Tastalyzer: Audiovisual Exploration of Urban and Rural Variations in Music Taste

Christine Bauer
Markus Schedl
[christine.bauer,markus.schedl]@jku.at
Johannes Kepler University Linz
Linz, Austria

Vera Angerer
Stefan Wegenkittl
[vkarl.mmt-m2017,stefan.wegenkittl]@fh-salzburg.ac.at
Salzburg University of Applied Sciences
Puch/Urstein, Austria

ABSTRACT

We present a browsing interface that allows for an audiovisual exploration of regional music taste around the world. We exploit a total of 10,758,121 geolocated tweets about music. The web-based geo-aware visualization and auralization called *Tastalyzer* enables exploring and analyzing music taste on a fine-grained geographical level, such as (i) comparing rural and corresponding urban music taste within an agglomeration (city) or (ii) comparing the music taste in a target region (agglomeration) to the taste of the country the region is part of and (iii) to the global music taste.

CCS CONCEPTS

• **Information systems → Multimedia information systems**; *Users and interactive retrieval*; • **Human-centered computing → Visualization systems and tools**; *Geographic visualization.*

KEYWORDS

music browsing; music taste; cultural differences; user interface; audiovisual exploration; Twitter

ACM Reference Format:
Christine Bauer, Markus Schedl, Vera Angerer, and Stefan Wegenkittl. 2019. Tastalyzer: Audiovisual Exploration of Urban and Rural Variations in Music Taste. In *Proceedings of the 27th ACM International Conference on Multimedia (MM ’19), October 21–25, 2019, Nice, France.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3343031.3350585

1 INTRODUCTION AND RELATED WORK

Various studies indicate that geography plays a role in music preferences [3, 6]. Most work concentrates on country-specific differences (e.g., [4, 8, 24]). Mizell [18], in contrast, reports regional differences within the United States and particularly compares urban and rural areas. For instance, musicals and operetta music are strongly preferred in the Northeast of the United States, and jazz in the West. Urban residents are far more likely to report a preference for classic rock than people residing in rural regions.

When it comes to geo-aware visualizations of music and music taste, related work is scarce and, typically, existing research remains on a coarse granularity level when visualizing the geospatial aspect. For instance, some visualizations [11, 23] focus on artist location; typically such work exploits and interlinks various information sources to retrieve geospatial information on artists and plot the geographical locations on a map. Park et al. [22] take a different approach to combining music and geospatial information; they start from geospatial positions and sonify the georeferenced data for the selected environment (based on ambient noise, surroundings, traffic, etc.). Then they discuss the musical interpretation of routes on a map. Hauger and Schedl [12] are among the first to visualize geospatial music listening patterns. They present a user interface that allows for exploring artist and genre popularity (in terms of playcounts) for various geographic regions across the globe.

In this work, we propose a new form of exploratory data analysis using our tool *Tastalyzer*—a geo-aware visualization and auralization to explore music taste. *The main scientific goal is to allow for an audiovisual in-depth investigation of differences in music taste around the world.* In contrast to previous research, our work addresses geographic differences in music taste from a *fine-grained geo-spatial perspective* and considers regions *across the globe.* More precisely, we investigate a total of 10,758,121 geolocated tweets about music 3,835 regions around the world; a region represents an *agglomeration* (manifested as a city) which is composed of an *urban area* and a *rural area.* We use the *ne_10m_urban_areas* dataset [19] to identify urban areas—all regions outside urban areas are considered rural—and a Voronoi tessellation [27] for assigning urban and rural areas the closest agglomeration.

The contribution of our work, differentiating Tastalyzer from existing systems, is summarized as follows: *(i)* It allows for comparing music taste in urban and rural areas. *(ii)* We use a modified version of *Hellinger distance* [16] to calculate the difference between urban and rural distributions of music taste. *(iii)* It provides an intuitive visualization, allowing to compare the music taste in a target region to the taste of the country that the region is part of and to the global music taste. *(iv)* For each region, it provides a compact—yet, concise—overview of a region’s music listening preferences as distribution over genres. *(v)* We measure the most representative songs for a region. *(vi)* We use a new dataset of 10,758,121 geolocated tweets about music; other datasets can be easily integrated.

2 MUSIC EXPLORATION IN TASTALYZER

In *Tastalyzer* we visualize the music taste of 4,389 urban and 4,768 rural areas in 5,070 agglomerations around the world.

Interface functionality. Tastalyzer allows for selecting a region of interest on a geospatial map. By clicking on a region, a popup opens containing relevant information for the region (Figure 1). Interactively, a user can select which details are shown in the graph;

ACM ISBN 978-1-4503-6889-6/19/10.
https://doi.org/10.1145/3343031.3350585

for instance, whether the region should be compared to the corresponding country at large or the entire global dataset.

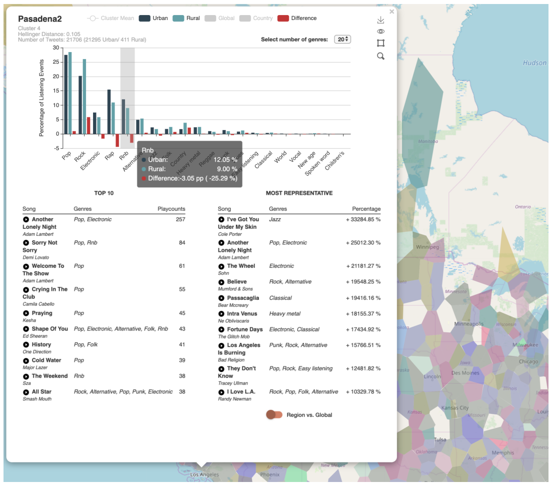

Figure 1: Bar chart showing distribution over genres (urban vs. rural) including popup with details for the genre "RnB".

A bar chart shows the region's music listening preferences as distribution over genres. The switches at the top of the popup allow for comparing the region's genre distribution to the distribution of the corresponding country the region is located in, the urban or rural area, or to the global genre distribution. Each of these data series can be added individually to the chart. Differences between two selected datasets (urban, rural, country, global) can be displayed. When hovering over a genre in the chart, a popup with the exact numbers for the genre and the selected data series will appear. Regions have been colored to code clusters of typical urban/rural differences so that geographic trends can easily be spotted.

Below the bar chart, there are two tables: The table on the left shows the top 10 songs with respect to playcounts in the selected region; the table on the right shows the most representative songs (top 10) for the selected region. Bot tables provide song title, artist, and all genre tags for the song. In addition, the table on the left shows the songs' playcounts in the respective region and the table on the right provides the percentage difference to the song's popularity in the region compared to its popularity in the corresponding country or on a global scope. A toggle switch at the bottom of the list allows for switching between a comparison to either the country or the global scope. Each song presented in the tables can be listened to by simply clicking the "play" button next to it.

Datasets. To investigate geospatial music taste, we used the *Twitter API* to obtain music-related tweets (by filtering w.r.t. hashtags commonly used to indicate music consumption). We considered geo-tagged tweets that were posted in the time period from 01/2011 to 04/2019. Our dataset is, thus, an extension of the *Million Musical Tweets Dataset (MMTD)* [13] from 2013. It consist of 4,389 urban areas with a total of 8,401,890 tweets and 4,768 rural areas with a total of 2,569,868 tweets from 5,070 different agglomerations.

To identify genres for each song in the dataset, we rely on a dictionary of 20 general music genres as used by AllMusic [1]. We gather the top user-generated tags using the Last.fm API [14], casefold tags and genres, and index the tags using the genre dictionary. Finally, each song is described by a bag-of-words representation of genres. Tastalyzer supports an easy embedding of arbitrary genre definitions. We plan to integrate more fine-grained music genre dictionaries; e.g., Freebase [10] or Spotify's microgenres [17].

We use the *ne_10m_urban_areas* dataset [19] by Natural Earth to classify urban areas. This dataset is based on population estimates by LandScan [21]. Tweets from inside those defined areas are considered urban, whereas all regions outside those areas are considered rural. Using a Voronoi tessellation [27], rural tweets are assigned to the closest agglomeration, so that we can compare urban and rural tweets for each agglomeration.

Clustering. To calculate the difference between urban and rural distributions of music taste, we use a modified version of *Hellinger distance* [16]. We use spectral clustering for grouping the differences in a 15-dimensional space.

Representative songs for a target region. We adapt the approach by Bauer and Schedl [5] to identify the songs most representative for a region. We first rescale the global (or country-specific) playcount to match the target region's range, and subsequently relate each song's playcount in the target region to the rescaled global (or country) value; normalized by relative frequencies.

Implementation details. The user interface is implemented using *ECMAScript 6-8* [7]. For the visualization on an interactive map, we use the Python package *Folium*, which allows for using *Leaflet.js* [15] in Python and integrating the world map by OpenStreetMap [20]. The diagrams are created with *ECharts* [2].

For the audiovisual exploration of music, we embed the Spotify Play Button [25]. We use Google's Firebase Cloud Functions [9] to send an authorized request to the Spotify Search Endpoint [26], requesting the selected song via song title and the artist's name. If the request was successful, the Spotify Play Button is embedded using the Spotify ID of the requested track.

3 CONCLUSIONS

In this paper, we presented a web-based geo-aware visualization and auralization called *Tastalyzer*, which facilitates the exploration of regional music taste around the world. In contrast to existing approaches, it enables the comparison of music taste on a fine-grained level, such as comparing the music taste in a target region to the taste of the corresponding country or to the global music taste. As a tool for scientific investigation, Tastalyzer allows for answering research questions such as the following, which are part of our ongoing research: How does music taste differ between urban and rural areas in terms of genre distribution? What are similarities and differences in people's music taste across countries or regions? How does music taste in a country's region differ from the average taste of the whole country or the global taste? Which are the most representative songs listened to in a certain region?

ACKNOWLEDGMENTS

A first version of Tastalyzer was developed by Simon Waldner in his Master's thesis at Salzburg University of Applied Sciences. This research is supported by the Austrian Science Fund (FWF): V579.

REFERENCES

[1] AllMusic 2019. https://www.allmusic.com
[2] Apache ECharts project 2019. JavaScript Library ECharts version 4.1.0.rc2. https://ecomfe.github.io/echarts-doc/public/en/index.html
[3] Young Min Baek. 2015. Relationship Between Cultural Distance and Cross-Cultural Music Video Consumption on YouTube. *Social Science Computer Review* 33, 6 (2015), 730–748. https://doi.org/10.1177/0894439314562184
[4] Christine Bauer and Markus Schedl. 2018. On the Importance of Considering Country-specific Aspects on the Online-Market: An Example of Music Recommendation Considering Country-Specific Mainstream. In *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS '18)*. 3647–3656. http://hdl.handle.net/10125/50349
[5] Christine Bauer and Markus Schedl. 2019. Global and country-specific mainstreaminess measures: Definitions, analysis, and usage for improving personalized music recommendation systems. *PLOS ONE* 14, 6 (2019), e0217389. https://doi.org/10.1371/journal.pone.0217389
[6] Oliver Budzinski and Julia Pannicke. 2017. Do preferences for pop music converge across countries?: Empirical evidence from the Eurovision Song Contest. *Creative Industries Journal* 10, 2 (2017), 168–187. https://doi.org/10.1080/17510694.2017.1332451
[7] Ecma International 2019. ECMAScript 6-8. https://www.ecma-international.org/publications/standards/Ecma-262.htm
[8] Bruce Ferwerda, Andreu Vall, Marko Tkalčič, and Markus Schedl. 2016. Exploring Music Diversity Needs Across Countries. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization (UMAP '16)*. ACM, New York, NY, USA, 287–288. https://doi.org/10.1145/2930238.2930262
[9] Google 2019. Cloud Functions for Firebase. https://firebase.google.com/docs/functions
[10] Google 2019. Freebase API. https://developers.google.com/freebase/
[11] Sten Govaerts and Erik Duval. 2009. A Web-based Approach to Determine the Origin of an Artist. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR '10)*, Keiji Hirata, George Tzanetakis, and Kazuyoshi Yoshii (Eds.). International Society for Music Information Retrieval, 261–266.
[12] David Hauger and Markus Schedl. 2012. Exploring Geospatial Music Listening Patterns in Microblog Data. In *Proceedings of the 10th International Workshop on Adaptive Multimedia Retrieval (AMR '12)*. Copenhagen, Denmark.
[13] David Hauger, Markus Schedl, Andrej Košir, and Marko Tkalčič. 2013. The million musical tweets dataset: What can we learn from microblogs. In *Proceedings of the 14th International Society for Music Information Retrieval Conference* (4–8) *(ISMIR '13)*, Alceu de Souza Britto Jr., Fabien Gouyon, and Simon Dixon (Eds.). International Society for Music Information Retrieval, 189–194. http://www.ppgia.pucpr.br/ismir2013/wp-content/uploads/2013/09/85_Paper.pdf
[14] Last.fm 2019. Last.fm API endpoint track.getTopTags. https://www.last.fm/api/show/track.getTopTags
[15] Leaflet 2019. JavaScript Library Leaflet. https://leafletjs.com
[16] Rémi Lebret and Ronan Collobert. 2015. Rehabilitation of count-based models for word vector representations. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics (CICLing '15)*. Springer, Cham, Germany, 417–429.
[17] Glenn McDonald. 2019. Every Noise at Once. http://everynoise.com
[18] Lee Mizell. 2005. *Music Preferences in the US: 1982-2002.* Technical Report. https://eric.ed.gov/?id=ED511715
[19] Natural Earth 2019. ne_10m_urban_areas_landscan dataset. https://www.naturalearthdata.com/http//www.naturalearthdata.com/download/10m/cultural/ne_10m_urban_areas_landscan.zip
[20] Oak Ridge National Laboratory 2019. OpenStreetMap contributors. https://www.openstreetmap.org
[21] Oak Ridge National Laboratory 2019. ORNL's LandScan. https://landscan.ornl.gov
[22] Sihwa Park, Seunghun Kim, Samuel Lee, and Woon Seung Yeo. 2010. Online Map Interface for Creative and Interactive MusicMaking. In *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME '10)*. Sydney, Australia, 331–334.
[23] Yves Raimond, Christopher Sutton, and Mark Sandler. 2008. Automatic Interlinking of Music Datasets on the Semantic Web. In *Proceedings of the Linked Data on the Web Workshop (LDOW '08)*. CEUR Workshop Proceedings. http://ceur-ws.org/Vol-369/
[24] Marcin Skowron, Florian Lemmerich, Bruce Ferwerda, and Markus Schedl. 2017. Predicting genre preferences from cultural and socio-economic factors for music retrieval. In *Proceedings of the European Conference on Information Retrieval (ECIR '17)*. Springer, Cham, Germany, 561–567.
[25] Spotify 2019. Spotify Play Button. https://developer.spotify.com/documentation/widgets/generate/play-button/
[26] Spotify 2019. Spotify Search. https://developer.spotify.com/console/search/
[27] Georges Voronoi. 1908. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik* 134 (1908), 198–287.