Intelligent User Interfaces for Social Music Discovery and Exploration of Large-scale Music Repositories

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

ABSTRACT
In this position paper, we address the question of how to make music search and discovery more appealing, more exciting, and more joyful. In particular, we argue to research methods that foster serendipitous encounters with music items and to integrate ways for social interaction while exploring music collections and discovering the gems in today's huge catalogs available through online streaming platforms. We identify two major challenges here: the need for (i) highly efficient clustering and information visualization techniques that scale to these music catalogs and (ii) novel user interfaces that explain the clustering of music items and provide means to make the exploration of music a social event.

ACM Classification Keywords
Information Systems: Information Interfaces and Presentation: User Interfaces; Information Systems: Information Interfaces and Presentation: Sound and Music Computing

Author Keywords
intelligent user interfaces; music search; music browsing; music discovery; interaction

INTRODUCTION
The lasting trend of online digital streaming (e.g., *Youtube*), web radio (e.g., *Pandora*), and automatic playlist generation (e.g., *Spotify*) has enabled listeners to access virtually all (Western) music in the world. While the research field of music information retrieval (MIR) is concerned with many exciting tasks, ranging from audio fingerprinting to score following to extracting musically meaningful concepts such as rhythm, harmony, or timbre, a large part of industrial research and development in the area of music retrieval is currently focused on recommender systems. This fact is evidenced, for instance, by the enormous success of respective benchmarking initiatives such as the *Yahoo! Music*-sponsored *KDD Cup 2011* [7] and the *Million Song Dataset Challenge* [24].

HUMANIZE'17, March 13, 2017, Limassol, Cyprus.

ACM ISBN 978-1-4503-4905-5/17/03 ...$15.00.
http://dx.doi.org/10.1145/3039677.3039678

While music recommender systems are booming, approaches to access music repositories by other means, in particular, browsing via visual and aural user interfaces, have received less attention in the past few years. Therefore, we propose to research **browsing and interaction methods that make the process of discovering new music more exciting and joyful**, particularly by fostering interesting and serendipitous musical experiences and by social exploration of music, together with friends, family, or acquaintances. To this end, we consider vital, and elaborate in the following, two research directions: (i) improving algorithms for clustering huge collections of music items and (ii) improving user interface to browse large-scale music repositories.

CHALLENGES
To reach this aim, we are facing two major challenges. First, we need to **scale unsupervised learning (clustering) and information visualization techniques to real-world-sized music repositories of millions of tracks**, which are common in today's music streaming services. Since the amount of music items that can be processed by existing research prototypes for music browsing is in the range of at most a few tens of thousands, solving this challenge will require extensive research on optimizing and enhancing these techniques. Second, we need to elaborate means for **social and collaborative exploration of music collections**. Researching and implementing them will lead to improvements in the music browsing experience. To the best of our knowledge, existing music browsing interfaces barely provide such collective experiences. We therefore need to research strategies for listener motivation, engagement, and interaction between listeners and between listeners and music items. In doing so, we should also review and borrow from current reearch in the areas of game theory and gamification.

In the following, the two main challenges introduced above are discussed. First, an overview about relevant and current literature is given, followed by a brief research proposal, in which we indicate possible research directions to address the respective challenge.

Clustering Millions of Music Items
Research on automatic clustering of similar data items, which are typically represented as high-dimensional feature vectors, has yielded a variety of well-established clustering techniques, including k-means [25], neural network approaches

