Evaluation of a Mood-Based Graphical User Interface for Accessing TV Archives

Jana Eggink
Evaluation of a Mood-Based Graphical User Interface for Accessing TV Archives

Jana Eggink

Abstract

This paper describes a quantitative user study of a mood-based graphical interface, designed to access a large archive of BBC TV programmes. Programmes were automatically classified on two mood dimensions, ranging from serious to humorous and from slow to fast-paced. More than 3200 people participated in the trial over a time period of three months, having access to 2400 different TV programmes. Direct feedback was positive but very limited. An analysis of the user behaviour however showed that about one third of all users seemed to like the experience, as they accessed the GUI multiple times. The analysis also revealed that users interacted more with the graphical interface than with the text-based searches. Behavioural patterns were consistent with a use for entertainment purposes rather than professional research, indicating the suitability of the interface for its intended purpose of providing archive access for non-professional users.

This document was originally published in the Proceedings of the ICME International Workshop on Affective Analysis in Multimedia (AAM), San Jose, USA, July 2013. © IEEE

Additional key words: Graphical User Interfaces, Usability Studies, Mood, Multimedia classification, Archive access
Evaluation of a Mood-Based Graphical User Interface for Accessing TV Archives

Jana Eggink

1 Introduction
The British Broadcasting Corporation (BBC) is one of the largest public broadcasters and has archives reaching back to the beginnings of radio and television. By its 100th anniversary in 2022 the BBC aims to open up its audio-visual archive for public access. Search in such a large archive is difficult, especially for the non-professional user. Some manually created metadata exists, most TV programmes have one or more genre labels assigned, additionally actors, presenters and producers are listed for some programmes. Detailed semantic labels exist for most factual programmes, aimed at professional archivists and programme makers to enable efficient reuse of material.

For drama and entertainment programmes the available information is more limited, therefore a project has been started to investigate the use of additional mood-based metadata. Mood can provide intuitive understanding, and complement the existing programme information. As manual labelling of such a large archive is far too costly, an earlier project developed an automatic mood classification system [3]. Classification results using only features derived from the audio-visual content were very promising, so as a next step we wanted to test the usefulness of mood in a graphical user interface.

In the following we will describe the user interface in more detail, as well as the user group and the available test material. The user interactions with the mood-based GUI have been tracked, and will be analysed in the remainder of this paper. To the best of our knowledge, there are very few publications of comparable studies. The Sound and Vision archive analysed the search behaviour of media professionals using their text based interface [4]. The most frequent queries here were for specific titles of programmes, mainly news and current affairs. This is different to the behaviour expected from non-professional users interacting with a graphical interface, making it difficult to draw conclusions from a direct comparison. A mood based graphical interface is used in a commercial setting for photograph and video retrieval at Moodstream by Getty Images [5], or by for music by [6], but we are not aware of any published evaluations of usage. Work by [7] focused on affective classification and visualisation of music videos. While similar in some respects to our proposed system, they focused on a very specific type of video material, and did not evaluate their system in a realistic scenario, leaving it unclear if and how users would want to use their interface given the choice.

2 Experimental Setting
The mood-based prototype GUI was made available for web browser access internally at the BBC, using content recorded via the BBC Redux system. This system was set up in 2007 and since then continually records BBC channels and makes the programmes available to a subsection of BBC employees for research purposes and broadcast compliance checking [1][2]. A selection of older programmes from the BBC archive is also available within the Redux system.

The GUI used automatically obtained mood values, based on an analysis of the audio and the video stream of a programme and machine learning techniques as described in [3]. In addition to the graphical display, the interface provided multiple options for metadata searches, and allowed direct access to the videos for viewing and downloading. Programmes were represented by pink dots in a two-dimensional plane, when hovering with the mouse over a dot the corresponding programme information and a preview picture were displayed in the lower part of window. The default view was based on mood only, with serious to humorous on the x-axis and slow to fast-paced on the y-axis, see Fig. 1. The choice of mood was based on [3], where these two moods showed high user agreement, higher than the frequently chosen happy-sad axis, and little correlation between them. The pure mood view could be changed to a timeline view, where the
Programme in outlier position, see Section 7.

Fig. 1. Mood GUI, pure mood view

Erroneous alignment, see Section 7.

Fig. 2. Mood GUI, timeline view

Fig. 3. Video page, programme and series information

Fig. 4. Video page, user feedback
date of broadcast was displayed on the x-axis, and any of the two mood dimensions on the y-axis, see Fig. 2.

Clicking on a programme dot took the user to the video page, which had an integrated video player and gave download options. The lower half of the window provided a choice between user feedback, programme and series information, and a display of the time varying features used to calculate the mood values, mainly intended for research purposes. The default view was the programme and series information, see Fig. 3. When a video finished playing, the user pressed pause or attempted to leave the page, the view automatically switched to the feedback view, Fig. 4. In addition to collecting this active user feedback we also tracked usage of the GUI, i.e. all user clicks or other active selections and searches.

