Approach for exploring a search map through audible landmarks in Virtual Reality

Florian Klein[1], Juliane Höbel-Müller[2], Sayantan Polley[2], Stephan Werner[1], Andreas Nürnberger[2]
[1] *Technische Universität Ilmenau, Electronic Media Technology Group*
[2] *Otto-von-Guericke-Universität Magdeburg, Data and Knowledge Engineering Group*

Introduction

Classical search interfaces based on keyword search and result lists (e.g. Google, Bing or Ecosia) are well suited for users who know what they are looking for. However, unlike keyword-based search, the goal of exploration is not just to present relevant or related items, but to allow the user to learn about the information space and its properties (cf. [1]). List-based interfaces with keyword search are of limited use when the volume and variety of items and metadata seem overwhelming. In contrast, exploratory information retrieval systems should support users with vague or evolving information needs [7]. For example, two-dimensional search maps of items can be useful as they convey information about the general structure and scope of a collection (cf. [10, 5, 6, 7]).

The Search Map

Low et al. propose a two-dimensional search map interface. It gives the impression of panning a global map of images [5], movie items [6] and scientific papers [7] by aligning inexpensive small maps showing local neighborhoods. By selectively jumping from one neighborhood to the next, the user is able to explore the entire item collection. The approach is intended to transfer knowledge about the content and relevance of individual items from one local map to the next. Grouping by similarity would help to assess the relevance of entire clusters once the relevance of a few representative items is known [6]. Furthermore, incidental discoveries would be more likely than standard browsing [6]. The prototype in [6] was modified and extended in [8] to a more interactive and global approach.

Designing interfaces for exploratory search is challenging, particularly in cases where the item feature space is very high-dimensional (cf. [7]). High-dimensional feature spaces make it difficult to display on a computer screen. To visualize similarities, the features are therefore reduced and projected into a two-dimensional display space. Hereby, the neighborhood relationships in the high-dimensional space must be preserved in the projection so that the user can perceive the similarity easily. While different dimension reduction techniques exist such as Self-Organizing Maps (SOMs) [3], Principle Component Analysis (PCA) [2], Multidimensional Scaling (MDS) techniques [4] and Autoencoders [11], none of the aforementioned fully preserve the structure of a collection with more than two dimensions.

Venna and Kaski [12] introduce a measure of trustworthiness and continuity. A search map is fully trustworthy if the closest neighbors of an item in display space are also neighbors in original space. A map is fully continuous if all neighbors in original space are visualized in display space. However, to the best of our knowledge, none of the dimensionality techniques and two-dimensional search maps are able to fully satisfy trustworthiness and continuity. Therefore, we pursue the idea of enriching a search map with an audible guidance technique in a large virtual environment instead of trying to mitigate dimensionality reduction errors.

We introduce the concept of "audible landmarks", which are sound cues that guide the user's attention to relevant items or clusters in the search map. The audible landmarks aim to facilitate the exploration and navigation of the search map and increase the user's confidence and satisfaction with the results. "Audible landmarks" adds the perspective of explainable Artificial Intelligence (AI) in search [18], where explanations support users to gain trust with "AI black box" systems. We present a conceptual system approach for audio-visual guidance of attention in a search map, which consists of three main components: a 360° virtual reality (VR) environment that displays the search map in a spherical layout, a spatial audio representation that assigns sounds to individual items and clusters based on their attributes and positions, and a sonification algorithm that controls the playback and mixing of the sounds according to the user's actions and preferences. We discuss the design considerations and challenges of our approach, as well as some possible applications and future directions. Our approach is motivated by the following research questions.

Research Questions

1 Are users more successful with audiovisual guided exploration than with visual guided exploration?

2 Do users prefer audiovisual or visual guidance?

3 Does a user's exploratory behavior differ between audiovisual and visual guided?

Technical Realization

We reduce the high-dimensional item feature space of a movie collection by learning a variational autoencoder network and use the latent space as new, two-dimensional information space [8]. Items are shown as gray semi-transparent disks on the search map. A selection of items is visually represented by thumbnails. The selection algorithm uses stream sampling that maximizes the minimum distance between thumbnails [9, 8]. One of the main challenges of our approach is to create a plausible acoustic representation of the individual items in the search map in VR. To achieve this, we use UNITY as a development environment. We use a map projection method to transform the flat search map, which is optimized for

