Personalized Retrieval and Browsing of Classical Music and Supporting Multimedia Material

Marko Tkalčič
Free University of Bolzano
Faculty of Computer Science
Piazza Domenicani 3
Bolzano, Italy
marko.tkalcic@unibz.it

Markus Schedl
Johannes Kepler University
Department of Computational Perception
Altenbergerstrasse 69
Linz, Austria
markus.schedl@jku.at

Cynthia C. S. Liem and Mark S. Melenhorst
Delft University of Technology
Mekelweg 4
Delft, The Netherlands
c.c.s.liem@tudelft.nl
m.s.melenhorst@tudelft.nl

ABSTRACT

This paper reports on three demonstrators developed to provide personalized, enhanced experiences of classical music, within the EU FP7 project "Performances as Highly Enriched aNd Interactive Concert eXperiences" (PHENICX)[1]: (i) an interface for accessing supplemental multimodal material about music items, (ii) a recommender system for visualizations of classical music, and (iii) a recommender system for tagging. The personalization in all three demos is achieved through modeling users' personality and musical sophistication. The links to the web interfaces are provided as well as the outcomes of quantitative and qualitative evaluations.

Categories and Subject Descriptors

Information systems [**Information retrieval**]: Music retrieval; Information systems [**Information retrieval**]: Recommender systems

1. INTRODUCTION

Classical music consumption, especially live performances, did not change much over the recent decades, despite the development of new technologies. The share of consumption on mobile devices and the buzz on social media is very limited [14]. In order to provide the listeners of classical music an with an enhanced experience that new technologies can offer (such as additional information and personalization) and lower the barriers for engagement with classical music [9], we performed a series of studies and implemented a set of demonstrators, presented here. These demonstrators are also targeted at attracting new audiences to classical music.

2. RELATED WORK

In order to implement personalized applications, such as adaptation of user interfaces [10], for end users the system should be able to acquire the user preferences. Implicit acquisition of such data in the domain of classical music has shown to be hard as users hardly leave any traces on their preferences about classical music on social media [14]. However, related work suggests that preferences are related to two proxy constructs, personality and music sophistication. Several studies have shown that preferences for music are correlated with user's personalities in terms of the five factor model (FFM) [13]. The FFM is a well-known model of personlity that takes into account individual differences in many areas [7]. Furthermore, algorithms for the implicit acquisition of personality from social media have been proposed [2]. The concept of music sophistication has been presented recently and it has shown good correlation with music preferences [11]. In an earlier study we have shown that both personality and music sophistication are correlated with the user preferences for supporting multimodal material for classical music [15].

3. DEMONSTRATORS

We present three demonstrators: (i) a personalized interface for accessing supplemental multimodal material, (ii) a recommender system for the visualization of classical music and (iii) a recommender system for tagging of classical music.

The PHENICX project's long-term goal was to develop services that acquire the user characteristics (personality and music sophistication) in an unobtrusive way, which appears to be supported by related work [2]. However, for the demonstrators shown here we acquired these characteristics through questionnaires. As the questionnaires used are lenghty (10 questions used in the TIPI (Ten Items Personality Questionnaire [4]) personality questionnaire, 44 questions used in the BFI (Big Five Inventory [7]) personality questionnaire and 38 questions used in the MSI (Music Sophistication Index [12]) questionnaire), we performed a set of initial studies where we identified the questionnaire items (i.e., questions) that are most informative. We set the number of these questions to two or three, in accordance to user experience guidelines. The choice of the questions is described in the following subsections on a per-demonstrator basis.

[1]`http://phenicx.upf.edu`

WOODSTOCK '97 El Paso, Texas USA

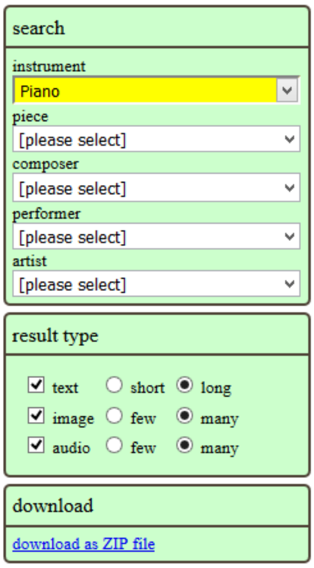

Figure 1: Part of the personalized user interface for the retrieval of supporting material.

3.1 Access to Supplemental Multimodal Material

This demonstrator is a web interface[2] that allows to retrieve supporting information about the composers, pieces, instruments and performers of a classical performance. The results of a query are presented in the form of text, images and audio.