3 GENERAL USAGE ANALYSIS

3.1 User base and content selection

The mood GUI went live on the 14th May 2012, at first advertised only within the BBC Research and Development department, with additional promotions on a wider internal event a week later, followed again a week later by a direct link on the Redux homepage. We analysed the data up to and including August 22nd, always excluding team members and other test users.

The trial participants were not explicitly invited or otherwise selected. The mood GUI was presented as an alternative way to access content, in addition to existing text based interfaces which allow searches in traditional metadata such as title, synopsis and broadcast date. In this sense, the trial was collecting data 'in the wild', trying to evaluate if users, when given a choice, would like to use a mood based interface to find TV programmes.

All TV content included had to be processed for mood classification. The computational resources available did not allow processing all available material; we therefore started the trial with around 1200 programmes and successively increased the number to twice that much at the end of the trial period. The selection of programmes had to be performed manually and included a mix of programmes selected by broadcast date to include older material, some well-known series to provide interesting content, and a random selection based on availability for a representative archive view.

3.2 Number of users over time

The number of people who accessed the Mood GUI during the trial period was large with 3206 different users\(^1\). Nearly a third (1013 participants) of these were returning users, who accessed the GUI for at least a second time after a break of at minimum 24 hours.

When looking at the number of people accessing the GUI every day, a clear peak can be seen around the 28th and 29th of May, when the link was first made available on the Redux homepage, see Fig. 5. Initial curiosity was obviously high, but even after being online for three months, still around 30 to 50 people were accessing the GUI every day. When counting the number of distinct users per week, a stable demand with around 200 to 300 users can be observed even at the end of the analysed trial period, meaning that it was not the same small group of people using the GUI every day.

We can also see at what time of day people were using the GUI most, cumulated over the entire period, see Fig. 6. Somewhat surprisingly usage continued throughout the night, even

---

\(^1\) The total number of potential users, i.e. BBC employees with a valid Redux account, is not known to the authors.
though access numbers were low between 3am and 7am. We assume these might have been people working shifts, or travelling in different time zones. A small peak in demand can be seen over lunchtime, with the most usage in the evening, times when people are most likely looking for entertainment.

4 MOOD CHART USE

After analysing the number of people using the mood GUI, we evaluated their search and browsing behaviour. There were multiple ways users could interact with the GUI, such as adjusting the sliders to zoom to specific sections on the mood or timeline-mood view, or searching various metadata fields including title, genre, actors and broadcast date.

4.1 Individual interactions

For a first analysis we counted all interactions, including multiple actions of the same type (e.g. search by title) by a single user. Adjusting the sliders on the 2-dimensional graphical display accounts for 5090 out of 8006 events, the majority of all interactions, see Fig. 7. This might be caused to some extent by the fact that there are two sliders to adjust, increasing the number of interactions, and by multiple slider adjustments for zooming into specific areas.

The next most frequent action was a selection of the axis, changing from a pure mood view as in Fig. 1 to the timeline view as in Fig. 2 and back. The mood view with serious/humorous versus slow/fast-paced was the default view, the most frequent axis selection was a change to timeline versus humour (565 times), followed by timeline versus pace (306 times), and a return to the pure mood view (111 times). The most commonly used search field was title (759 times), followed by genre (656), while searches for people (280) or specific date ranges (239) were less common.

4.2 Interactions by number of users

The adjustment of the sliders often led to multiple consecutive actions by the same user. For another view on user behaviour we analysed how many individual users performed a specific action, i.e. counting each user only once per action. Slider adjustment is still the most commonly performed action, with 828 people having tried this feature, see Fig. 8. Overall, the relative popularity of interactions remains unchanged, with 541 people having used the axis selection at least once, and title and genre search being used by 374 and 289 different people respectively.

4.3 Metadata searches

We also looked at the searches within the provided metadata, pooling results by user, i.e. counting each user-search combination only once. All text-based search fields offered auto-complete options during typing, selecting possible completions based on all matching entries in the database. This had the advantage that searches could be matched without having to account for variations in spelling, although free text entries were also possible.

The most common title search was for Doctor Who, alternatively with the spelling of the old or new series (Doctor or Dr Who), followed by Mad Men and Horizon, see Table 1. Among the genre searches, 83 different people search for ‘films’, making it the most common search term, see Table 2. Even though overall searches for title were more frequent than those for genre, genre searches

\[\text{Note that the results should in no way be used to determine the absolute popularity of BBC programmes, especially as the archive is not yet complete.}\]
offered fewer options so that the numbers for individual search terms are higher. Other popular options for genre searches were comedy, drama and corresponding subgenres, followed by factual and documentary.