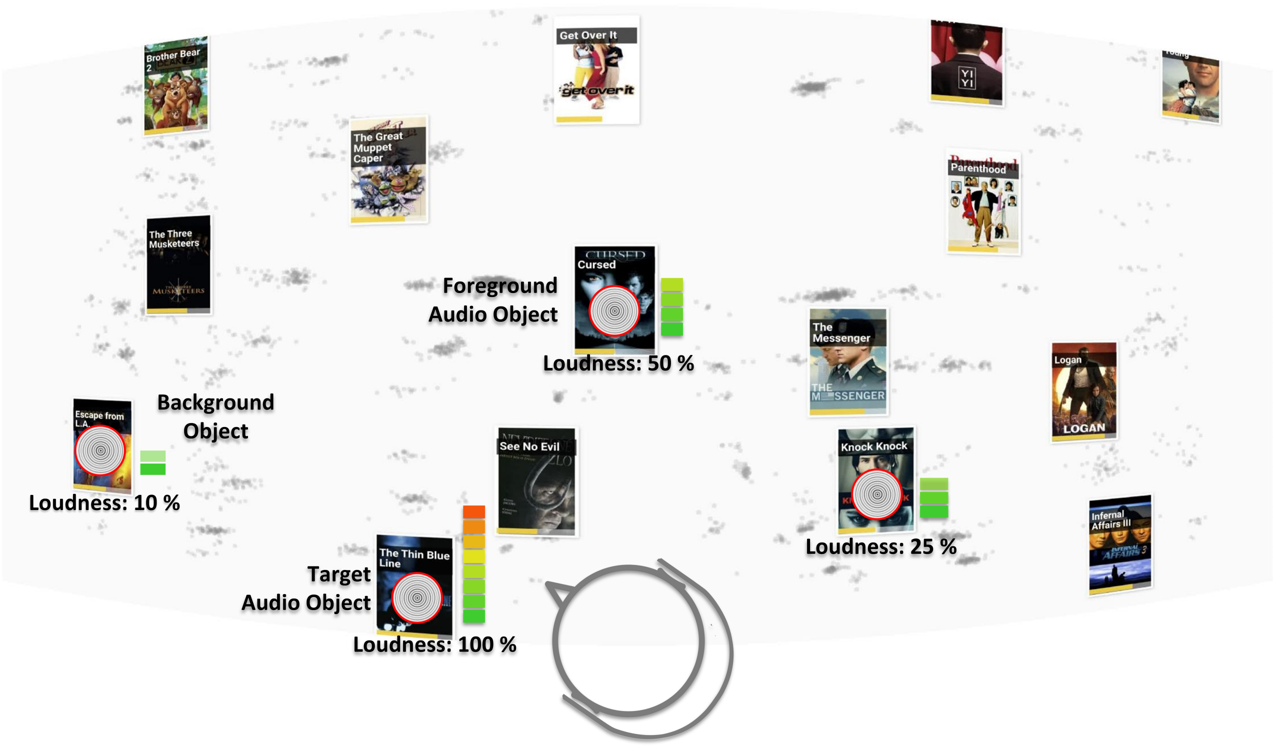

Figure 1: Example search map from a movie database as seen in a virtual reality application. In this example loudness of the items is adjusted according to angular difference between head pose and item direction.

a screen projection, to display it on the inner side of a sphere. For our project, we use an azimuthal equidistant projection, which preserves distances from a chosen center point. This projection also minimizes distortion near the poles. The map itself is limited in its vertical extent to approx. +/-30° viewing range of the user.

In its horizontal extension it is limited to approx. 270° viewing angle. On the one hand, this serves to support a more comfortable head rotation range of the user during the search task. On the other hand, it is intended to prevent the ends of the map, which does not rotate in itself, from coinciding.

To enhance the functionality of our application, we also implement spatial audio objects on the characteristic items (movie pictures). Spatial audio is used to create a sense of direction, distance, and depth for sound sources in space. We used spatial audio objects to attach sounds as landmarks to specific locations and items on the map.

Figure 1 shows an example visualization of a search map as seen by a user wearing a head-mounted display. The user is surrounded by the map and visually and acoustically explores the search items. Different acoustic manipulation can be realized to support the exploration of the map, as described in the following section.

Two different spatial audio auralization approaches are under investigation: A) Google Resonance Audio SDK [15] and B) pyBinSim [16]. Google Resonance Audio SDK is a plugin for UNITY [17] that provides spatial audio features such as binaural rendering, occlusion, and source directivity. pyBinSim is a Python-based tool that allows real-time binaural rendering of dynamic scenes using binaural room impulse responses. Both methods allow us to place sound objects at any distance and direction. Solution A) is better suited for distribution and its more efficient in terms of resources. However it only relies on simulated room acoustics and there are only limited options to control the auralization (such as head-related transfer functions and headphone compensation filters). Solution B) gives us better control of such parameters and enables us to create acoustics of real rooms, however its more a research tool and not suited for software distribution.

To evaluate our application, we implement user interactions and behavior measurements using pointing and tracking. We want to measure user behavior by recording head pose direction, reaction time, and subjective feedback.

