Evaluating & Updating Music Genre Bars

Swaroop Panda, Shatarupa Thakurta Roy

Indian Institute of Technology Kanpur

Music Genre Bars are horizontally stacked bar charts that represent soft music genres. They have been deployed in a music streaming application allowing the user to search songs while adjusting genre proportions. In this work, we present a method to evaluate these deployed music genre bars. The evaluation basically consists of two parts; evaluating the genre bars in and of itself and evaluating the genre bars embedded in the streaming application. Both the parts of the evaluation include conducting case studies with the users. The insights obtained from the analysis of the results are then used for a design update of the deployed music genre bars.

Music Genre Bars

The authors in [1] proposed Music Genre Bars. Music Genre Bars are horizontally stacked bar charts that represent soft music genres. Soft music genres are weighted multiple genre annotations to songs; for example a song belongs to 20% blues, 30% country and 50% jazz instead of completely belonging to blues or country or jazz. In the genre bar, each stacked bar represents a particular genre.

These genre bars have been deployed in a music streaming application prototype. In the streaming application the genre bars are provided with sliders so as to allow the user to change the genre proportions by changing the length of the stacked bars (where the bar length is commensurate to the genre proportion). Changing genre proportions further results in finding the nearest corresponding songs from a given playlist using euclidean distance.

[1] Swaroop Panda and Shatarupa Thakurta Roy. Music genre bars. *arXiv preprint arXiv:2103.00129*, 2021
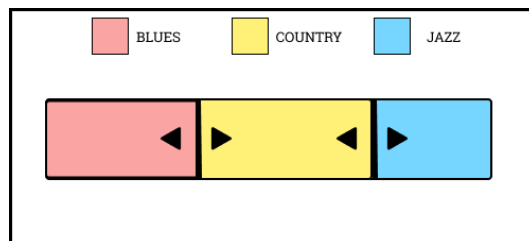

Figure 1: Music Genre Bars across three genres of Blues, Country, Jazz.

In this work we propose an evaluation method for these deployed Music Genre Bars. The evaluation is in two parts; evaluation of the genre bars in and of itself and and evaluation of the bars embedded in the music streaming application prototype. After analysis of the responses from the evaluations, the insights are incorporated into the streaming application prototype and the design of Music Genre Bars

is updated.

Evaluating Music Genre Bars

Evaluating music genre bars involves two components; evaluating the bars in and of themselves and evaluating the bars embedded and situated within the interface. This two part evaluation derives from the definition and the stated functionalities of the Music Genre bars; to a) represent genre proportions b) allow the user to change proportions within the streaming application.

Evaluating the Bars

The evaluation of these stacked bar charts involves conducting user studies. The basic idea of the user study is verify and ensure whether the users are able to identify proportions from the bars correctly. The user studies were basically digital surveys (containing a questionnaire) distributed among participants who were graduate students and researchers in different universities.

The questions in the questionnaire were specifically developed to ascertain whether users would correctly match a set of numerical proportions to their corresponding visual counterparts (in this case, horizontally stacked bar charts). The questions were directed in both ways; requiring the users to identify the correct visual charts from the numerical proportions and the numerical proportions from visual charts. More specifically the questions included, *"Choose a stacked bar chart that most likely represents the proportion A: 30%, B:40%, C: 30%."* and *"What proportion does the given stacked bar chart most likely represent?"*. The questionnaire was distributed and responses were obtained and analysed.

The questions used and the analysis of the responses in the survey were motivated from the ten primitive analysis task types from Amar et al. [2]. Amar et al. suggest that these ten low-level analysis tasks (Retrieve Value, Filter, Compute Derived Value, Find Extremum, Sort, Determine Range, Characterize Distribution, Find Anomalies, Cluster, Correlate) *capture people's activities while employing information visualization tools for understanding data*. In this case, the low level analysis would resemble correlating (*Given a set of data cases and two attributes, determine useful relationships between the values of those attributes.* as defined by [3]) the visual proportions with the numerical and vice versa. Low-level tasks are useful for the evaluation of the stacked bar charts because no specific (or high-level) tasks are specified at this stage. As we shall see in the next section, tasks only become specified when the stacked bar charts are embedded in a music streaming application.

[2] Robert Amar, James Eagan, and John Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 111–117. IEEE, 2005

[3] Robert Amar, James Eagan, and John Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 111–117. IEEE, 2005

Evaluating for low level tasks thus also enables other applications of the the genre bars.

From the survey questionnaire, it was discovered that most of the users could correctly identify proportions in both the directions. After a combined analysis of thirty-six respondents to the survey, it was discovered that around 96.7% of the respondents correctly matched numerical proportions with a given visual chart while 96.1% of the respondents successfully matched correct visual charts with a given numerical proportion. Percentages, from descriptive statistics, are used because there are no high-level or specific tasks (which could require statistically significant or other similar metrics) that the bars are supposed to perform. Thus the evaluation is via-negativa; ensuring that the genre bars are *not wrong* in conveying proportions to the users and are thus available for a variety of high-level or specific tasks.

Evaluating the Interface

This streaming application prototype, containing the genre slider (as described in [4]), is evaluated using user studies obtained from heuristic evaluation [5]. Heuristic evaluation presents a set of guidelines or rules of thumb for usability inspection and detecting "usable" problems in a user interface.

To validate the streaming application prototype a small user study with eleven users is conducted. The user study is developed in a usability study software called Maze. The user study basically requires users to complete certain actions or tasks and answer some questions on the given prototype using the software. These actions include moving the slider and changing the proportions of the genres while the questions were related to the overall usability of the slider and the interface. Specifically the users were asked increase or decrease the proportions and were asked for the affordance (how the object can be used) of the slider.

