

Intelligent user interfaces for music discovery

Monday, 14 November 2022 21:47



Knees, P., et al. (2020). Intelligent User Interfaces for Music Discovery. *Transactions of the International Society for Music Information Retrieval*, 3(1), pp. 165–179. DOI: <https://doi.org/10.5334/tismir.60>

OVERVIEW ARTICLE

Intelligent User Interfaces for Music Discovery

Peter Knees*, Markus Schedl† and Masataka Goto‡

Assisting the user in finding music is one of the original motivations that led to the establishment of Music Information Retrieval (MIR) as a research field. This encompasses classic Information Retrieval inspired access to music repositories that aims at meeting an information need of an expert user. Beyond this, however, music as a cultural art form is also connected to an entertainment need of potential listeners, requiring more intuitive and engaging means for music discovery. A central aspect in this process is the user interface.

In this article, we reflect on the evolution of MIR-driven intelligent user interfaces for music browsing and discovery over the past two decades. We argue that three major developments have transformed and shaped user interfaces during this period, each connected to a phase of new listening practices. Phase 1 has seen the development of *content-based music retrieval interfaces* built upon audio processing and content description algorithms facilitating the automatic organization of repositories and finding music according to sound qualities. These interfaces are primarily connected to personal music collections or (still) small commercial catalogs. Phase 2 comprises interfaces incorporating *collaborative and automatic semantic description* of music, exploiting knowledge captured in user-generated metadata. These interfaces are connected to collective web platforms. Phase 3 is dominated by *recommender systems* built upon the collection of online music interaction traces on a large scale. These interfaces are connected to streaming services.

We review and contextualize work from all three phases and extrapolate current developments to outline possible scenarios of music recommendation and listening interfaces of the future.

Keywords: music access; music browsing; user interfaces; content-based MIR; community metadata; recommender systems

With its origins in Information Retrieval research, a fundamental goal of Music Information Retrieval (MIR) as a dedicated research field in the year 2000 was to develop technology to assist the user in finding music, information about music, or information in music (Byrd and Fingerhut, 2002). Since then, also driven by the developments in content-based analysis, semantic annotation, and personalization, intelligent music applications have had significant impact on people's interaction with music. These applications comprise "active music-listening interfaces" (Goto and Dannenberg, 2019) which augment the process of music listening to increase engagement of the user and/or give deeper insights into musical aspects also to musically less experienced listeners. For accessing

pieces from music collections, and discovering new music, interfaces based on querying, visual browsing, or recommendation have facilitated new modes of interaction.

Revisiting an early definition of MIR by Downie (2004) as "a multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world's vast store of music accessible to all" 16 years later, reveals that these developments were in fact intended and implanted into MIR from the beginning. Given the music industry landscape and how people listen to music today,¹ this visionary definition has undoubtedly stood the test of time.

In this paper, we reflect on the evolution of MIR-driven user interfaces for music browsing and discovery over the past two decades—from organizing personal music collections to streaming a personalized selection from "the world's vast store of music". Therefore, we connect major developments that have transformed and shaped MIR research in general, and user interfaces in particular,

* Faculty of Informatics, TU Wien, Vienna, AT

[†] Institute of Computational Perception, Johannes Kepler University Linz, AT

[‡] National Institute of Advanced Industrial Science and Technology (AIST), JP

Corresponding author: Peter Knees (peter.knees@tuwien.ac.at)

to prevalent and emerging listening practices at the time. We identify three main phases that have each laid the foundation for the next and review work that focuses on the specific aspects of these phases.

First, we investigate the phase of growing digital personal music collections and interfaces built upon intelligent audio processing and content description algorithms in Section 2. These algorithms facilitate the automatic organization of repositories and finding music in personal collections, as well as commercial repositories, according to sound qualities. Second, in Section 3, we investigate the emergence of collective web platforms and their exploitation for listening interfaces. The extracted user-generated metadata often pertains to semantic descriptions and complements the content-based methods that facilitated the developments of the preceding phase. This phase also constitutes an intermediate step towards exploitation of collective listening data, which is the driving force behind the third, and ongoing phase, which is connected to streaming services (Section 4). Here, the collection of online music interaction traces on a large scale and their exploitation in recommender systems are defining elements. Extrapolating these and other ongoing developments, we outline possible scenarios of music recommendation and listening interfaces of the future in Section 5.

Note that the phases we identify in the evolution of user interfaces for music discovery also reflect the "three ages

with the title list and mere text searches based on bibliographic information are useful enough to browse the whole collection to choose pieces to listen to. However, as the accessible collection grows and becomes largely unfamiliar, such simple interfaces become insufficient (Cunningham et al., 2017; Cunningham, 2019), and new research approaches targeting the retrieval, classification, and organization of music emerge.

"Intelligent" interfaces for music retrieval become a research field of interest with the developments in content-based music retrieval (Casey et al., 2008). A landmark in this regard is the development of query by humming systems (Kageyama et al., 1993) and search engines indexing sound properties of loudness, pitch, and timbre (Wold et al., 1996) that initiate the emancipation of music search systems from traditional text- and metadata-based indexing and query interfaces. While interfaces are still very much targeted at presenting results in sequential order according to relevance to a query, starting in the early 2000s, MIR research proposes several alternatives to facilitate music discovery.

2.1 Map-based music browsing and discovery

Interfaces that allow content-based searches for music retrieval are useful when people can formulate good queries and especially when users are looking for a particular work, but sometimes it is difficult to come up with an appropriate query when faced with a huge

of MIR" as described by Herrera (2018). Herrera refers to these three phases as "the age of feature extractors", "the age of semantic descriptors" and "the age of context-aware systems", respectively. We further agree on the already ongoing "age of creative systems" that builds upon MIR to facilitate new interfaces that support creativity as we discuss in Section 5. We believe that this strong alignment gives further evidence of the pivotal role of intelligent user interfaces in the development of MIR. While user interfaces, especially in the early phases, were often mere research prototypes, their development is tightly intertwined with ongoing trends. Thus, they provide essential gauges to the state of the art, and, even beyond, give perspective of what *could* be possible.

2. Phase 1: Content-Based Music Retrieval Interfaces

The late 1990s see two pivotal developments. On one hand, the Internet gets established as mainstream communication medium and distribution channel. On the other hand, technological advances in encoding and compression of audio signals (most notably mp3) allow for distribution of hi-fi audio content via the Internet and lead to the development of high capacity portable music players (Brown et al., 2001; Bull, 2006). This impacts not only the music industry, but also initiates a profound change in the way people "use" music (North et al., 2004).

At the time, the most popular and conventional interfaces for such music access display the list of bibliographic information (metadata) such as titles and artist names. When the number of musical pieces in a personal music collection is not large, music interfaces

music collection and vague search criteria. Interfaces for music browsing and discovery are therefore proposed to let users encounter unexpected but interesting musical pieces or artists. Visualization of a music collection is one way to provide users with various bird's-eye views and comprehensive interactions. The most popular visualization is to project musical pieces or artists onto a 2D or 3D space ("map") by using music similarity. 2D visualizations also lend themselves to being applied on tabletop interfaces for intuitive access and interaction (e.g. Julià and Jordà, 2009). The trend of spatially arranging collections for exploration can be seen throughout the past 20 years and is still unbroken, cf. **Figure 1**.

One of the earliest interfaces is *GenreSpace* by Tzanetakis et al. (2001) that visualizes musical pieces with genre-specific colors in a 3D space (see **Figure 1(a)** for a greyscale image). Coloring of each piece is determined by automatic genre classification. The layout of pieces is determined by principal component analysis (PCA), which projects high-dimensional audio feature vectors into 3D positions.

Another early interface by Pampalk et al. (2002) called *Islands of Music* visualizes musical pieces on a 2D space representing an artificial landscape, cf. **Figure 1(b)**. It uses a self-organizing map (SOM) to arrange musical pieces so that similar pieces are located near each other, and uses a metaphor of "islands" that represent self-organized clusters of similar pieces. The denser the regions (more items in the same cluster), the higher the landscape (up to "mountains" for very dense regions). Sparse regions are represented by the ocean. Several extensions of the Islands of Music idea were proposed in the following

years. An aligned SOM is used by Pampalk et al. (2004) to enable a shift of focus between clusterings created for different musical aspects. This interface provides three different views corresponding to similarities based on three aspects: (1) timbre analysis, (2) rhythm analysis, and (3) metadata like artist and genre. A user can smoothly change focus from one view to another while exploring how the organization changes. Neumayer et al. (2005) propose a method to automatically generate playlists by drawing a curve on the SOM visualization.

The *nepTune* interface presented by Knee et al. (2006), as shown in **Figure 1(c)**, enables exploration of music collections by navigating through a three-dimensional artificial landscape. Variants include a mobile version (Huber et al., 2012) and a larger-scale version using a growing hierarchical self-organizing map (Dittenbach et al., 2001) that automatically structures the map

For example, *Artist Map* by van Gulik and Vignoli (2005) is an interface that enables users to explore and discover artists. This interface projects artists onto a 2D space and visualizes them as small dots with genre-specific, tempo-specific, or year-specific colors, cf. **Figure 1(d)**. This visualization can also be used to create playlists by drawing paths and specifying regions.