Acoustic Manipulations for Exploration Support

To enhance the user experience of exploring virtual search maps spatial acoustic manipulations are applied. The system dynamically adjusts the loudness, directivity, and reverberation of audio objects based on the head pose of the user regarding the individual items in the search map. The system aims to create an intuitive and plausible auditory scene that supports awareness of the items and exploration in the search map. For this different acoustic manipulations are applied:

Loudness manipulation - The volume of each audio object is adjusted according to the angle difference between head orientation and item direction. Item closer to the current head orientation are louder than others. With the availability of eye tracking data, this modification can be coupled with gaze direction, too.

Directivity manipulation - Depending on the direction of the user's head-pose to the audio objects, the object's volume is adjusted frequency dependent. This can be done by mimicing the directivity of natural sound sources in an opposite manner. With increasing angle distance, high frequencies are attenuated first and in the following also middle to lower frequencies. This approach can be seen as a extension of the loudness manipulation.

Reverberation manipulation - To create a spatial auditory impression, reverberation is added for each audio object. Different approaches are under consideration. Similar to the "directivtiy manipulation", objects which have a greater angle distance could be rendered more reverberant. This is to create a better spatial localization of the objects and an environmental context. Another approach is to assign a certain reverberation characteristic to groups of objects. The idea is to put similar items in the same acoustic room and thus provide some classification for the user.

The intended manipulations must be evaluated in a future user study. The efficiency of the search task as success yes/no, the time to complete the task, a visualization of the head pose during the search task (heat map) and a questionnaire to query the preferred method are used as test criteria. We suppose that our approach significantly improves the user experience of exploring search maps through acoustic manipulations. The study itself is not part of this paper and will be conducted separately.

References

[1] Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Information Review 13(5), 407–424.

[2] Jolliffe, I. T. (1986). Principal component analysis. Springer Verlag.

[3] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological Cybernetics, 43(1), 59–69.

[4] Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, 1–27.

[5] Low, T., Hentschel, C., Stober, S., Sack, H., Nürnberger, A. (2014). Visual berrypicking in large image collections. In: Proc. of the 8th Nordic Conf. on HCI. 1043–1046. NordiCHI '14, ACM.

[6] Low, T., Hentschel, C., Stober, S., Sack, H., Nürnberger, A. (2017). Exploring Large Movie Collections: Comparing Visual Berrypicking and Traditional Browsing. In: Amsaleg, L., Gumundsson, G., Gurrin, C., Jónsson, B., Satoh, S. (eds) MultiMedia Modeling. MMM 2017. Lecture Notes in Computer Science, vol. 10133. Springer, Cham.

[7] Low, T., Hentschel, C., Polley, S., Das, A., Sack, H., Nürnberger, A., and Stober, S. (2019). The ISMIR Explorer: A Visual Interface for Exploring 20 Years of ISMIR Publications. In: Proc. of the 20th Int. Soc. for Music Information Retrieval Conf. (ISMIR).

[8] Low, T. (2022). Towards combining search and exploration - escaping the filter bubble through map-based exploration. Dissertation, Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg.

[9] Minack, E., Siberski, W., Nejdl, W. (2011). Incremental Diversification for Very Large Sets: A Streaming-Based Approach. In: Proc. of the 34th Int. Conf. on Research and Development in Information Retrieval. SIGIR. 2011, 585–594.

[10] Rodden, K., Basalaj, W., Sinclair, D., Wood, K. (2001). Does organisation by similarity assist image browsing? In: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems. 190–197. CHI '01, ACM.

[11] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.

[12] Venna, J., Kaski, S. (2006). Local multidimensional scaling. Neural Networks, 19, 889–899.

[13] Venna, J., Peltonen, J., Nybo, K., Aidos, H., Kaski, S. (2010). Information retrieval perspective to nonlinear dimensionality reduction for data visualization. Journal of Machine Learning Research, 11, 451–490.

[14] Werner, S., Klein, F., Neidhardt, A., Sloma, U., Schneiderwind, C., Brandenburg, K. (2021). Creation of Auditory Augmented Reality Using a Position Dynamic Binaural Synthesis System Technical Components, Psychoacoustic Needs, and Perceptual Evaluation. Appl. Sci., 11, 1150.

[15] Google Resonance Audio SDK; Spatial audio for AR/VR; `https://resonance-audio.github.io/resonance-audio` (last seen 31st March 2023).

[16] Neidhardt, A., Klein, F., Knoop, N. and Köllmer, T. (2017). Flexible Python Tool for Dynamic Binaural Synthesis Applications Applications, 142nd Convention of the Audio Engineering Society, Berlin, Germany.

[17] Google Resonance Audio SDK; Runtime and development environment and game engine; `https://www.unity.com` (last seen 31st March 2023).

[18] Polley, S., Koparde, R., Gowri, A., Perera, M. and Nürnberger, A. (2021). Towards trustworthiness in the context of explainable search, In: Proc. of the 44th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval.