MUSIC GALAXY HITCHHIKER: 3D WEB MUSIC NAVIGATION THROUGH AUDIO SPACE TRAINED WITH TAG AND ARTIST LABELS

Dongwoo Suh[1] **Kyungyun Lee**[1] **Jongpil Lee**[2] **Jiyoung Park**[2] **Juhan Nam**[2]
[1] School of Computing, KAIST [2] Graduate School of Culture Technology, KAIST
{kehops,kyungyunlee2393,richter,jypark527,juhannam}@kaist.ac.kr

ABSTRACT

MUSIC GALAXY HITCHHIKER is an interactive web application that presents 3D-navigable visualization of audio feature space. The feature space was computed via two deep convolutional neural networks (DCNN), each supervised by semantic music tags and artist labels. The application recommends music with two different search modes: *SEARCH BY TAG* and *SEARCH BY SONG*. It enhances user experiences with the recommendation results by showing not only the list of songs but also visual similarity in the feature space.

1. INTRODUCTION

Commercial music streaming services exploit various music search and recommendation strategies to provide users with better experiences. The majority of services, however, simply returns a list of songs as a result of the search and recommendation. The under-the-hood analysis that determines the retrieved songs and their relations are usually hidden to users. The goal of web application, MUSIC GALAXY HITCHHIKER, is to intuitively demonstrate the results of content-based song analysis and provide visual interface for music listening. With this application, users can search with tag or song and visually assure the retrieved results by observing their similarity in the 3-D visual space. They also can freely navigate the visual space where similar songs are closely clustered.

A decent number of music browsers that exploit visualization of song quality and their relations have been proposed so far [1, 2, 4]. Our application is similar to *Songrium* [1] in that both use web-based platforms and content-based analysis to determine the geometric positions of songs. However, we focus on leveraging more general song features learned from large-scale supervised learning, and provide various visualization techniques to show song similarity in 3-D space and explain semantic descriptions of songs.

2. METHOD

MUSIC GALAXY HITCHHIKER provide two different search modes and 3-D navigation for music browsing as shown in Figure 1.

2.1 Search By Tag

Search By Tag mode allows users to find songs with tags. The search query can be composed of up to two tags. Currently, we allow only pre-defined 50 tags that represent genre and mood such as *rnb* and *happy*. The similarity is calculated by taking cosine distance between the query vector, a transformation of input query into a 50 dimensional one-hot vector, and the 50-dimensional tag vector of each song that represents the prediction probabilities of the tags inferred by the tag-based DCNN model. This mode also provides two types of real-time visualization techniques, *stacked graph* and *tag bubble*, as shown in Figure 2. Each of them dynamically displays tag prediction values while the song is being played back.

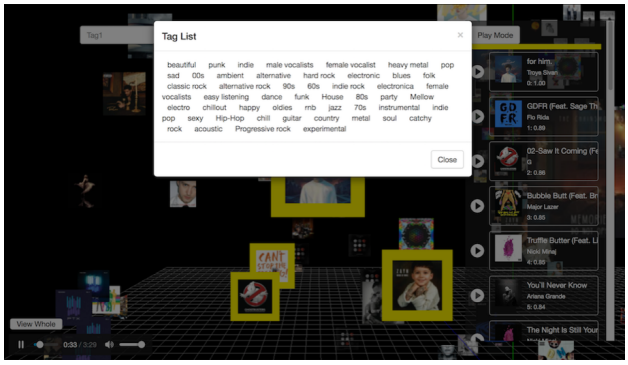
2.2 Search By Song

Search By Song mode takes a song as a query and provides a list of songs similar to the query song. In particular, our system provides two sets of lists based on different content analysis models. One is a tag-based model that utilizes a DCNN and semantic song description labels [3]. The other is an artist-based model that utilizes a DCNN and artist labels [5]. Whenever users click on a song on the 3-D space, new searched lists are generated from the song query. One list from the artist-based model is located on the left side and the other list from the tag-based model is on the right side. Figure 1(b) shows that the selected songs from each model are highlighted with red and yellow boundaries, respectively, and those selected by both models are colored in orange.

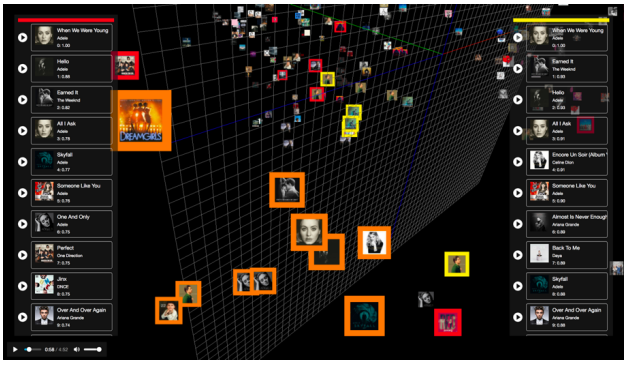
2.3 3D Navigation

Music browsing can be regarded as exploring the universe of music. Our application allows users to actively navigate through the song space and find favorite songs by chance from the similarity-based list of music recommendation. We used three.js [1] to map each song onto a 3-D space. The 3-D position was calculated from the tag prediction

[1] https://threejs.org/

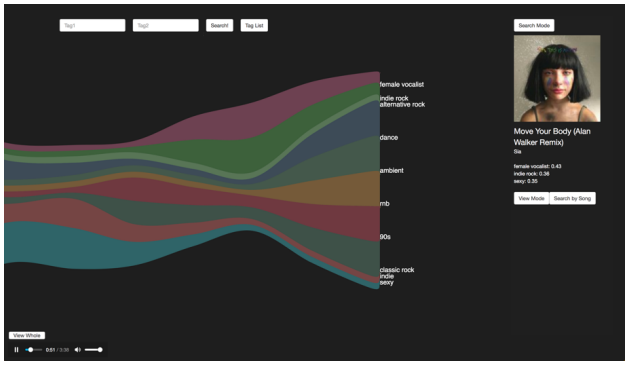


(a) Search by Tag mode that shows available tags

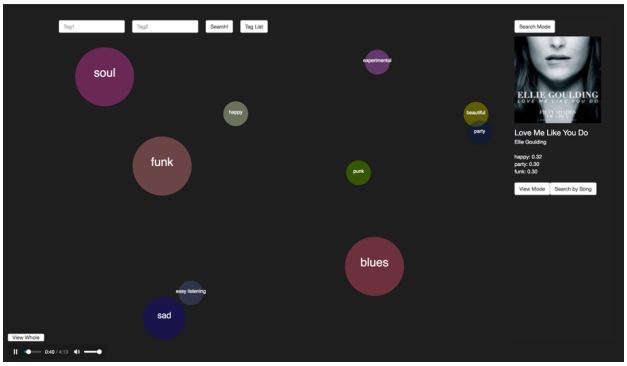


(b) Search by Song mode with highlighted search results

Figure 1: Two different search modes of MUSIC GALAXY HITCHHIKER



(a) Stacked graph



(b) Tag bubble

Figure 2: Real-time tag visualizations

probability vector using t-Distributed Stochastic Neighbor Embedding (t-SNE).

3. DATA AND SYSTEM DESCRIPTION

We chose the audio files for music database from the Melon [2] chart and Billboard's Hot-100 chart [3]. We took up to 5 songs from each artist to ensure diversity, having a total of 897 songs. We also obtained album cover pictures to represent each song on the 3-D space. The system is composed of client and server. The server calculates the similarities of the music with user query. It also sends audio and the metadata of the recommended music to the user side. The client works as a proxy server for any REST API requests to the server instance. The browser renders the web components and stores the UI states.

4. CONCLUSION

We proposed a web-based music application that provides various visualizations of content-based analysis and recommendation. We believe this visual approach is more intuitive and persuasive than simply showing a list of songs. In the future, we plan to combine context-based recommendation techniques to achieve personalization. Also, we will support natural sentences as an input query.

[2] http://www.melon.com/

[3] http://www.billboard.com/charts/hot-100

5. REFERENCES

[1] Masahiro Hamasaki, Masataka Goto, and Tomoyasu Nakano. Songrium: A music browsing assistance service with interactive visualization and exploration of protect a web of music. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, pages 523–528, New York, NY, USA, 2014. ACM.

[2] Paul Lamere and Douglas Eck. Using 3d visualizations to explore and discover music. In *in Int. Conference on Music Information Retrieval*, 2007.

[3] Jongpil Lee, Jiyoung Park, Keunhyoung Luke Kim, and Juhan Nam. Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms. *Sound and Music Computing Conference (SMC)*, pages 220–226, 2017.

[4] Elias Pampalk, Simon Dixon, and Gerhard Widmer. Exploring music collections by browsing different views. *Comput. Music J.*, 28(2):49–62, June 2004.

[5] Jiyoung Park, Jongpil Lee, Jangyeon Park, Jung-Woo Ha, and Juhan Nam. Representation learning of music using artist labels. *arXiv preprint arXiv:1710.06648*, 2017.