In the *Search Inside the Music* application, Lamere and Eck (2007) use a three-dimensional MDS projection, cf. **Figure 1(e)**. 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.

Other examples use, e.g., metaphors of a "galaxy" or "cosmos," or extend visualizations with additional information. *MusicGalaxy* by Stober and Nürnberger (2010), for example, is an exploration interface that uses a similarity-preserving projection of musical pieces onto

into hierarchically linked individual SOMs (Sneha et al., 2011a). Lübbbers and Jarke (2009) present a browser employing multi-dimensional scaling (MDS) and SOMs to create 3-dimensional 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.

Another SOM-based browsing interface is *Globe of Music* by Leitich and Topf (2007), which maps songs to a sphere instead of a plane by means of a GeoSOM (Wu and Takatsuka, 2006). Mörchen et al. (2005) employ an emergent SOM and the U-map visualization technique (Ultsch and Siemon, 1990) to color-code similarities between neighboring clusters. Vembu and Baumann (2004) incorporate a dictionary of musically related terms to describe similar artists.

While the above interfaces focus on musical pieces, interfaces focusing on artists have also been investigated.

a 2D galaxy space. It takes timbre, rhythm, dynamics, and lyrics into account in computing the similarity and uses an adaptive non-linear multi-focus zoom lens that can simultaneously zoom multiple regions of interest while most interfaces support only a single region zooming, cf. **Figure 1(f)**. The related metaphor of a “planetarium” has been used in *Songrium* by Hamasaki et al. (2015). *Songrium* is a public web service for interactive visualization and exploration of *web-native music* on video sharing services.² It uses similarity-preserving projections of pieces onto both 2D and 3D galaxy spaces and provides various functions: analysis and visualization of derivative works, and interactive chronological visualization and playback of musical pieces, cf. **Figure 1(g)**.

Vad et al. (2015) apply t-SNE (van der Maaten and Hinton, 2008) to mood- and emotion-related descriptors, which they infer from low-level acoustic features. The result of the data projection is visualized on a 2D map, around which the authors build an interface to support the creation of playlists by drawing a path and by area selection, as can be seen in **Figure 1(h)**.

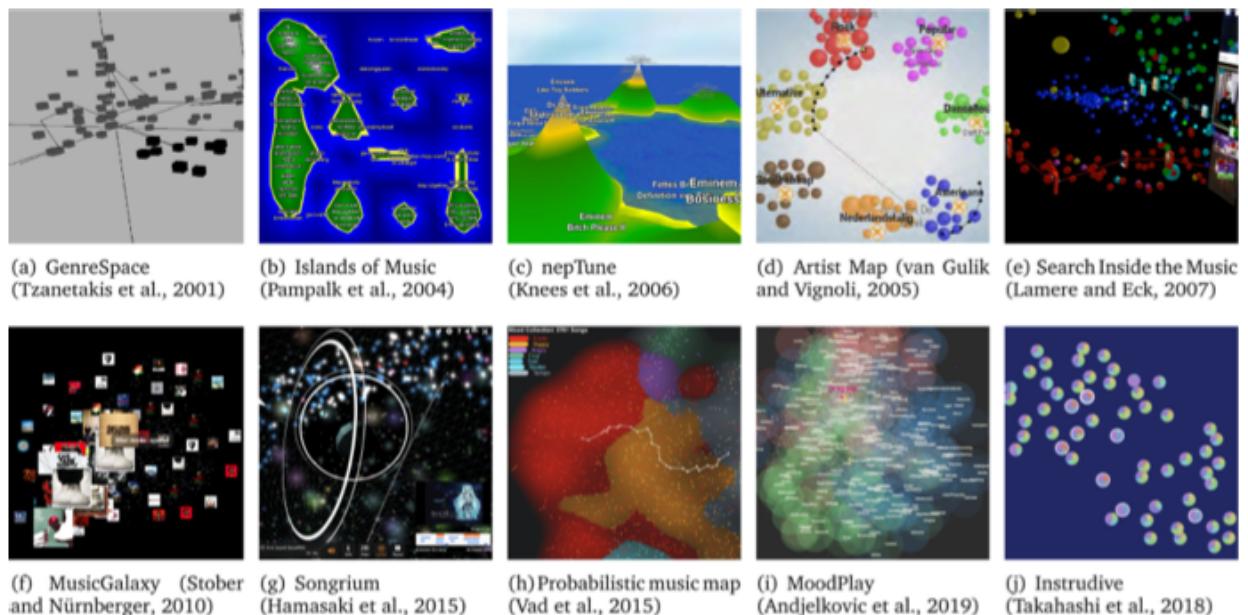


Figure 1: Examples of map-based music browsing interfaces based upon dimensionality reduction techniques.

MoodPlay by Andjelkovic et al. (2019) uses correspondence analysis on categorical mood metadata to visualize artists in a latent mood space, cf. **Figure 1(i)**. More details on the interactive recommendation approach facilitated through this visualization can be found in Section 4.1.

Instrudive by Takahashi et al. (2018) enables users

constructing “mixtapes” from given start and end tracks (Flexer et al., 2008). *VocalFinder* by Fujihara et al. (2010) enables content-based retrieval of songs with vocals that have similar vocal timbre to the query song.

Visualization of a music collection is not always necessary to develop music interfaces. Stewart et al. (2008) present an interface that uses only sound auralization and

to browse and listen to musical pieces by focusing on instrumentation detected automatically. It visualizes each musical piece as a multicolored pie chart in which different colors denote different instruments, cf. **Figure 1(j)**. The ratios of the colors indicate relative duration in which the corresponding instruments appear in the piece.

2.2 Content-based filtering and sequential play

When a collection of music becomes huge, it is not feasible to visualize all pieces in the collection. Other types of interfaces that visualize a part of the music collection instead of the whole have also been proposed. An example is *Musicream* by Goto and Goto (2009), a user interface that focuses on inducing active user interactions to discover and manage music in a huge collection. The idea behind Musicream is to see if people can break free from stereotyped thinking that music playback interfaces must be based on lists of song titles and artist names. To satisfy the desire "I want to hear something," it allows a user to unexpectedly come across various pieces similar to ones that the user likes. As shown in **Figure 2(a)**, disk icons representing pieces flow one after another from top to bottom, and a user can select a disk and listen to it. By dragging a favorite disk in the flow, the user can easily pick out other pieces similar to the query disk (attach similar disks) by using content-based similarity. In addition, to satisfy a desire like "I want to hear something my way," Musicream gives a user greater freedom of editing playlists by generating a playlist of playlists. Since all operations are automatically recorded, the user can also visit and retrieve a past state as if using a time machine.

The *FM4 Soundpark Player* by Gasser and Flexer (2009) makes content-based suggestions by showing up to five similar tracks in a graph-like manner, cf. **Figure 2(b)**, and

haptic feedback to explore a large music collection in a two or three-dimensional space.

The article "*Reinventing the Wheel*" by Pohle et al. (2007) reveals that a single-dial browsing device can be a useful interface for musical pieces stored on mobile music players. The whole collection is ordered in a circular locally-consistent playlist by using the Traveling Salesman algorithm so that similar pieces can be arranged adjacently. The user may simply turn the wheel to access different pieces. This interface also has the advantage of combining two different similarity measures, one based on timbre analysis and the other based on community metadata analysis. **Figure 2(c)** shows an extended implementation of this concept by Schnitzer et al. (2007) on an Apple iPod, the most popular mobile listening device at the time.

2.3 Summary of Phase 1

Phase 1 is strongly connected to browsing interfaces that make use of features extracted from the signal and present repositories in a structured manner to make them accessible. As many of these developments are rooted in the early years of MIR research, they often reflect the technological state of the art in terms of content descriptors, with the discovery interface attached as a communication vehicle to present the capabilities of the underlying algorithms. Thus, a user experience (UX) beyond the possibility of experiencing a novel, alternative view on collections or being assisted in the task of creating playlists is not the focus of these discovery interfaces. Consequently, user-centric evaluations of the interfaces are scarce and often only anecdotal.

Later works put more emphasis on evaluation of the proposed interfaces. Findings include that while users initially expect to find genre-like structures on maps, other organisation criteria like mood are perceived positively

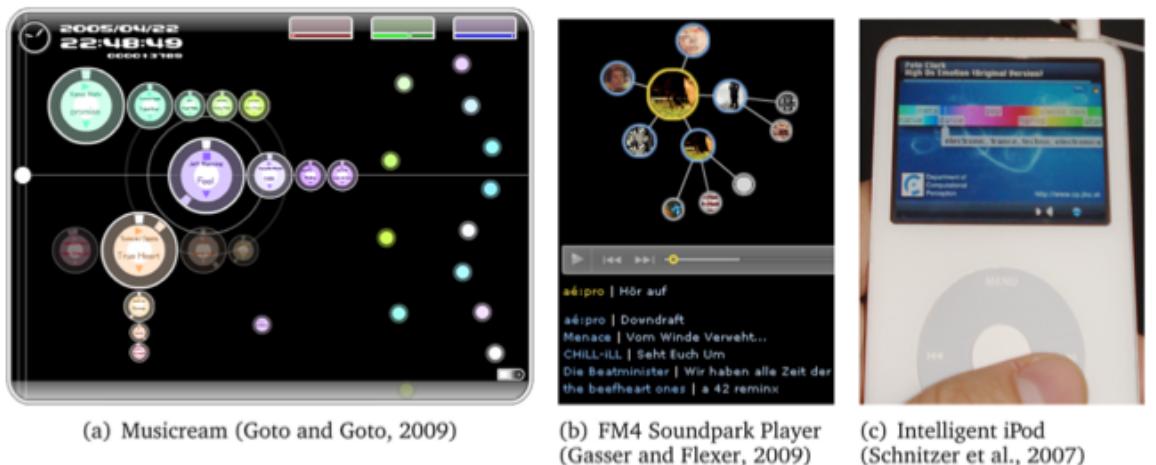


Figure 2: Interfaces for sequential exploration of collections based on content similarity.

for exploration, rediscovery, and playlist generation, once they become familiar (Vad et al., 2015; Andjelkovic et al., 2019).

3. Phase 2: Collaborative and Automatic Semantic Description

While content-based analysis allowed for unprecedented views on music collections based on sound, interfaces built solely upon the extracted information were not able to "explain" the music contained or give semantically meaningful support for orientation within the collections. That is, while they are able to capture qualities of the sound of the contained music, they largely neglect existing concepts of music organization, such as (sub-)genres, and how people use music, e.g., according to mood or activity (Lonsdale and North, 2011; Ferwerda et al., 2015). This and other cultural information is however typically found on the web and ranges from user-generated tags to unstructured bits of expressed opinions (e.g., forum posts or comments in social media) to more detailed reviews and encyclopedic articles (containing, e.g., biographies and discography release histories). In MIR, this type of data is often referred to as community metadata or music context data (Knees and Schedl, 2013).

These online "collaborative efforts" of describing music are resulting in a rich vocabulary of semantic labels ("folksonomy") and have shaped music retrieval interfaces towards music information systems starting around 2005. A very influential service at this time, both as a music information system and a source for semantic social tags, is Last.fm.³ In parallel, platforms like Audioscrobbler, which merged with Last.fm in 2005, take advantage of users being increasingly always connected to the Internet and tracking listening events for the sake of identifying listening patterns and making recommendations, leading to the phase of automatic playlisting and music recommendation (cf. Section 4). In this section, we focus on semantic labels, such as social tags (Lamere, 2008), describing musical attributes as well as metadata and descriptors of musical reception, as a main driver of MIR research and music interfaces.

3.1 Collaborative platforms and music information systems

With music related information being ubiquitous on the web, dedicated web platforms that provide background knowledge on artists emerge, e.g. the AllMusic Guide,⁴ depending on editorial content. Using new technologies, such music information systems can, however, also be built by aggregating information extracted from various sources, such as knowledge bases (Raimond et al., 2007; Raimond, 2008) or web pages (Schedl et al., 2011b), or by taking advantage of the "wisdom of the crowd" (Surowiecki, 2004) and building collaborative platforms like the above mentioned Last.fm.

A central feature of Last.fm is to allow users to tag their music, ideally resulting in a democratic ground truth (Mai, 2011) of what could be considered the semantic

information are noisy and non-trustworthy information as well as data sparsity and cold start issues mostly due to popularity biases (cf. Lamere, 2008).

MIR research during this phase therefore deals extensively with auto-tagging, i.e., automatically inferring semantic labels from the audio signal of a music piece (or related data), to overcome this shortcoming (e.g. Eck et al., 2008; Bertin-Mahieux et al., 2008; Whitman and Ellis, 2004; Turnbull et al., 2007a; Kim et al., 2009; Sordo, 2012; Mandel et al., 2011).

Alternative approaches to generate semantic labels involve human contributions. TagATune by Law et al. (2007) is a game that pairs players across the Internet who try to determine whether they are listening to the same song by typing tags, cf. Figure 3(b). In return for entertaining users, TagATune has collected interesting tags for a database of songs. Other examples of interfaces that were designed to collect useful information while engaging with music are MajorMiner by Mandel and Ellis (2008) (see Figure 3(c)), Listen Game by Turnbull et al. (2007b), Herdit by Barrington et al. (2009), and Moodswings by Kim et al. (2008) (cf. Section 4.1).

A more traditional way to obtain musically informed labels is to have human experts, e.g. trained musicians, manually label music tracks according to predefined musical categories. This approach is followed by the Music Genome Project,⁵ and serves as the foundation of Pandora's automatic radio stations (cf. Section 4). In the Music Genome Project, according to Prockup et al. (2015), "the musical attributes refer to specific musical components comprising elements of the vocals, instrumentation, sonority, and rhythm."

As a consequence of these efforts, during this phase, the question of how to present and integrate this information

Music » Led Zeppelin » Tracks » Stairway to Heaven » Tags

Tags

60s 70s 80s acoustic alternative amazing awesome awesome guitar jams ballad ballads beautiful best song ever best songs ever blues blues rock british chill chillout classic **classic rock** classics cool epic favorite favorite songs favorites favourite favourite songs favourites folk folk rock genius great guitar guitar solo guitar virtuoso **hard rock** heavy metal jimmy page **led zeppelin** legend legendary love masterpiece melancholic mellow metal oldies progressive progressive rock psychedelic psychedelic rock **rock** rock and roll rock ballad rock ballads sad slow soft rock stairway to heaven

(a) Last.fm tags of Led Zeppelin
(source: <https://musicmachinery.com/tag/lastfm/>)

TagATune

Describe the tune ... Listening to the same tune? **Correct** 60 points Partner **Correct** 60 points

your description: male vocal medieval music quartet two females your partner's description: guitar solo no vocals

Your partner has chosen:

MajorMiner

Describe this clip

Your tags: **classical, harp, female, sad, love, fiddle, violin**

New clip Game summary

Tag colors: 2 points, 1 point, no points yet that could be 0, 0 points

Inter-ECG Cancer Thesis 4 pages 100% © 2007 MajorMiner, Inc.

(b) TagATune (Law et al., 2007) (c) MajorMiner (Mandel and Ellis, 2008)

Figure 3: Exemplary sources for human-generated semantic

(Vural, 2011) or what could be considered the semantic dimensions of the corresponding tracks, cf. **Figure 3(a)**. However, typical problems arising with this type of

Figure 3: Exemplary sources for human-generated semantic annotations. (a) collaborative tags; (b) and (c) games with a purpose or crowdsourcing.

into interfaces was secondary to the question of how to obtain it, as will become obvious next.

3.2 Visual interfaces

With the trend towards web-based interfaces, visualizations and map-based interfaces integrating semantic information have been proposed. This semantic information comprises tags typically referring to genres and musical dimensions such as instrumentation, as well as geographical data and topics reflecting the lyrical content.

MusicRainbow by Pampalk and Goto (2006) is a user interface for discovering unknown artists, which follows the above idea of a single-dial browsing device but features informative visualization. As shown in **Figure 4**, artists are mapped on a circular rainbow where colors represent different styles of music. Similar artists are automatically mapped near each other by using the traveling salesman algorithm and summarized with word labels extracted from artist-related web pages. A user can rotate the rainbow by turning a knob and find an interesting artist by referring to the word labels. The **nepTune** interface shown in **Figure 1(c)** also provides a mode that integrates text-based information extracted from artist web pages for supporting navigation in the 3D environment. To this end, labels referring to genres, instruments, origins, and eras serve as landmarks.

Other approaches explore music context data to visualize music over real geographical maps, rather than computing a clustering based on audio descriptors. For instance, Govaerts and Duval (2009) extract geographical information from biographies and integrate it into a visualization of radio station playlists, cf. **Figure 5**. Hauger and Schedl (2012) extract listening events and location information from microblogs and visualize both on a world map.

Lyrics are also important elements of music. By using semantic topics automatically estimated from lyrics, new types of visual interfaces for lyrics retrieval can be achieved. **LyricsRadar** by Sasaki et al. (2014) is a lyrics retrieval interface that uses latent Dirichlet allocation (LDA) to analyze topics of lyrics and visualizes the topic ratio for each song by using the topic radar chart. It then

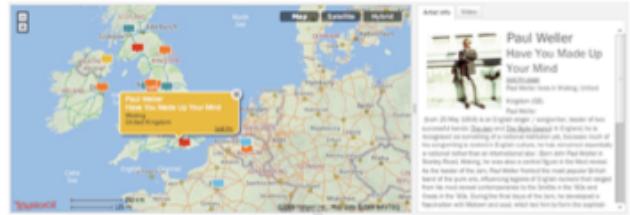


Figure 5: Automatically enriched information system by Govaerts and Duval (2009).

enables a user to find her favorite lyrics interactively. **Lyric Jumper** by Tsukuda et al. (2017) is a lyrics-based music exploratory web service that enables a user to choose an artist based on topics of lyrics and find unfamiliar artists who have a similar profile to her favorite artist. It uses an advanced topic model that incorporates an artist's profile of lyrics topics and provides various functions such as topic tendency visualization, artist ranking, artist recommendation, and lyric phrase recommendation.

3.3 Summary of Phase 2

The second phase of music discovery interfaces gives emphasis to textual representations in interfaces to convey semantic features of the music tracks to the user. This gives the user deeper insights into the individual tracks and allows for exploration through specific facets, rather than structuring repositories and identifying neighboring tracks based on a similarity function integrating various aspects. On the user's side, these interfaces require a more active exploration and selection of relevant properties when browsing.

With the integration of semantic information from structured, semi-structured, and unstructured sources, traditional retrieval paradigms become again more relevant in the music discovery process (cf. Bischoff et al., 2008). At the same time, the extracted music information as well as the data collected during interaction with collaborative platforms can be exploited to facilitate passive discovery, leading to Phase 3.

4. Phase 3: Recommender Interfaces and Continuous Streaming

With ubiquitous Internet connection and the development of computer and entertainment systems to be always online, physical music collections have lost relevance to many people, as virtually all music content is available at all times.⁶

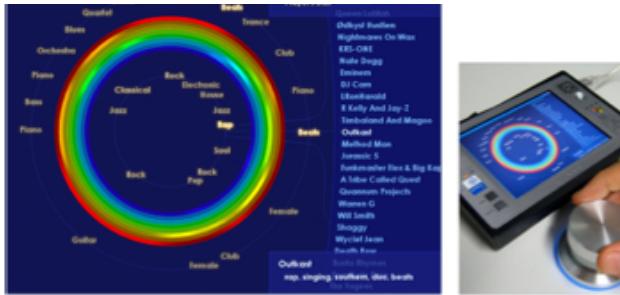


Figure 4: MusicRainbow: An artist discovery interface that enables a user to actively browse a music collection by using audio-based similarity and web-based labeling.

In essence, subscription streaming services like Spotify, Pandora, Deezer, Amazon Music and Apple Music have transformed the music business and music listening alike.

A central element to these services is the aspect of personalization, i.e., providing foremost a user-tailored view onto the available collections of allegedly tens of millions of songs. Discovery of music is therefore also performed by the system, based on the user profile of past interactions, rather than just by the user herself.

Music recommendation typically models personal preferences of users by using their listening histories or explicit user feedback (e.g. Slaney and White, 2007; Celma, 2010). It then generates a set of recommended musical

pieces or artists for each user. This recommendation can be implemented by using collaborative filtering based on users' past behaviors, and exhibits patterns of music similarity not captured by content-based approaches (Slaney, 2011). When the playback order of recommended pieces is important, automatic playlist generation is also used (e.g. McFee and Lanckriet, 2011; Hariri et al., 2012; Bonnin and Jannach, 2014).

The main challenges of this type of algorithm are, as in all other domains of recommender systems, cold start problems. The approach taken to remedy these is again to integrate additional information on the music items to be recommended, i.e. facets of content and metadata as applied in the earlier phases, by building hybrid recommenders on top of pure collaborative filtering. Additionally, context-awareness plays an important role, for instance to recommend music for daily activities (Wang et al., 2012).

This still ongoing phase starts around 2007 and sees further boosts around 2010 and 2015, with an unbroken upward trend. An overview of aspects, techniques and challenges of music recommender systems is described by Schedl et al. (2015). Therefore, in this section, we do not elaborate on the basics of music recommender systems. Instead, we highlight again interfaces that focus on personalization and user-centric aspects (Section 4.1) and the recent trend to introduce psychologically-inspired user models in recommender algorithms (Section 4.2), as we consider these to be the bridge to future intelligent music listening interfaces.

4.1 Recommender interfaces

Although most related studies have focused on methods and algorithms of music recommendation and playlist generation, or user experiences of recommender systems, some studies focus on interfaces.

More recently, studies have focused on the design of user-centric recommender interfaces to account for individual preferences and control of the recommendation process. Jin et al. (2018) investigate the impact of different control elements for users to adapt recommendations, while aiming at preventing cognitive overload. One finding is that users with high musical sophistication index (Müllensiefen et al., 2014) not only appreciate higher control over recommendations but also perceive adapted recommendations to be of higher quality, leading to higher acceptance. The impact of personal characteristics on preferences of visual control elements is further investigated by Millecamp et al. (2018). Again, participants with high musical sophistication index, as well as Spotify power users, showed strong preference for control via a radar chart over traditional sliders for adapting recommendation parameters for discovery of music, cf. **Figure 6**. Kamekhosh et al. (2020) investigate the implications of recommender techniques on the discovery of music in playlist building. They find that recommendations displayed in visual playlist building tools are actively incorporated by users and even impact the choices made in playlist creation when recommendations are not directly incorporated.

Overall, these interfaces and studies about interfaces show a clear trend towards personalization and user-centric development, integrating aspects of personality and affect (cf. Knees et al., 2019). This observation is further supported by works dealing with psychologically-inspired music recommendation as described next.

4.2 Psychologically-inspired music recommendation

Recently, music recommender research is experiencing a boost in topics related to psychology-informed recommendation. In particular the psychological concepts of personality and affect (mood and emotion) are

MUSICsun by Rampaik and Goto (2007) is a user interface for artist recommendation. A user first puts favorite artist names into a “sun” metaphor, a circle in the center of the screen, and then obtains a ranked list of recommended artists. The sun is visualized with some surrounding “rays” that are labeled with words to summarize the query artists in the sun. By interactively selecting a ray, the user can look at and listen to the corresponding recommended artists.

MoodPlay by Andjelkovic et al. (2019) is an interactive music recommender system that uses a hybrid recommendation algorithm based on mood metadata and audio content, cf. Section 2.1. A user first constructs a profile by entering favorite artist names and then obtains a ranked list of recommended artists, highlighted in a latent mood space visualization, cf. **Figure 1(i)**. The centroid of profile artist positions is used to recommend nearby artists. The change of a user's preference is interactively modeled by moving in this space and its trail is used to recommend artists.

In *MoodSwings* (Kim et al., 2008), users try to match each other while tracing the trajectory of music through a 2D emotion space. The users' input provides metadata on the emotional impression of songs as it changes over time.

increasingly integrated into prototypes. The motivation for this is that while listening to music both personality traits and affective states have been shown to influence music preferences strongly (Rentfrow and Gosling, 2003; Ferwerda et al., 2017; Schedl et al., 2018).

Lu and Tintarev (2018) propose a system that re-ranks results of a collaborative filtering approach according



Figure 6: Personalized control over recommendations (Millecamp et al., 2018).

to the degree of diversity each song contributes to the recommendation list. Since previous studies showed that personality is most strongly correlated with musical key, genre, and number of artists, the authors implement diversity through these features and adjust results depending on the listener's personality. Fernández-Tobías et al. (2016) propose a personality-aware matrix factorization approach that integrates a latent user factor describing users' personality in terms of the Big Five/OCEAN model with the 5 factors openness, conscientiousness, extraversion, agreeableness, and neuroticism (John et al., 1991). Deng et al. (2015) propose an emotion-aware recommender for which they extract music listening information and emotions from posts in Sina Weibo,⁷ a popular Chinese microblogging service, adopting a lexicon-based approach (Chinese dictionaries and emoticons). *FocusMusicRecommender* by Yakura et al. (2018) recommends and plays back musical pieces suitable to the user's current concentration level estimated from the user's behavior history.

4.3 Summary of Phase 3

The still ongoing third phase of music discovery interfaces is driven by machine learning methods to predict the "right music" at the "right time" for each user. To this end, user profiles consisting of previous interactions, as well as

from recognizing non-standard and ambiguously pronounceable terms like artist names from spoken language to context and intention-aware disambiguation of utterances, e.g. to identify the intended version of a song.

In terms of recommendation approaches this signifies a renaissance of knowledge-based recommender systems (Burke, 2000) and increasing integration of music knowledge graphs (Oramas et al., 2016), enabling conversational interaction and techniques like “critiquing”, an iterative process of evaluation and modification of recommendations based on the characteristics of items (Chen and Pu, 2012), and a need for story generation techniques (Behrooz et al., 2019). An example showcasing some of these techniques is the music recommender chatbot MusicBot by Jin et al. (2019). MusicBot features user-initiated and system-suggested critiquing which have positive impact on user engagement as well as on diversity in discovery. MusicRoBot by Zhou et al. (2018) is another conversational music recommender built upon a music knowledge graph.