such as self-organizing maps (SOM) [17, 18], spectral methods such as principal components analysis (PCA) [12, 14], or multidimensional scaling (MDS) [5, 2]. More recently, a variety of non-linear manifold learning techniques has been proposed. These techniques assume that the high-dimensional input features lie on a non-linear submanifold of the Euclidean space. Respective methods aim at mapping the input data to a two- or three-dimensional output space while preserving the neighborhood structure of the items on the manifold. State-of-the-art techniques include isometric feature mapping (Isomap) [39], locally linear embedding (LLE) [31], and Laplacian Eigenmaps [1]. Isomap aims at preserving geodesic distances between data items on the manifold, where these distances are approximated by Euclidean distances for neighboring items, or accumulations of Euclidean distances among the shortest path between faraway items. LLE exploits local symmetries of linear reconstructions of data items by first modeling each data point in the input space as a linear combination of its nearest neighbors and subsequently using the linear coefficients to determine an output configuration that optimally reconstructs the data items. Laplacian Eigenmaps aim to preserve local intrinsic geometry. A weighted adjacency graph for neighboring data items is constructed and the output configuration is found by spectral decomposition of the corresponding graph Laplacian. Other recent techniques include probabilistic clustering methods such as t-distributed stochastic neighborhood embedding (t-SNE) [42], which models data items in the input space by Gaussians and in the output space by Student's t-distributions. The algorithm then tries to minimize the Kullback-Leibler divergence between the joint probabilities of data items in the input space and in the output space. T-SNE has particularly been developed to visualize high-dimensional data by alleviating the crowding problem that affects the projection of high-dimensional feature vectors to two or three dimensions.

Research Proposal

While there exists a variety of different high-dimensional data projection, clustering, and visualization techniques, little attention has been given so far to develop methods that scale to large item collections, even less to the very task of analyzing and adapting existing methods to deal with specificities of data used in the MIR domain. Addressing these issues, in our opinion, research should focus on elaborating methods that can deal with data points of high dimensionality, which are particularly present in computational audio- and web-based music features, as well as with data sets containing large amounts of items, which is common in online music catalogs. To provide some examples, typical audio-based features that describe properties such as timbre or rhythm of a music piece easily consist of feature vectors reaching thousands of dimensions [37]. In web-based music context modeling, which usually describe music items by term weight vectors, the dimensionalities are even higher, often reaching hundreds of thousands [36].

We therefore require (i) scalable dimensionality reduction and feature selection techniques for feature preparation, (ii) efficient means to estimate similarity between data items, which is key for employing machine learning on a large scale, and (iii) data projection and clustering techniques that improve over the current state of the art in terms of clustering quality and computational complexity. Furthermore, these techniques should also be able to deal with the inhomogeneity of music-related features, in particular when comparing audio-based descriptors to features mined from web pages or social media.

Interfaces for Social Music Exploration

User interfaces to explore a music collection typically apply some kind of metaphor to create a visual representation of the collection. This metaphor is often built upon a map-based visualization, created from data projection or clustering of music items according to music descriptors. One of the most popular is the *Islands of Music* metaphor introduced by Pampalk in [28], which employs a SOM for clustering and visualizes dense clusters as islands and mountains, whereas sparsely populated regions of the map are depicted as oceans. Quite a few later works follow and extend this idea. In [32], Schedl proposes a hierarchical extension of the two-dimensional map. An aligned SOM is used by Pampalk et al. in [29] to enable a seamless shift of focus between clusterings created for different musical aspects, for instance, between a SOM created only on rhythm features and one created only on timbre features. Neumayer et al. propose a method to automatically generate playlists by drawing a curve on the map [27]. Leitich and Topf propose in [20] the *Globe of Music* interface, which maps songs to a sphere instead of a plane by means of a GeoSOM [44]. Mörchen et al. [26] employ an emergent SOM and the U-map visualization technique [40] to color-code similarities between neighboring clusters. Vembu and Baumann incorporate a dictionary of musically related terms to describe similar artists [43]. In addition to the above approaches that create two-dimensional maps, three-dimensional browsing interfaces that employ SOM variants are also available. The *nepTune* interface presented by Knees et al. in [16] enables exploration of music collections by navigating through a three-dimensional artificial landscape, like in a computer game. A version for mobile devices is proposed by Huber et al. in [13]. A hierarchical variant that employs a growing hierarchical self-organizing map (GHSOM) [6] automatically structures the music collection into hierarchically interlinked individual SOMs [35]. Lübbers and Jarke present in [23] a browser employing MDS and SOM to create 3-D landscapes. In contrast to the *Islands of Music* metaphor, they use an inverse height map, meaning that agglomerations of songs are visualized as valleys, while clusters are separated by mountains. Their interface further enables the user to adapt the landscape by building or removing mountains, which triggers an adaptation of the underlying similarity measure.

