deep networks for facial expression recognition with crowd-sourced label distribution. In *ACM International Conference on Multimodal Interaction (ICMI)*, 2016.