As a result, the predominant notion of a music discovery interface being a graphical user interface might lose relevance as interaction moves to a different modality. In this setting, the trends towards *context-awareness* and *personalization*, also on the level of individual personality

potentially any other source of information on the user, such as context data or personality features, are exploited

Current commercial platforms and their interfaces are designed to cover a variety of use cases, by providing applications with different foci. As different usage scenarios and user intents require different types of recommendation strategies, the user is given the choice as to which focus is best suited in the current situation, by offering different applications to select from. For instance, discovery of new tracks (e.g. as in Spotify's Release Radar) requires a different strategy than rediscovery of known tracks (e.g. as in Daily Mixes) and a personalized radio station for *Workout* will have different selection criteria than a radio station for *Chill*. In addition, platforms integrate many functions of traditional terrestrial radio stations as well, including promotion of artists, and therefore also provide manually curated discovery, e.g. by means of non-personalized radio stations or playlists. Hence, music discovery interfaces have moved away from a one-size-fits-all approach to a suite of applications catering to different listening needs and access paradigms.

5. The Next Phase: The Future of Intelligent Music User Interfaces

Just as technological developments have enabled and shaped the nature of music access in the past – from audio compression to always-online mobile devices – the future will be no different in this regard.

One direction that has already been taken is the streaming of music via so-called *smart speakers* like Amazon Echo, Google Home, or Apple HomePod, controlled via voice through *personal assistants* like Alexa, Google Assistant, or Siri, respectively (Dredge, 2018). For music recommendation, this poses new challenges

traits, gain even more importance. This amplifies the already central challenge to accurately infer a user's intent in an action (listening, skipping, etc.), i.e., to uncover the reasons why humans indulge in music, from the comparatively limited signal that is received (Hu et al., 2008; Jannach et al., 2018).

On the other hand, we see the developments in the realm of *music generation and variation* algorithms. These algorithms create new musical content by learning from large repositories of examples, cf. recent work by Google Magenta⁸ (Roberts et al., 2018a, b; Huang et al., 2019) and OpenAI,⁹ and/or with the help of informed rules and templates, e.g., in automatic video soundtrack creation or adaptive video game music generation. An important development in this research direction is again to give the user agency in the process of creation ("co-creation"). For instance, a personalization approach to melody generation is taken in *MidiMe* by Dinculescu et al. (2019), cf. **Figure 7(c)**. *Cococo* by Louie et al. (2020) is a controlled music creation tool for completion of compositions, giving high-level control of the generative process to the user.

In the long run, we expect the borders of these domains to blur, i.e., there will be no difference in accessing existing, recorded music and music automatically created by the system tailored to the listener's needs. More concretely, as discussed as one of grand challenges in MIR by Goto (2012), we envision music streaming systems that deliver preferred content based on the user's current state and situational context, automatically change existing music content to fit the context of the user, e.g., by varying instruments, arrangements, or tempo of the track, and even create new music based on the given setting. One of the earliest approaches to customize or personalize existing music is "*music touch-up*" by Goto (2007).

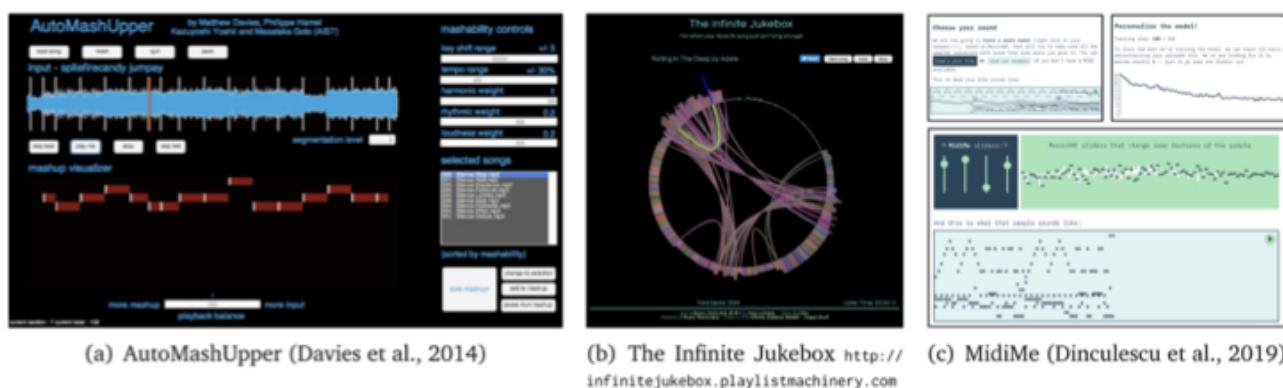


Figure 7: Interfaces highlighting the confluence of music listening and music (co-)creation.

Further examples are *Drumix* by Yoshii et al. (2007) and *AutoMashUpper* by Davies et al. (2014), cf. **Figure 7(a)**. Lamere's *Infinite Jukebox*¹⁰ can also be seen as an example in this direction, cf. **Figure 7(b)**.

With the current knowledge of streaming platforms about a user's preferences, context sensing devices running the music apps, and first algorithms to variate and generate content, the necessary ingredients for such a development seem to be available already. These developments, along with the increasing interest in the role of Artificial Intelligence (AI) in arts in general, will have a larger impact than just a technological one, raising questions of legal matters regarding ownership and intellectual property (Sturm et al., 2019) or the perception and value of art, especially AI-created art (Hong and Curran, 2019). Research in these areas therefore needs to consider a variety of stakeholders.

6. Conclusions and Discussion

We identified three phases of listening culture and discussed corresponding intelligent interfaces. Interfaces pertaining to the first phase focus on structuring and visualizing smaller scale music collections, such as personal collections or early digital sales repositories. In terms of research prototypes, this phase is most driven by content-based MIR algorithms. The second phase deals with web-based interfaces and information systems, with a strong focus on textual descriptions in the form of collaborative tags. MIR research during this phase therefore deals with automatic tagging of music and utilization of tag information in interfaces. Finally, the third and current phase is shaped by lean-back experiences driven by automatic playlist algorithms and personalized recommendation systems. MIR research is therefore shifting towards exploitation of user interaction data, however always with a focus on integration of content-based methods, community metadata, user information, and contextual information of the user. While the former three strategies are typically applied to remedy cold start problems, the integration of context-awareness often amplifies them.

The overview given in this paper focuses on academic interfaces over the past 20 years; however, it is interesting to observe that today's most successful commercial platforms bear little resemblance to the prototypes

discussed. Instead, traditional list or "spreadsheet" views showing the classic metadata fields title, artist, album, and track length still seem to constitute the state of the art in displaying music throughout most applications. This discrepancy between academic work and commercial services affects mostly the interfaces, as the underlying methods for content feature extraction, metadata integration, and recommendation can all be found in similar forms in existing systems. This raises the question whether academic interfaces do not meet users' desiderata for a music application or if commercial interfaces are missing out on beneficial components.

Lehtiniemi and Holm (2013) have investigated different types of music discovery interfaces and summarized user comments regarding desired features for an "ultimate" music player: "a streaming music service with a large music collection and a mobile client; support for all three modes of music discovery (explorative, active and passive); easy means for finding new music (e.g. textual search, 'get similar' type of functionality and mood-based music search); music recommendations with surprising and unexpected results; links to artist videos, biography and other related information; storing, editing and fine-tuning playlists; adapting to user's own musical taste; support for social networking services; contextual awareness; and customizable and aesthetic look."

We can see that commercial interfaces tick many boxes from this list, but we can also see how the discussed interfaces from all three phases relate to these aspects and have left their footprints in current systems. While map-based interfaces from Phase 1 see no adoption in current commercial systems, the concepts of similarity-based retrieval, playlist generation, and sequential play are still key elements. From Phase 2, facets of music information systems, such as biographical and related data, can be found in active exploration scenarios, for instance when focusing on the discovery of the work of a specific artist. The aspect of personalization in Phase 3, which is also the basis for serendipitous results in recommendations, is the central feature of current systems. The trends towards context-awareness and adaptive interfaces are ongoing.

As integration of all these requirements is far from trivial and beyond the scope of typical research prototypes, new developments make increasing use of existing and familiar interface elements, e.g. by including or mimicking user

interface elements from Spotify (Millecamp et al., 2018; Jin et al., 2018; Liang and Willemsen, 2019). Nonetheless, research prototypes will continue to fall short of providing the full music platform experience. A notable exception and example of a comprehensive application originating

the 20th International Society for Music Information Retrieval Conference, pages 303–310, Delft, The Netherlands. ISMIR.
Bertin-Mahieux, T., Eck, D., Maillet, F., & Lamere, P. (2008). Autotagger: A model for predicting social

from research, which is successfully adopted outside of lab conditions, is *Songrium* by Hamasaki et al. (2015), which integrates several levels of discovery functions and active music-listening interfaces into a joint application.