While the above works employ some variant of a SOM, another common data projection technique in this context is MDS. For instance, the *MusicGalaxy* interface [38] by Stober and Nürnberger applies a pivot-based MDS. The authors integrate multiple sources of similarity information, giving the user control over their influence. In the *Search Inside the Music* application [19], Lamere and Eck use a three-dimensional MDS projection. Their interface provides different views that arrange images of album covers according to the output of the MDS, either in a cloud, a grid, or a spiral. Lillie applies a linear variant of MDS to project multi-dimensional music descriptors to a plane in the *MusicBox* framework [22]. The user can further create playlists by selecting individual tracks or by drawing a path in the visualization space.

Vad et al. [41] apply t-SNE [42] to mood- and emotion-related descriptors, which they infer from low-level acoustic features. The result of the data projection is visualized on a two-dimensional map, around which the authors build an interface to support the creation of playlists by drawing a path and by area selection.

Other recent approaches visualize music over real geographical maps, rather than computing a clustering based on audio descriptors. For instance, Celma and Nunes extract location information, such as place of birth, from *Wikipedia* to place artists or bands on a world map [4]. Govaerts and Duval extract geographical information from biographies and integrate it into a visualization of radio station playlists [9]. Hauger and Schedl extract listening events and location information from microblogs and visualize both on a world map [11]. They further integrate a similarity measure that infers artist proximity from co-occurrences in playlists.

While placing music items on a map or sphere, arranged in clusters, is certainly the most widely employed method for building visual user interfaces to music collections, there also exists work that employs other visualization approaches. In the *Musicream* interface proposed by Goto and Goto in [8], music of different styles form different streams of discs that drop out of faucets. The listener can simply pick out a piece to listen to or collect similar pieces by dragging a seed song into one of the streams to create a playlist. Pampalk and Goto propose the *MusicSun* interface [30] that fosters exploration of music collections by descriptive terms and item similarities. The descriptors are visualized as sun rays around a set of core seed songs. When the user selects a ray, the system provides a list of recommendations that can be adjusted based on audio and web similarity. The *Songrium* interface [10] by Hamasaki et al. is a collection of web applications designed to enrich the music listening experience. It offers various ways to browse music, e.g., graph-based visualization of songs using audio similarity for placement or cover song exploration in a solar system-like structure.

Research Proposal

Although the presented state-of-the-art approaches offer many interesting possibilities to browse music collections, they are limited in at least two aspects. First, particularly map-based interfaces usually do not explain why certain groups of items are deemed similar by the applied data projection or clustering algorithms. Research on extracting and integrating meaningful music descriptors into the browsing interfaces therefore needs to be conducted. Interestingly, while there exists a large body of work on content-based [3, 34] and context-based [15] feature extraction for music, the resulting descriptors are rarely integrated in existing user interfaces in a way that provides semantic clues to the users. For instance, an intelligent music browsing interface should tell the listener that two music pieces are mapped to nearby positions on a map (or put into the same cluster) because they use the same instrumentation or have a similar rhythmic structure. Respective approaches are likely to break new ground in improving the understanding of musical concepts, especially for non-music experts. User interfaces providing such information might therefore even exhibit some pedagogic value. This could be realized, among others, by gamification of search tasks and multi-faceted information visualization to illustrate various musical properties such as timbre, rhythm, melody, and harmony.

Second, almost no work has been devoted so far to making the visual exploration of music a social event, i.e., to address not only the human–music interaction, but also the human–human interaction during music browsing. This comes as a surprise since music consumption (and performance) has always been a highly social event. In particular in former times music and dance was unequivocally tied together. Nowadays, in the era of social media, users share their music preferences and listening behaviors through social networks, but are offered very few tools to explore music spaces together. Therefore, research should be invested into the many facets of joint exploration of music collections. Such research may include the development of new models for social exploration, incorporating latest findings on motivation mechanisms and immersion, and their integration into prototypical applications. In this context, the evaluation of these models and applications is vital. User-centric evaluation of music retrieval and recommendation algorithms is frequently mentioned as one large shortcoming of (academic) research in the field [33, 21]. Currently, most MIR approaches are evaluated either based on a ground truth that is assumed to be objective or via direct user feedback that is typically very simple in nature, e.g. asking the users whether they like or dislike the recommended music items. We argue that these single-faceted evaluation approaches fall short of investigating and understanding the listeners' needs and requirements for music retrieval systems. Today, crowd-sourced evaluation platforms, e.g., *Amazon Mechanical Turk*, provide easy access to thousands of potential human evaluators and therefore offer good opportunities to perform experiments on a larger scale than in previous times. This is particularly important for academic researchers who do not have access to millions of users like the big players such as *Spotify*, *Apple*, or *Google* which can easily conduct large-scale A/B testing experiments. We are sure that assessing aspects such as serendipity, discoverability, familiarity, or simply pleasure and entertainment, in well-designed user studies will help building better, more user friendly and entertaining interfaces for music discovery and exploration.