From the analysis of the user study results of the prototype, it emerged that the genre slider initially seemed intriguing to some users. This was ascertained from the variety of different answers to a relevant question. To the question of *What is the slider for?*, answers ranged from *"For changing songs"*, *"For sliding genres"*, *"changing sizes"* to *"For genre switching"*. In order to address this and to improve upon the prototype, the genre slider was labelled and annotated with the corresponding proportion and the genre so as to provide clarity behind the purpose and the intent of the slider and also to display the exact proportion of the respective genre.

The analysis also indicated that most of the users correctly discov-

[4] Swaroop Panda and Shatarupa Thakurta Roy. Music genre bars. *arXiv preprint arXiv:2103.00129*, 2021
[5] Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256, 1990

ered the affordance of the slider; with most users choosing to either scroll or drag the slider. The analysis indicated that most of the users on average (around 95%) also completed the genre changing tasks (as specified in the Maze software in the user study) successfully. This was based on the task completion analysis provided by the software. Other metrics from the Maze software included average duration for the task completion and misclick rate. Analysis and insights from these are potential topics for the next design update for the Music Genre Bars.
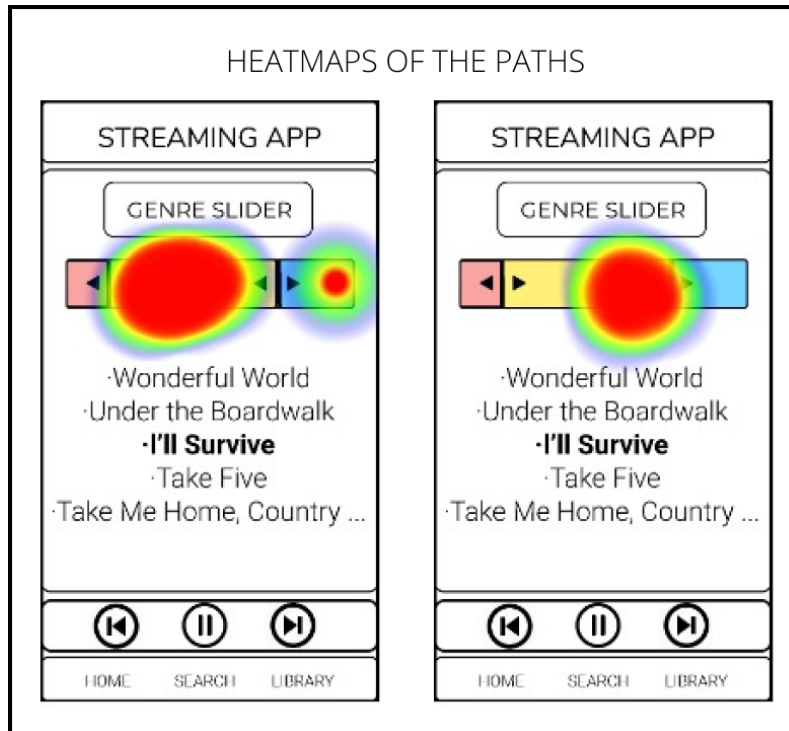

Figure 2: The heat maps of the user activities for modifying user proportions on the music streaming application from Maze.

A Design Update

We come to define a design update as a new version of an existing design that has emerged from incorporating insights from user evaluation (as in this case), or adding a new feature, or improving the visual form, performance or anything else.

The design update to the Music Genre Bars in the streaming application includes annotating proportions and the corresponding genres on the individual stacked bars; thus allowing the users to precisely view the proportions that are chosen.
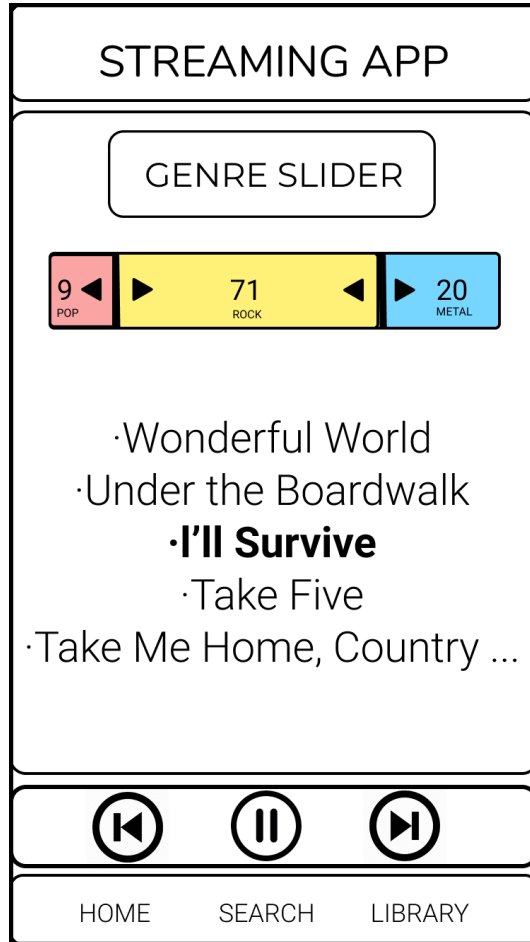
References

Robert Amar, James Eagan, and John Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 111–117. IEEE, 2005.

Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256, 1990.

Swaroop Panda and Shatarupa Thakurta Roy. Music genre bars. *arXiv preprint arXiv:2103.00129*, 2021.