To sum up, the evolution of music discovery interfaces has led to the current situation of access to virtually all music catalogs by means of streaming services. On top of that, these services are providing a suite of applications catering to different listening needs and situations. The trend of personalizing listening experiences leads us to believe that, in the not too distant future, music listening will not only be a matter of delivering the right music at the right time, but also of generating and "shaping" the right music for the situation the user is in. We will therefore see a confluence of music retrieval and (interactive) music generation. Beyond this, the topics of explainable recommendations and control over recommendations are gaining importance. Given these exciting perspectives, research in MIR and intelligent user interfaces for music discovery and listening will undoubtedly remain an exciting field to work on.

Notes

- ¹ cf. IFPI Global Music Report 2020 (<https://gmr.ifpi.org>).
- ² <https://songrium.jp>.
- ³ <https://last.fm>.
- ⁴ <https://www.allmusic.com>.
- ⁵ <https://www.pandora.com/about/mpg>.
- ⁶ Interestingly, this has not changed the attitude towards "personal collections" which are nowadays accessed online and frequently organized in playlists (Hagen, 2015; Lee et al., 2016; Cunningham et al., 2017).
- ⁷ <http://weibo.com>.
- ⁸ <https://magenta.tensorflow.org>.
- ⁹ <https://jukebox.openai.com>.
- ¹⁰ <http://infinitejukebox.playlistmachinery.com>.

Competing Interests

The authors have no competing interests to declare.

References

- Andjelkovic, I., Parra, D., & O'Donovan, J.** (2019). Moodplay: Interactive music recommendation based on artists' mood similarity. *International Journal of Human-Computer Studies*, 121, 142–159. Advances in Computer-Human Interaction for Recommender Systems. DOI: <https://doi.org/10.1016/j.ijhcs.2018.04.004>
- Barrington, L., O'Malley, D., Turnbull, D., & Lanckriet, G.** (2009). User-centered design of a social game to tag music. In *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP 2009)*, pages 7–10. DOI: <https://doi.org/10.1145/1600150.1600152>
- Behrooz, M., Mennicken, S., Thom, J., Kumar, R., & Cramer, H.** (2019). Augmenting music listening experiences on voice assistants. In *Proceedings of tags from acoustic features on large music databases*. *Journal of New Music Research*, 37(2), 115–135. DOI: <https://doi.org/10.1080/09298210802479250>
- Bischoff, K., Firat, C. S., Nejdl, W., & Paiu, R.** (2008). Can all tags be used for search? In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pages 193–202, New York, NY, USA. Association for Computing Machinery. DOI: <https://doi.org/10.1145/1458082.1458112>
- Bonnin, G., & Jannach, D.** (2014). Automated generation of music playlists: Survey and experiments. *ACM Computing Surveys*, 47(2), 26:1–26:35. DOI: <https://doi.org/10.1145/2652481>
- Brown, B. A. T., Geelhoed, E., & Sellen, A.** (2001). The use of conventional and new music media: Implications for future technologies. In Hirose, M., editor, *Human-Computer Interaction INTERACT '01: IFIP TC13 International Conference on Human-Computer Interaction*, pages 67–75. IOS Press.
- Bull, M.** (2006). Investigating the culture of mobile listening: From Walkman to iPod. In O'Hara, K. and Brown, B., editors, *Consuming Music Together: Social and Collaborative Aspects of Music Consumption Technologies*, pages 131–149. Springer Netherlands, Dordrecht. DOI: https://doi.org/10.1007/1-4020-4097-0_7
- Burke, R.** (2000). Knowledge-based recommender systems. In *Encyclopedia of Library and Information Systems*, volume 69, pages 180–200. Marcel Dekker, New York, NY, USA.
- Byrd, D., & Fingerhut, M.** (2002). The history of ISMIR – a short happy tale. *D-Lib Magazine*, 8(11).
- Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M.** (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96(4), 668–696. DOI: <https://doi.org/10.1109/JPROC.2008.916370>
- Celma, O.** (2010). *Music Recommendation and Discovery – The Long Tail, Long Tail, and Long Play in the Digital Music Space*. Springer. DOI: <https://doi.org/10.1007/978-3-642-13287-2>
- Chen, L., & Pu, P.** (2012). Critiquing-based recommenders: Survey and emerging trends. *User Modeling and User-Adapted Interaction*, 22(1), 125–150. DOI: <https://doi.org/10.1007/s11257-011-9108-6>
- Cunningham, S. J.** (2019). Interacting with personal music collections. In Taylor, N. G., Christian-Lamb, C., Martin, M. H., & Nardi, B., editors, *Information in Contemporary Society*, pages 526–536, Cham. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-15742-5_50
- Cunningham, S. J., Bainbridge, D., & Bainbridge, A.** (2017). Exploring personal music collection behavior. In Choemprayong, S., Crestani, F., & Cunningham, S. J., editors, *Digital Libraries: Data, Information, and Knowledge for Digital Lives*, pages 295–306, Cham.

- Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-70232-2_25
- Davies, M. E. P., Hamel, P., Yoshii, K., & Goto, M.** (2014). AutoMashUpper: Automatic creation of multi-song music mashups. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(12), 1726–1737. DOI: <https://doi.org/10.1109/TASLP.2014.2347135>
- Deng, S., Wang, D., Li, X., & Xu, G.** (2015). Exploring user emotion in microblogs for music recommendation. *Expert Systems with Applications*, 42(23), 9284–9293. DOI: <https://doi.org/10.1016/j.eswa.2015.08.029>
- Dinculescu, M., Engel, J., & Roberts, A.** (2019). MidiMe: Personalizing a MusicVAE model with user data. In *NeurIPS Workshop on Machine Learning for Creativity and Design*.
- Dittenbach, M., Merkl, D., & Rauber, A.** (2001). Hierarchical clustering of document archives with the growing hierarchical self-organizing map. In *Proceedings of the International Conference on Artificial Neural Networks (ICANN 2001)*. DOI: https://doi.org/10.1007/3-540-44668-0_70
- Downie, J. S.** (2004). The scientific evaluation of music information retrieval systems: Foundations and future. *Computer Music Journal*, 28, 12–23. DOI: <https://doi.org/10.1162/014892604323112211>
- Dredge, S.** (2018). Everybody's talkin': Smart speakers & their impact on music consumption. Technical report, Music Ally report to BPI and ERA.
- Eck, D., Lamere, P., Bertin-Mahieux, T., & Green, S.** (2008). Automatic generation of social tags for music recommendation. In *Advances in Neural Information Processing Systems 20 (NIPS 2007)*.
- Fernández-Tobías, I., Braunhofer, M., Elahi, M., Ricci, F., & Cantador, I.** (2016). Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction*, 26(2–3), 221–255. DOI: <https://doi.org/10.1007/s11257-016-9172-z>
- Ferwerda, B., Tkalcic, M., & Schedl, M.** (2017). Personality traits and music genres: What do people prefer to listen to? In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP 2017)*, pages 285–288. DOI: <https://doi.org/10.1145/3079628.3079693>
- Ferwerda, B., Yang, E., Schedl, M., & Tkalcic, M.** (2015). Personality traits predict music taxonomy preferences. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 2241–2246, New York, NY, USA. Association for Computing Machinery. DOI: <https://doi.org/10.1145/2702613.2732754>
- Flexer, A., Schnitzer, D., Gasser, M., & Widmer, G.** (2008). Playlist generation using start and end songs. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 303–308. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-70232-2_25
- information retrieval. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3), 638–648. DOI: <https://doi.org/10.1109/TASL.2010.2041386>
- Gasser, M., & Flexer, A.** (2009). FM4 Soundpark: Audio-based music recommendation in everyday use. In *Proceedings of the 6th Sound and Music Computing Conference (SMC 2009)*.
- Goto, M.** (2007). Active music listening interfaces based on signal processing. In *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1441–1444. DOI: <https://doi.org/10.1109/ICASSP.2007.367351>
- Goto, M.** (2012). Grand challenges in music information research. In Muller, M., Goto, M., & Schedl, M., editors, *Dagstuhl Follow-Ups: Multimodal Music Processing*, pages 217–225. Dagstuhl Publishing.
- Goto, M., & Dannenberg, R. B.** (2019). Music interfaces based on automatic music signal analysis: New ways to create and listen to music. *IEEE Signal Processing Magazine*, 36(1), 74–81. DOI: <https://doi.org/10.1109/MS.2018.2874360>
- Goto, M., & Goto, T.** (2009). Musicream: Integrated music-listening interface for active, flexible, and unexpected encounters with musical pieces. *IPSJ (Information Processing Society of Japan) Journal*, 50(12), 2923–2936. DOI: <https://doi.org/10.2197/ipsjjip.17.292>
- Govaerts, S., & Duval, E.** (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 2009)*.
- Hagen, A. N.** (2015). The playlist experience: Personal playlists in music streaming services. *Popular Music and Society*, 38(5), 625–645. DOI: <https://doi.org/10.1080/03007766.2015.1021174>
- Hamasaki, M., Goto, M., & Nakano, T.** (2015). Songrium: Browsing and listening environment for music content creation community. In *Proceedings of the 12th Sound and Music Computing Conference (SMC 2015)*, pages 23–30.
- Hariri, N., Mobasher, B., & Burke, R.** (2012). Context-aware music recommendation based on latent topic sequential patterns. In *Proceedings of the 6th ACM Conference on Recommender Systems (RecSys 2012)*, pages 131–138. DOI: <https://doi.org/10.1145/2365952.2365979>
- Hauger, D., & Schedl, M.** (2012). Exploring geospatial music listening patterns in microblog data. In *Proceedings of the 10th International Workshop on Adaptive Multimedia Retrieval (AMR 2012)*.
- Herrera, P.** (2018). *MIRages: An account of music audio extractors, semantic description and context-awareness, in the three ages of MIR*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain.
- Hong, J.-W., & Curran, N. M.** (2019). Artificial intelligence, artists, and art: Attitudes toward artwork produced by humans vs. artificial intelligence. *ACM Transactions on Interactive Intelligence and Learning Systems*, 10(1), 1–23.

