Chapter 1
Introduction to Music Similarity and Retrieval

Traditionally, electronically searching for music, whether in collections of thousands (private collections) or millions of tracks (digital music resellers), is basically a database lookup task based on meta-data. For indexing a collection, existing music retrieval systems make use of arbitrarily assigned and subjective meta-information like genre or style in combination with objective meta-data like artist name, album name, track name, record label, or year of release. On top of that, often, the hierarchical scheme (genre–) artist–album–track is then used to allow for browsing within the collection. While this may be sufficient for small private collections, in cases where most contained pieces are not known a priori, the unmanageable amount of pieces may easily overstrain the user and impede the discovery of desired music. Thus, a person searching for music, e.g., a potential customer, must already have a very precise conception of the expected result which makes retrieval of desired pieces from existing systems impractical and unintuitive.

Obviously, the intrinsic problem of these indexing approaches is the limitation to a rather small set of meta-data, whereas neither the musical content nor the cultural context of music pieces is captured. Archival and retrieval of music is historically a librarian’s task, and structure and format of databases are optimized for access by experts. Today, the majority of users are not experts—neither in database search nor in terms of musical education. When searching for music, particularly when trying to discover new music, users rarely formulate their queries using bibliographic terms but rather describe properties like emotion or usage context [207]. Therefore, different search and retrieval scenarios become more important.

Retrieval systems that neglect musical, cultural, and personal aspects are far away from the manifold ways that people organize, deal with, and interact with music collections—often expressed in information retrieval (IR) terms, these systems neglect their users’ music information needs [90, 258]. For music, information needs can be quite distinct from standard text-related information needs. Music as a media is heavily intertwined with pop culture as well as with hedonistic and recreational activities. The need to find music might not be as much one that is targeted at
information but merely one targeted at pure entertainment. Thus, one could argue that for most popular and mainstream music, the average user accessing a music information system has primarily an entertainment need (cf. [34, 325]).

1.1 Music Information Retrieval

As a response to the challenges, specifics, and needs of retrieval in the music domain, the area of research known as music information retrieval (MIR) has evolved in the 1990s and emancipated itself as a dedicated field at the beginning of the millennium with the organization of the ISMIR\(^1\) conference series [54]. Among others, MIR is researching and developing intelligent methods that aim at extracting musically meaningful descriptors either directly from the audio signal or from contextual sources. These descriptors can then be used, e.g., to build improved interfaces to music collections. In this section, we give an overview of the field of MIR. We start by looking into definitions found in the literature and proceed by describing the predominant retrieval paradigms found in MIR, illustrated by exemplary tasks and applications. We round this overview up by pointing to research areas of MIR that go beyond traditional IR tasks.

In the literature, one can find several definitions of MIR—each focusing on specific aspects. We give a selection of these in order to sketch the bigger picture. In an early definition, Futrelle and Downie emphasize the multi- and interdisciplinarity of MIR and its origins in digital library research:

MIR is a(n) ... interdisciplinary research area encompassing computer science and information retrieval, musicology and music theory, audio engineering and digital signal processing, cognitive science, library science, publishing, and law. Its agenda, roughly, is to develop ways of managing collections of musical material for preservation, access, research, and other uses. [138]

Later on, Downie highlights research on content analysis, i.e., the automatic extraction of music descriptors from the audio signal itself, interfaces, and infrastructure:

MIR is 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. [110]

Finally, in our own definition, we highlight the multimodality of the field:

MIR is concerned with the extraction, analysis, and usage of information about any kind of music entity (e.g., a song or a music artist) on any representation level (for example, audio signal, symbolic MIDI representation of a piece of music, or name of a music artist). [401]