ACKNOWLEDGMENTS

This research is supported by the Austrian Science Fund (FWF): P25655.

REFERENCES

1. M. Belkin and P. Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15(6):1373–1396, June 2003.

2. I. Borg and P. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005.
3. M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-Based Music Information Retrieval: Current Directions and Future Challenges. *Proceedings of the IEEE*, 96:668–696, April 2008.
4. O. Celma and M. Nunes. GeoMuzik: A geographic interface for large music collections. In *ISMIR 2008: 9th International Conference on Music Information Retrieval – Late-Breaking/Demo Session*, 2008.
5. T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman & Hall, 1994.
6. M. Dittenbach, D. Merkl, and A. Rauber. Hierarchical clustering of document archives with the growing hierarchical self-organizing map. In *Proceedings of the International Conference on Artificial Neural Networks (ICANN 2001)*, Vienna, Austria, 2001.
7. G. Dror, N. Koenigstein, Y. Koren, and M. Weimer. The Yahoo! Music Dataset and KDD-Cup'11. *JMLR: Proceedings of KDD-Cup 2011 competition*, 18:3–18, October 2012.
8. M. Goto and T. Goto. Musicream: Integrated Music-Listening Interface for Active, Flexible, and Unexpected Encounters with Musical Pieces. *Information Processing Society of Japan*, 50(12):2923–2936, December 2009.
9. S. Govaerts and E. Duval. A Web-based Approach to Determine the Origin of an Artist. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, October 2009.
10. M. Hamasaki, M. Goto, and T. Nakano. Songrium: A Music Browsing Assistance Service with Interactive Visualization and Exploration of a Web of Music. In *WWW'14 Companion*, Seoul, Korea, April 2014.
11. D. Hauger and M. Schedl. Exploring Geospatial Music Listening Patterns in Microblog Data. In *Proceedings of the 10th International Workshop on Adaptive Multimedia Retrieval (AMR 2012)*, Copenhagen, Denmark, October 2012.
12. H. Hotelling. Analysis of a Complex of Statistical Variables Into Principal Components. *Journal of Educational Psychology*, 24:417–441 and 498–520, 1933.
13. S. Huber, M. Schedl, and P. Knees. nepDroid: An Intelligent Mobile Music Player. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR 2012)*, Hong Kong, June 2012.
14. I. T. Jolliffe. *Principal Component Analysis*. Springer, New York, NY, USA, 1986.
15. P. Knees and M. Schedl. A Survey of Music Similarity and Recommendation from Music Context Data. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 10(1), 2013.
16. P. Knees, M. Schedl, T. Pohle, and G. Widmer. An Innovative Three-Dimensional User Interface for Exploring Music Collections Enriched with Meta-Information from the Web. In *Proceedings of the 14th ACM International Conference on Multimedia*, Santa Barbara, CA, USA, October 23–27 2006.
17. T. Kohonen. Self-Organizing Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43:59–69, 1982.
18. T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Germany, 3rd edition, 2001.
19. P. Lamere and D. Eck. Using 3D Visualizations to Explore and Discover Music. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR'07)*, 2007.
20. S. Leitich and M. Topf. Globe of Music - Music Library Visualization using GeoSOM. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR'07)*, Vienna, Austria, 2007.
21. C. C. Liem, M. Müller, D. Eck, G. Tzanetakis, and A. Hanjalic. The Need for Music Information Retrieval with User-centered and Multimodal Strategies. In *Proceedings of the 1st International ACM Workshop on Music Information Retrieval with User-centered and Multimodal Strategies (MIRUM)*, pages 1–6, Scottsdale, AZ, USA, 2011.
22. A. Lillie. MusicBox: Navigating the space of your music. Masters thesis, Massachusetts Institute of Technology, September 2008.
23. D. Lübbers and M. Jarke. Adaptive Multimodal Exploration of Music Collections. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, October 2009.
24. B. McFee, T. Bertin-Mahieux, D. P. Ellis, and G. R. Lanckriet. The Million Song Dataset Challenge. In *Proceedings of the 21st International World Wide Web Conference (WWW): 4th International Workshop on Advances in Music Information Research (AdMIRe)*, Lyon, France, April 2012.
25. J. B. McQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematics, Statistics and Probability*, volume 1, pages 281–297, 1967.
26. F. Mörchen, A. Ultsch, M. Nöcker, and C. Stamm. Databionic Visualization of Music Collections According to Perceptual Distance. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, London, UK, September 11–15 2005.