- Fujihara, H., Goto, M., Kitahara, T., & Okuno, H. G.** (2010). A modeling of singing voice robust to accompaniment sounds and its application to singer identification and vocal-timbre-similaritybased music

- Hu, Y., Koren, Y., & Volinsky, C.** (2008). Collaborative filtering for implicit feedback datasets. In *Proceedings*

- of the 8th IEEE International Conference on Data Mining (ICDM), pages 263–272. DOI: <https://doi.org/10.1109/ICDM.2008.22>
- Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., Dai, A., Hoffman, M., Dinculescu, M., & Eck, D.** (2019). Music transformer: Generating music with long-term structure. In *Proceedings of the 7th International Conference on Learning Representations (ICLR 2019)*.
- Huber, S., Schedl, M., & Knees, P.** (2012). nepDroid: An intelligent mobile music player. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ACM ICMR 2012)*. DOI: <https://doi.org/10.1145/2324796.2324862>
- Jannach, D., Lerche, L., & Zanker, M.** (2018). Recommending based on implicit feedback. In Brusilovsky, P. and He, D., editors, *Social Information Access: Systems and Technologies*, pages 510–569. Springer International Publishing, Cham. DOI: https://doi.org/10.1007/978-3-319-90092-6_14
- Jin, Y., Cai, W., Chen, L., Htun, N. N., & Verbert, K.** (2019). MusicBot: Evaluating critiquing-based music recommenders with conversational interaction. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19*, pages 951–960, New York, NY, USA. Association for Computing Machinery. DOI: <https://doi.org/10.1145/3357384.3357923>
- Jin, Y., Tintarev, N., & Verbert, K.** (2018). Effects of personal characteristics on music recommender systems with different levels of controllability. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys 2018)*, pages 13–21. DOI: <https://doi.org/10.1145/3240323.3240358>
- John, O. P., Donahue, E. M., & Kentle, R. L.** (1991). *The Big Five Inventory – Versions 4a and 54*. University of California, Berkeley, Institute of Personality and Social Research. DOI: <https://doi.org/10.1037/t07550-000>
- Julia, C. F., & Jordà, S.** (2009). SongExplorer: A tabletop application for exploring large collections of songs. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*.
- Kageyama, T., Mochizuki, K., & Takashima, Y.** (1993). Melody retrieval with humming. In *Proceedings of the 1993 ACM SIGART Conference on Music Information Retrieval (ISMIR 1993)*.
- Knees, P., & Schedl, M.** (2013). A survey of music similarity and recommendation from music context data. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 10(1). DOI: <https://doi.org/10.1145/2542205.2542206>
- Knees, P., Schedl, M., Ferwerda, B., & Laplante, A.** (2019). User awareness in music recommender systems. In Augstein, M., Herder, E., & Wörndl, W., editors, *Personalized Human-Computer Interaction*, pages 223–252. DeGruyter, Berlin, Boston. DOI: <https://doi.org/10.1515/9783110552485-009>
- Knees, P., Schedl, M., Pohle, T., & Widmer, G.** (2006). 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 (ACM Multimedia 2006)*. DOI: <https://doi.org/10.1145/1180639.1180652>
- Lamere, P.** (2008). Social tagging and music information retrieval. *Journal of New Music Research*, 37(2), 101–114. DOI: <https://doi.org/10.1080/09298210802479284>
- Lamere, P., & Eck, D.** (2007). Using 3D visualizations to explore and discover music. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*.
- Law, E. L. M., von Ahn, L., Dannenberg, R. B., & Crawford, M.** (2007). TagATune: A game for music and sound annotation. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, pages 361–364.
- Lee, J. H., Kim, Y.-S., & Hubbles, C.** (2016). A look at the cloud from both sides now: An analysis of cloud music service usage. In *Proceedings of the 17th International Society for Music Information Retrieval Conference*, pages 299–305, New York City, United States. ISMIR.
- Lehtiniemi, A., & Holm, J.** (2013). Designing for music discovery: Evaluation and comparison of five music player prototypes. *Journal of New Music Research*, 42(3), 283–302. DOI: <https://doi.org/10.1080/09298210.2013.796997>
- Leitich, S., & Topf, M.** (2007). Globe of Music – music library visualization using GeoSOM. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*.

- 1993 International Computer Music Conference (ICMC 1993), pages 349–351.
- Kamekhosh, I., Bonnin, G., & Jannach, D.** (2020). Effects of recommendations on the playlist creation behavior of users. *User Modeling and User-Adapted Interaction*, 30, 285–322. DOI: <https://doi.org/10.1007/s11257-019-09237-4>
- Kim, J. H., Tomasik, B., & Turnbull, D.** (2009). Using artist similarity to propagate semantic information. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*.
- Kim, Y. E., Schmidt, E. M., & Emelle, L.** (2008). MoodSwings: A collaborative game for music mood label collection. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 231–236.
- Liang, Y., & Willemsen, M. C.** (2019). Personalized recommendations for music genre exploration. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '19*, pages 276–284, New York, NY, USA. Association for Computing Machinery. DOI: <https://doi.org/10.1145/3320435.3320455>
- Lonsdale, A. J., & North, A. C.** (2011). Why do we listen to music? a uses and gratifications analysis. *British Journal of Psychology*, 102(1), 108–134. DOI: <https://doi.org/10.1348/000712610X506831>
- Louie, R., Coenen, A., Huang, C. Z., Terry, M., & Cai, C. J.** (2020). Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA. Association

- for Computing Machinery. DOI: <https://doi.org/10.1145/3313831.3376739>
- Lu, F., & Tintarev, N.** (2018). A diversity adjusting strategy with personality for music recommendation. In *Proceedings of the 5th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, co-located with ACM Conference on Recommender Systems (RecSys 2018)*.
- Lübbbers, D., & Jarke, M.** (2009). Adaptive multimodal exploration of music collections. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*.
- Mai, J.-E.** (2011). Folksonomies and the new order: Authority in the digital disorder. *Knowledge Organization*, 38(2), 114–122. DOI: <https://doi.org/10.5771/0943-7444-2011-2-114>
- Mandel, M. I., & Ellis, D. P.** (2008). A web-based game for collecting music metadata. *Journal of New Music Research*, 37(2), 151–165. DOI: <https://doi.org/10.1080/09298210802479300>
- Mandel, M. I., Pascanu, R., Eck, D., Bengio, Y., Aiello, L. M., Schifanella, R., & Menczer, F.** (2011). Contextual tag inference. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 7S(1), 32:1–32:18. DOI: <https://doi.org/10.1145/2037676.2037689>
- McFee, B., & Lanckriet, G.** (2011). The natural language of playlists. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*.
- Millicamp, M., Htun, N. N., Jin, Y., & Verbert, K.** (2018). Controlling Spotify recommendations: Effects of personal characteristics on music recommender user interfaces. In *Proceedings of the 26th Conference Systems and Technology*, 8(2). DOI: <https://doi.org/10.1145/2926718>
- Pampalk, E., Dixon, S., & Widmer, G.** (2004). Exploring music collections by browsing different views. *Computer Music Journal*, 28(2), 49–62. DOI: <https://doi.org/10.1162/014892604323112248>
- Pampalk, E., & Goto, M.** (2006). MusicRainbow: A new user interface to discover artists using audiobased similarity and web-based labeling. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*.
- Pampalk, E., & Goto, M.** (2007). MusicSun: A new approach to artist recommendation. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*.
- Pampalk, E., Rauber, A., & Merkl, D.** (2002). Content-based organization and visualization of music archives. In *Proceedings of the 10th ACM International Conference on Multimedia (MM 2002)*, pages 570–579, Juan les Pins, France. DOI: <https://doi.org/10.1145/641007.641121>
- Pohle, T., Knees, P., Schedl, M., Pampalk, E., & Widmer, G.** (2007). "Reinventing the Wheel": A novel approach to music player interfaces. *IEEE Transactions on Multimedia*, 9(3), 567–575. DOI: <https://doi.org/10.1109/TMM.2006.887991>
- Prockup, M., Ehmann, A. F., Gouyon, F., Schmidt, E., Celma, Ò., & Kim, Y. E.** (2015). Modeling genre with the Music Genome Project: Comparing human-labeled attributes and audio features. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Málaga, Spain.
- Raimond, Y.** (2008). *A Distributed Music Information System*. PhD thesis, Queen Mary University of London,