The query is defined through a set of dropdown menus, as shown in Fig. 1. The search query itself (the upper frame of the user interface) is not personalized. The personalization affects the result type preferences (the result type frame in the user interface). More specifically, for the two multimodal types analyzed (i.e. text and image) the personalization affects the length and amount of the results. Some users prefer long text over short and some prefer many images over few. For example, people who score high on openness, agreeableness, conscientiousness and extraversion tend to show a positive correlation with consumption, interestingness and novelty, hence meaning that they prefer to consume more of the material, find it interesting and novel.

The choice of the two questions was done by observing how the user answers to the music sophistication and personality questions correlate with the preferences for the content. We observe that among the music sophistication questions the one with the highest absolute value of the correlation was *How many classical concerts do you attend per year?*. Among the personality questions the one with the highest absolute correlation was *I myself as someone who is reserved, quiet.*

In order to personalize the web interface for the supporting multimedia material, we implemented a recommender system that adapts the result types (i.e. the length of the text and the number of the images shown) to the end user. The personalization is done by recommending the result types that are most popular in the cluster of users the active user belongs to, where clusters are created by categorizing users according to their answers to the two selected questions.

In order to evaluate the satisfaction of users with the recommended settings, we performed a user study. The users were asked to rate the personalized interface through the question *I like the way the material is presented.* In the next step, the participants were shown a random snapshot of the interface and were asked to provide a rating. In the final step the participants were shown both snapshots next to each other and were asked to select the one they preferred. In total we recruited 96 participants through Amazon Mechanical Turk.

The results have shown that in 63 (out of 96) of the cases the participants preferred the personalized snapshot, which corresponds to 66%. The higher preference for the personalized interface was confirmed also in the comparison of the ratings given to the two snapshots. The mean rating for the personalized snapshot was 3.29 (on a scale from 1 to 5, where 1 was the lowest and 5 the highest score) compared to the mean rating for the random snapshot of 2.92. The Wilcoxon signed-rank test [3] showed that the mean difference was significant ($p = 0.01$).

3.2 Recommendation of Visualizations

This demonstrator[3] recommends a music visualization to the user in the form of a streamed video within the web browser. In order to enable users to get additional insight into music, a set of music visualizations (called Score Follower, Structure Layout and Orchestra Layout) have been developed within the PHENICX project. The visualizations are depicted in Figs. 2, 3 and 4, respectively. In the initial user study we observed that people with different personalities tend to prefer (in terms of pragmatic quality (PQ) scores, as defined by [5]) different visualizations.

Figure 2: The *Score Follower* visualization.

Similarly to the first demonstrator, we recommended to each user the visualization that has the highest average rating among the users that are in the same cluster. To this end we (i) choose the BFI questions for user clustering and (ii) ranked the visualizations within each user cluster.

The two BFI questions that account for most of the variance in the sample of users from the initial study were *I see myself as someone who is helpful and unselfish with others* and *I see myself as someone who tends to be disorganized.* We clustered the users into four clusters based on the median value along each variable.

After answering the two initial questions the user is shown a page where the three visualizations are presented in such

[2]http://bird.cp.jku.at/phenicx_mmsupp

[3]http://bird.cp.jku.at/phenicx_visrecsys

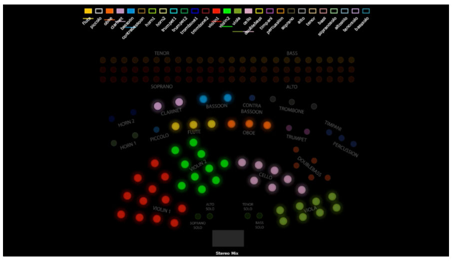

Figure 3: The *Orchestra Layout* visualization.

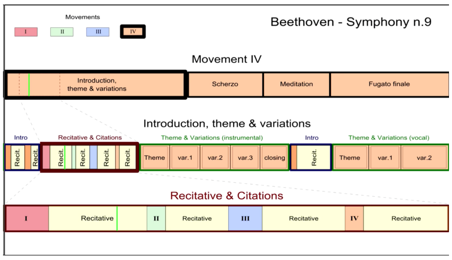

Figure 4: The *Structure visualization.*

a way that the top-ranked is in the central position and the other two are rendered smaller at the bottom of the interface (see Fig. 5). The user can switch from the recommended visualization to any of the other two.