27. R. Neumayer, M. Dittenbach, and A. Rauber. PlaySOM and PocketSOMPlayer, Alternative Interfaces to Large Music Collections. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, London, UK, September 2005.
28. E. Pampalk. Islands of Music: Analysis, Organization, and Visualization of Music Archives. Master's thesis, Vienna University of Technology, Vienna, Austria, 2001. http://www.oefai.at/∼elias/music/thesis.html.
29. E. Pampalk, S. Dixon, and G. Widmer. Exploring Music Collections by Browsing Different Views. *Computer Music Journal*, 28(3), 2004.
30. E. Pampalk and M. Goto. MusicSun: A New Approach to Artist Recommendation. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, September 2007.
31. S. Roweis and L. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, 2000.
32. M. Schedl. An Explorative, Hierarchical User Interface to Structured Music Repositories. Master's thesis, Vienna University of Technology, Austria, December 2003.
33. M. Schedl, A. Flexer, and J. Urbano. The neglected user in music information retrieval research. *Journal of Intelligent Information Systems*, 41:523–539, December 2013.
34. M. Schedl, E. Gómez, and J. Urbano. Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval*, 8(2–3):127–261, 2014.
35. M. Schedl, C. Höglinger, and P. Knees. Large-Scale Music Exploration in Hierarchically Organized Landscapes Using Prototypicality Information. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR)*, Trento, Italy, April 2011.
36. M. Schedl, T. Pohle, P. Knees, and G. Widmer. Exploring the Music Similarity Space on the Web. *ACM Transactions on Information Systems*, 29(3), July 2011.
37. K. Seyerlehner, G. Widmer, and T. Pohle. Fusing Block-Level Features for Music Similarity Estimation. In *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx)*, Graz, Austria, September 2010.
38. S. Stober and A. Nürnberger. MusicGalaxy - An Adaptive User-Interface for Exploratory Music Retrieval. In *Proceedings of 7th Sound and Music Computing Conference (SMC'10)*, pages 382–389, Barcelona, Spain, July 2010.
39. J. B. Tenenbaum, V. Silva, and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, 2000.
40. A. Ultsch and H. P. Siemon. Kohonen's Self-Organizing Feature Maps for Exploratory Data Analysis. In *Proceedings of the International Neural Network Conference (INNC 1990)*, pages 305–308, Dordrecht, the Netherlands, 1990. Kluwer Academic Publishers.
41. B. Vad, D. Boland, J. Williamson, R. Murray-Smith, and P. B. Steffensen. Design and evaluation of a probabilistic music projection interface. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, Spain, October 2015.
42. L. van der Maaten and G. Hinton. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.
43. S. Vembu and S. Baumann. A Self-Organizing Map Based Knowledge Discovery for Music Recommendation Systems. In *Proceedings of the 2nd International Symposium on Computer Music Modeling and Retrieval (CMMR)*, Esbjerg, Denmark, 2004.
44. M. Wu, Y. and. Takatsuka. Spherical self-organizing map using efficient indexed geodesic data structure. *Neural Networks*, 19(6–7):900—910, July–August 2006.