Table 1. Most frequent title searches

<table>
<thead>
<tr>
<th>Frequent Titles</th>
<th>Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor Who</td>
<td>21</td>
</tr>
<tr>
<td>Dr Who</td>
<td>16</td>
</tr>
<tr>
<td>Mad Men</td>
<td>12</td>
</tr>
<tr>
<td>Horizon</td>
<td>12</td>
</tr>
<tr>
<td>Family Guy</td>
<td>11</td>
</tr>
<tr>
<td>Sherlock</td>
<td>11</td>
</tr>
<tr>
<td>Have I Got News for You</td>
<td>11</td>
</tr>
<tr>
<td>Tomorrow’s World</td>
<td>11</td>
</tr>
<tr>
<td>Eastenders</td>
<td>9</td>
</tr>
<tr>
<td>Panorama</td>
<td>8</td>
</tr>
</tbody>
</table>

Generally, the main access and interaction with the mood GUI seemed to be via the graphical display, indicated by the frequent zooming via slider adjustment and switching between the pure mood and the timeline-mood view (which was much more frequently used than a direct search for specific date ranges). This indicates that people used the interface for browsing rather than specific searches, consistent with the intended use case of opening up the archives for non-professional users. The time of day when people most used the system with a strong peak in the evening and a small one during lunch time is also consistent with the use for entertainment rather than focussed research, as is the most common search for the genre ‘films’.

Table 2. Most frequent genre searches

<table>
<thead>
<tr>
<th>Frequent Genres</th>
<th>Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>films</td>
<td>83</td>
</tr>
<tr>
<td>comedy</td>
<td>42</td>
</tr>
<tr>
<td>drama</td>
<td>36</td>
</tr>
<tr>
<td>drama/scifiandfantasy</td>
<td>24</td>
</tr>
<tr>
<td>drama/crime</td>
<td>22</td>
</tr>
<tr>
<td>comedy/sitcoms</td>
<td>22</td>
</tr>
<tr>
<td>documentaries</td>
<td>20</td>
</tr>
<tr>
<td>factual</td>
<td>12</td>
</tr>
<tr>
<td>drama/thriller</td>
<td>12</td>
</tr>
<tr>
<td>documentary</td>
<td>12</td>
</tr>
</tbody>
</table>

5 VIDEO PAGE USE

Analysing the usage of the video page gives more information about the actual viewing behaviour of trial participants. This webpage appeared after a user selected an individual programme episode
by clicking on a pink dot in the mood GUI and allowed watching or downloading the content directly from Redux\textsuperscript{3}, see Fig. 3.

Both play and pause events were very frequent, with around 6500 and 5500 events respectively, making up the basic functionality of the video page. Nearly 250 programme selections were based on selecting another episode from the same series as the current one, a feature that was introduced about two weeks after the initial mood GUI went online. In total, about half of the programmes available at the end of the trial were watched at least once (1204 out of 2429 programmes).

5.1 Most Popular Programmes

The most frequently watched programme was an episode of ‘Never Mind the Buzzcocks’, watched by 258 different users, followed by an episode of ‘Torchwood’, see Table 3. Surprisingly, the 5\textsuperscript{th} most watched programme was an old episode of ‘Holiday Weatherview’. It seems somewhat unlikely that nearly 50 people were interested in watching the same outdated weather forecast. A closer look at the graphical display shows that this programme had been classified as the most serious one of all programmes, giving it a distinctive and highly visible location on the mood chart, see the highlighted dot in Fig. 1. Similar outlier positions were occupied by all of the top three most watched programmes, although at least ‘Never Mind the Buzzcocks’ seemed to have a general appeal beside this, as multiple episodes appear in the top 10 list. Clearly, location on a graphical interface influences people’s viewing behaviour. Even if they don’t necessarily watch the entire programme, their attention is directed by the position of individual programmes.

Table 3. Most frequently watched programmes, based on individual episodes

<table>
<thead>
<tr>
<th>Frequent Programmes</th>
<th>Watched</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never Mind the Buzzcocks</td>
<td>258</td>
</tr>
<tr>
<td>Torchwood</td>
<td>90</td>
</tr>
<tr>
<td>Dr Finlay’s Casebook</td>
<td>74</td>
</tr>
<tr>
<td>An Evening in with David Attenborough</td>
<td>55</td>
</tr>
<tr>
<td>Holiday Weatherview</td>
<td>49</td>
</tr>
<tr>
<td>Would I Lie to You?</td>
<td>46</td>
</tr>
<tr>
<td>Never Mind the Buzzcocks</td>
<td>36</td>
</tr>
<tr>
<td>Morecambe and Wise</td>
<td>33</td>
</tr>
<tr>
<td>Never Mind the Buzzcocks</td>
<td>32</td>
</tr>
<tr>
<td>Till Death Us Do Part</td>
<td>32</td>
</tr>
</tbody>
</table>

5.2 Archive material versus new programmes

The project was started with the goal in mind of opening up the archives. There was no prompting of the users towards archive or new material, so we could analyse the users’ choices assuming they reflect the users true preferences. We divided the programmes into pre- and post-Redux material, broadcast before or after 2007 respectively. This division was based on the broadcasting and not the production date; archive material re-broadcast after 2007 would therefore be classified as ‘new’.