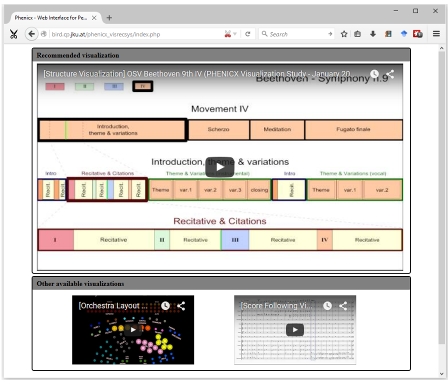

Figure 5: Rendering of the visualizations with the top ranked visualization bigger and positioned on top. The second ranked is positioned bottom-left and the third ranked is bottom-right.

We performed the evaluation of the recommender system for visualization through a user study. In total we had 79 participants recruited through Amazon Mechanical Turk. The participants were shown the snapshots of the three visualizations in a row next to each other. The order was randomly assigned for each participant. Then the participants were instructed to watch each visualization for at least 20 seconds. Finally, the participants were asked to rank the visualizations according to their preferences.

We computed the normalized discounted cumulative gain (NDCG) to assess the quality of the personalized rankings. We compared the mean NDCG of the recommended visualization rankings ($ndcg\_rec$) to the mean NDCG of randomly ranked visualizations ($ndcg\_rand$). The obtained values were $ndcg\_rec = 0.87$ and $ndcg\_rand = 0.82$. The independent t-test showed that the difference of the means was significant ($p = 0.03$). The significant difference means that our system's better average NDCG metric was not due to chance, but due to a systematically better prediction of the personalized recommender for visualization.

3.3 Tag Recommender System

The idea in this demonstrator[4] is to allow the user to capture certain moments by tagging the music during or after the concert. These tags represent the personal impressions of the users and it is possible to share them. The sharing is done through a recommender system that suggests the tags of similar users.

Tag recommender systems are popular in social media. The most widely used approach is to recommend tags that similar users have provided [8]. Various user similarity measures are used in the related work, depending on the domain and data available [8, 6]. We employ a variant of the classic K-nearest neighbors (K-NN) approach (see also [1]) for collaborative filtering to recommend tags similar users have provided, where similarity is computed based on personality.

The steps of the recommender system are to (1) compute the likelihood that a user would enter a tag at a given time in the piece, (2) rank these tags and (3) present them to the user.

In the initial study we choose two questions from the BFI questionnaire that correlate best with the *active engagement* factor of music sophistication. Two two top ranked questions were *I see myself as someone who is sophisticated in art, music, or literature* and *I see myself as someone who has an assertive personality.* Again, we clustered the users into four clusters based on the median value along each of the two variables. The likelihood of a user generating a tag at a given time is calculated by taking into account the tags that similar users have given around the same time. The user similarity measure is a binary function that returns one if the users are in the same cluster, zero otherwise. We set the time window to 20 seconds.

The top three recommended tags are shown as clickable buttons in the user interface. The user can, at any given time, submit a tag either by entering it through the textbox (and press Enter) or click on a button with a recommended tag. The recommended tags change through time as the music piece progresses.

In order to evaluate the tag recommender system, we performed a user study in two steps. In the first step we collected free-form tags. In the second step these tags were used as a basis for the recommendations. In the first step the participants were shown the textbox only, without any recommendations. In total we had 22 participants recruited through Amazon Mechanical Turk. We choose a 2-minute-long segment of the Eroica streamed from YouTube.[5] Each participant had to enter at least 5 tags. After data cleaning there were 128 valid tags (5.8 tags per user on average).

A qualitative analysis of the tags showed that the participants gave either (a) tags related to the performance or (b)

[4] `http://bird.cp.jku.at/phenicx_tagrecsys`

[5] `https://www.youtube.com/embed/gr2Hrq_GE68?enablejsapi=1&vq=hd1080`

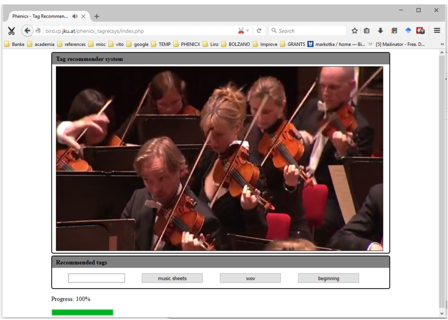

Figure 6: The interface for tagging classical music is composed of the video of the classical music piece (top) and the tagging interface (bottom) with the textbox for freetext tagging and the recommended tags.

tags related to the participant's perception.

In the second step we enabled the recommendations. The participants (not the same as in the previous step) were divided into two groups. The first group received recommended tags according to the presented algorithm, while the second (control) group received random recommendations from the pool of all tags collected in step one. In total we recruited 31 participants for the first group and 26 participants for the control group.