- on User Modeling, Adaptation and Personalization*, UMAP '18, pages 101–109, New York, NY, USA. Association for Computing Machinery. DOI: <https://doi.org/10.1145/3209219.3209223>
- Mörchen, F., Ultsch, A., Nöcker, M., & Stamm, C.** (2005). Databionic visualization of music collections according to perceptual distance. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*.
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L.** (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE*, 9(2), 1–23. DOI: <https://doi.org/10.1371/journal.pone.0089642>
- Neumayer, R., Dittenbach, M., & Rauber, A.** (2005). PlaySOM and PocketSOMPlayer, alternative interfaces to large music collections. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*.
- North, A. C., Hargreaves, D. J., & Hargreaves, J. J.** (2004). Uses of music in everyday life. *Music Perception*, 22(1), 41–77. DOI: <https://doi.org/10.1525/mp.2004.22.1.41>
- Oramas, S., Ostuni, V.C., Noia, T.D., Serra, X., & Sciascio, E. D.** (2016). Sound and music recommendation with knowledge graphs. *ACM Transactions on Intelligent*
- UK.
- Raimond, Y., Abdallah, S., Sandler, M., & Giasson, F.** (2007). The Music Ontology. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria.
- Rentfrow, P. J., & Gosling, S. D.** (2003). 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. DOI: <https://doi.org/10.1037/0022-3514.84.6.1236>
- Roberts, A., Engel, J., Oore, S., & Eck, D.** (2018a). Learning latent representations of music to generate interactive musical palettes. In *Proceedings of the 2018 ACM Workshop on Intelligent Music Interfaces for Listening and Creation (MILC 2018)*.
- Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D.** (2018b). A hierarchical latent vector model for learning long-term structure in music. In *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*, pages 4364–4373.
- Sasaki, S., Yoshii, K., Nakano, T., Goto, M., & Morisihima, S.** (2014). LyricsRadar: A lyrics retrieval system based on latent topics of lyrics. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 585–590.

- Schedl, M., Gómez, E., Trent, E., Tkalcic, M., Eghbal-Zadeh, H., & Martorell, A.** (2018). On the interrelation between listener characteristics and the perception of emotions in classical orchestra music. *IEEE Transactions on Affective Computing*, 9, 507–525. DOI: <https://doi.org/10.1109/TAFFC.2017.2663421>
- Schedl, M., Höglinger, C., & Knees, P.** (2011a). Large-scale music exploration in hierarchically organized landscapes using prototypicality information. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ACM ICMR 2011)*. DOI: <https://doi.org/10.1145/1991996.1992004>
- Schedl, M., Knees, P., McFee, B., Bogdanov, D., & Kaminskas, M.** (2015). Music recommender systems. In Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B., editors, *Recommender Systems Handbook*, pages 453–492. Springer, 2nd edition. DOI: https://doi.org/10.1007/978-1-4899-7637-6_13
- Schedl, M., Widmer, G., Knees, P., & Pohle, T.** (2011b). A music information system automatically generated via web content mining techniques. *Information Processing & Management*, 47, 426–439. DOI: <https://doi.org/10.1016/j.ipm.2010.09.002>
- Schnitzer, D., Pohle, T., Knees, P., & Widmer, G.**
- Information Retrieval Conference (ISMIR 2018)*, pages 561–568.
- Tsukuda, K., Ishida, K., & Goto, M.** (2017). Lyric Jumper: A lyrics-based music exploratory web service by modeling lyrics generative process. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, pages 544–551.
- Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G.** (2007a). Towards musical query-by-semanticdescription using the CAL500 data set. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR 2007)*. DOI: <https://doi.org/10.1145/1277741.1277817>
- Turnbull, D., Liu, R., Barrington, L., & Lanckriet, G.** (2007b). A game-based approach for collecting semantic annotations of music. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria.
- Tzanetakis, G., Essl, G., & Cook, P.** (2001). Automatic musical genre classification of audio signals. In *Proceedings of the 2nd International Symposium on Music Information Retrieval (ISMIR 2001)*, pages 205–210.

- (2007). One-touch access to music on mobile devices. In *Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia (MUM 2007)*, pages 103–109. DOI: <https://doi.org/10.1145/1329469.1329483>
- Slaney, M.** (2011). Web-scale multimedia analysis: Does content matter? *IEEE MultiMedia*, 18(2), 12–15. DOI: <https://doi.org/10.1109/MMUL.2011.34>
- Slaney, M., & White, W.** (2007). Similarity based on rating data. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, pages 479–484.
- Sordo, M.** (2012). *Semantic Annotation of Music Collections: A Computational Approach*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain.
- Stewart, R., Levy, M., & Sandler, M.** (2008). 3D interactive environment for music collection navigation. In *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*.
- Stober, S., & Nürnberg, A.** (2010). MusicGalaxy – an adaptive user-interface for exploratory music retrieval. In *Proceedings of the 7th Sound and Music Computing Conference (SMC 2010)*, pages 23–30.
- Sturm, B. L. T., Iglesias, M., Ben-Tal, O., Miron, M., & Gómez, E.** (2019). Artificial intelligence and music: Open questions of copyright law and engineering praxis. *Arts*, 8(3). DOI: <https://doi.org/10.3390/arts8030115>
- Surowiecki, J.** (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday.
- Takahashi, T., Fukayama, S., & Goto, M.** (2018). Instrudive: A music visualization system based on automatically recognized instrumentation. In *Proceedings of the 19th International Society for Music*
- Ultsch, A., & Siemon, H. P.** (1990). Kohonen's selforganizing feature maps for exploratory data analysis. In *Proceedings of the International Neural Network Conference (INNC 1990)*, pages 305–308.
- Vad, B., Boland, D., Williamson, J., Murray-Smith, R., & Steffensen, P. B.** (2015). Design and evaluation of a probabilistic music projection interface. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*, pages 134–140.
- van der Maaten, L., & Hinton, G.** (2008). Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.
- van Gulik, R., & Vignoli, F.** (2005). Visual playlist generation on the artist map. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pages 520–523.
- Vembu, S., & Baumann, S.** (2004). 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 2004)*. DOI: https://doi.org/10.1007/978-3-540-31807-1_9
- Wang, X., Rosenblum, D., & Wang, Y.** (2012). Context-aware mobile music recommendation for daily activities. In *Proceedings of the 20th ACM International Conference on Multimedia (ACM Multimedia 2012)*, pages 99–108. DOI: <https://doi.org/10.1145/2393347.2393368>
- Whitman, B., & Ellis, D. P. W.** (2004). Automatic record reviews. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, pages 470–477.
- Wold, E., Blum, T., Keislar, D., & Wheaton, J.** (1996). Content-based classification, search, and retrieval of audio. *IEEE MultiMedia*, 3(3), 27–36. DOI: <https://doi.org/10.1109/93.556537>

- Wu, Y., & Takatsuka, M.** (2006). Spherical selforganizing map using efficient indexed geodesic data structure. *Neural Networks*, 19(6–7), 900–910. DOI: <https://doi.org/10.1016/j.neunet.2006.05.021>
- Yakura, H., Nakano, T., & Goto, M.** (2018). Focus-MusicRecommender: A system for recommending music to listen to while working. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces (ACM IUI 2018)*, pages 7–17. DOI: <https://doi.org/10.1145/3172944.3172981>
- Yoshii, K., Goto, M., Komatani, K., Ogata, T., & Okuno, H. G.** (2007). Drumix: An audio player with functions

- of realtime drum-part rearrangement for active music listening. *IPSJ (Information Processing Society of Japan) Journal*, 48(3), 1229–1239. DOI: <https://doi.org/10.2197/ipsjdc.3.134>
- Zhou, C., Jin, Y., Zhang, K., Yuan, J., Li, S., & Wang, X.** (2018). MusicRoBot: Towards conversational context-aware music recommender system. In Pei, J., Manolopoulos, Y., Sadiq, S., & Li, J., editors, *Database Systems for Advanced Applications*, pages 817–820, Cham. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-91458-9_55

How to cite this article: Knees, P., Schedl, M., & Goto, M. (2020). Intelligent User Interfaces for Music Discovery. *Transactions of the International Society for Music Information Retrieval*, 3(1), pp. 165–179. DOI: <https://doi.org/10.5334/tismir.60>

Submitted: 19 March 2020

Accepted: 02 September 2020

Published: 16 October 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Ju[*Transactions of the International Society for Music Information Retrieval* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 