In total, 1057 archive and 1371 new programmes were accessible via the mood GUI. We define a programme as being watched by a user after at least one ‘play’ event. As a programme could be watched by multiple people, we could have more ‘watched’ events than programmes available. Both the archive and the new programmes were watched with similar frequency, in both cases the number of user-programme watching events is nearly 60\% of the number of available

\textsuperscript{3} Note that due to a programming bug information about downloads was not correctly saved and therefore we will be analysing online viewing behaviour only.
programmes. This shows a clear interest in pre-2007 programmes, also indicated by the relatively frequent title search for classic comedies like ‘Morecambe and Wise’.

6 USER FEEDBACK

After each programme viewing event, on pause or when leaving the video page, the user was asked to give feedback. This included a free text field for comments, or to suggest additional moods for a programme, as well as sliders to give mood ratings based on their personal perception and a like or dislike rating. Despite the large number of people having accessed the GUI, the amount of feedback was very limited. Only 45 users left any feedback at all, they provided new moods or like ratings for 58 programmes and 11 free text entries. The free text feedback included comments about the mood navigation experience in general, or specific aspects of the UI, all were positive.

Some additional mood suggestions for particular programmes were given, all mentioned only once: curiosity, tension, bored, uninterested, happy, engaged, nostalgic, surreal, sadness, anger, horror, camp, pessimistic. Some suggestions like happy seem were similar to the existing mood dimensions, i.e. humorous. Others, like surreal or horror, seem to describe different perceptual dimensions. None of these were mentioned frequently, so that additional moods were not added to the interface.

7 CONCLUSIONS AND FUTURE WORK

The first and most important conclusion is that people were interested and enjoyed using the mood-based GUI. About one third of all users returned, so while the GUI obviously wasn’t perfect for everyone, it still has a potentially very large user base. This is especially encouraging as the mood GUI only provided access to a small selection of programmes from the Redux system, while other text-based interfaces allowed access to the entire content. Some of the non-returning users might have been disappointed by the fact that their favourite programmes were missing, but this would require further research.

Based on the usage analysis of the mood GUI, we conclude that people used it primarily to find programmes they wanted to watch for entertainment, and for this a graphical interface that facilitates browsing can be useful. People interacted more with the graphical interface than with the traditional text-based searches, and when using the textual searches most frequently looked for the genre ‘films’. The time of day when the GUI was most frequently used also points towards use for entertainment, rather than for professional research as needed in programme production.

An important point is that people did not provide active feedback, even though the GUI automatically switched to the feedback page after each video viewing event. This means that it is absolutely necessary to collect implicit user feedback and track their usage behaviour for any type of quantitative analysis.

In terms of GUI design we could see that the placement of individual programmes on the two-dimensional mood interface had a noticeable influence on people’s viewing choices. Outliers clearly attracted interest, a fact that is worth considering when designing any type of graphical interface that places items in a two or three dimensional space.

No comments were made about the mood classification not being accurate enough. An early version of the feature extraction and classification code contained a bug that caused a number of programmes to be classified at near identical values of approximately 1.74 on the serious/humorous scale, leading to visible ‘straight line’, see Fig. 1. Colleagues within the research and development department commented on this; if any of the other users noticed this slightly odd alignment of programmes they did not provide any feedback about it.

Overall, the mood-based GUI received positive feedback and attracted a large number of users, providing proof of concept that mood can be a useful way of interacting with audio-visual archives. Part of our future work will be to evaluate the usability when a larger number of programmes is integrated. We are also planning to enlarge the potential user base by providing an alternative mood-based interface to the BBCs catch-up service, the iPlayer.
8 REFERENCES
http://www.bbc.co.uk/blogs/bbcinternet/2008/10/history_of_the_bbc_redox_proje.html,
accessed 19.11.2012
[3] J. Eggink and D. Bland: “A Large Scale Experiment for Mood-Based Classification of TV
Programmes”, Proc. IEEE Int. Conf. on Multimedia and Expo, ICME2012, Melbourne,
Australia, July 2012
Professionals at an Audiovisual Archive: A Transaction Log Analysis”, Journal of the American
Society for Information Science and Technology, vol. 61 no. 6, pp. 1180-1197, June 2010