The results show that the average satisfaction was higher in the first group that received tags recommended by our system (3.81) than in the second (3.26). The t-test showed that the difference was significant ($p = 0.008$).

Furthermore we analyzed the shares of the number of tags given through the textbox (custom tags) and those given through a button (recommended tags). Interestingly, in the first group, which was given recommended tags, this share is higher (46%) than in the second group, which was given random tags (37%). It seems that the recommended tags inspired participants to come up with additional tags.

We investigated whether the position of the button plays a role in the preferences. In the first group (recommended tags) the distribution of clicks for the left, mid and right buttons was 46, 40 and 27, respectively, while for the second group (control group) it was 57, 41 and 35, respectively. This suggests that the location of the recommended items plays a role.

Finally, we carried out a qualitative analysis of the tags in the first group (recommended tags) to see what was the relation between the participants' personality and the type of tags they were giving. We coded each tag as *performance-related* or as *perception-related*. Among all users the *performance-related* tags account for 80%, while this figure is different for each user cluster, which indicates that users with different personalities tend to prefer different kinds of tags.

4. CONCLUSION

In this paper, we have presented three demonstrators that show how to personalize different aspects of classical music-related information to the user requirements. The demos have also been evaluated through a set of user studies that confirmed the benefit of the personalized systems over non-personalized ones, from the user experience perspective.

5. ACKNOWLEDGMENTS

This research is supported by the EU-FP7 project no. 601166 ("PHENICX") and the Austrian Science Fund (FWF): P25655.

6. REFERENCES

[1] C. Desrosiers and G. Karypis. A Comprehensive Survey of Neighborhood-based Recommendation Methods. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 107–144. Springer US, Boston, MA, 2011.

[2] G. Farnadi, G. Sitaraman, S. Sushmita, F. Celli, M. Kosinski, D. J. Stillwell, S. Davalos, M.-F. Moens, and M. D. Cock. Computational Personality Recognition in Social Media. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research (UMUAI)*, (Special Issue on Personality in Personalized Systems), 2016.

[3] A. Field, J. Miles, and Z. Field. *Discovering Statistics Using R*. Sage Publications, 2012.

[4] S. D. Gosling, P. J. Rentfrow, and W. B. Swann. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6):504–528, dec 2003.

[5] M. Hassenzahl. The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, 19(4):319–349, dec 2004.

[6] R. Jäschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme. Tag Recommendations in Folksonomies. *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'07)*, 4702:506–514, 2007.

[7] O. John and S. Srivastava. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin and O. P. John, editors, *Handbook of personality: Theory and research*, volume 2, pages 102–138. Guilford Press, New York, second edition, 1999.

[8] L. B. Marinho, A. Nanopoulos, L. Schmidt-Thieme, R. Jäschke, A. Hotho, G. Stumme, and P. Symeonidis. Social Tagging Recommender Systems. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 615–644. Springer US, Boston, MA, 2011.

[9] M. S. Melenhorst and C. C. S. Liem. Put the Concert Attendee in the Spotlight. A User-Centered Design and Development Approach for Classical Concert Applications. *ISMIR 2015*, pages 800–806, 2015.

[10] V. G. Motti and J. Vanderdonckt. A computational framework for context-aware adaptation of user interfaces. In *IEEE 7th International Conference on Research Challenges in Information Science (RCIS)*, pages 1–12. IEEE, may 2013.

[11] D. Müllensiefen, B. Gingras, J. Musil, and L. Stewart. The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one*, 9(2):e89642, jan 2014.

[12] D. Müllensiefen, B. Gingras, L. Stewart, and J. Ji. Goldsmiths Musical Sophistication Index (Gold-MSI) v1.0: Technical Report and Documentation Revision 0.3. Technical report, University of London Goldsmiths, 2013.

[13] P. J. Rentfrow and S. D. Gosling. The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, 84(6):1236–1256, 2003.

[14] M. Schedl and M. Tkalčič. Genre-based Analysis of Social Media Data on Music Listening Behavior. In R. Zimmerman and Y. Yu, editors, *Proceedings of the First International Workshop on Internet-Scale Multimedia Management - WISMM '14*, pages 9–13, New York, New York, USA, 2014. ACM Press.

[15] M. Tkalčič, B. Ferwerda, D. Hauger, and M. Schedl. Personality Correlates for Digital Concert Program Notes. In *UMAP 2015, Lecture Notes On Computer Science 9146*, volume 9146, pages 364–369. 2015.