

# Chapter 1

## Introduction

### 1.1 Motivation

Music and emotions have always been interwoven. Would we still listen to music if it didn't affect us emotionally? Would a composer create music without wanting to express emotions? Emotion is one of the main elements considered when people listen to music as well as when they create it.

Through the development of computer technology, particularly machine learning and content analysis, automatic emotion detection in music files has become possible. Once taught to recognize emotions, computers can exceed human capabilities in the quantity and accuracy of performed analyses of compositions. More and more frequently, systems that search Internet music databases have been adding the select an emotion option to the basic search parameters, which include such things as title, composer, genre, etc.

As a professional musician, I have always been fascinated with expressing emotions through music. Also, analysis of musical compositions taking emotions into account provides us with interesting new insights into their construction. How did, Beethoven, for example, shape the emotions of his compositions so that they are now considered masterpieces? How do the compositions of one composer differ emotionally from another? Why do some compositions affect us with a whole range of emotions while others only one? Can the way an emotion is shaped over time in a musical composition be seen and visualized? These are the questions I tried to answer in this work.

The aim of this book is to present the stages of building automated systems for music emotion recognition. This includes conducting experiments on various music file formats and using different approaches, in the direction of creating emotion maps of musical pieces. Another objective is to indicate some uses for the obtained emotion maps, in the form of systems detecting patterns in the course of emotions or systems for comparing musical pieces, taking into account the shaping of the emotions.

This book presents the particular stages of my research on emotion detection in music. At first, I studied emotions in MIDI files using the categorical approach, which was connected with creating my own MIDI features for detecting emotions. Then, I conducted experiments on recognizing emotion classes in audio files with features extracted using audio analysis tools tailored to Music Information Retrieval. The next stage was applying the dimensional approach to studying audio files and the creation of emotion maps on the arousal-valence emotion plane, which visualized the emotional structure of musical pieces over time. The results of the last stages introduce new, and nowhere before presented by other authors, research on comparing different performances of the same composition using emotion tracking, and finding performances that are more and less similar. The applications of the presented emotion maps of music files can vary widely, and this work does not exhaust them all, but just initiates them.

## 1.2 Organization of This Book

This book is divided into three parts. Part I focuses on representations of emotions in music as well the process of creating music data sets. The content presented in this part is intensively used in the remaining two parts, devoted to emotion detection in MIDI files in Part II and emotion detection in audio files in Part III.

In Chap. 2, I explain two popular approaches used to describe emotions, categorical and dimensional. Different models based on a discrete number of classes as well as models specifying emotion type using an axis on the emotional space are presented. The selected emotion models discussed here, which include four basic emotions—happy, angry, sad and relaxed—in the categorical approach and Russell’s model in the dimensional, were then used in later experiments.

In Chap. 3, I present the process of creating ground truth data for emotion detection in MIDI and audio files. The process of music file annotation by music experts with a university music education is described. The collected ground truth is used in the remaining chapters.

Chapter 4 opens Part II, which focuses on emotion detection in MIDI files. This chapter presents a set of features extracted from MIDI files, assembled into four groups: rhythm, harmony, harmony-rhythm, and dynamic. It also introduces feature calculation methods and their potential to individually discriminate between emotion categories.

Chapter 5 puts forward emotion detection in classical music pieces in MIDI format; a hierarchical categorical model of emotions consisting of two levels was used. During feature selection, the most useful MIDI features were found for building a classifier recognizing four emotions.

Chapter 6 is the first chapter of Part III, which presents issues connected with emotion detection in audio files. This chapter focuses on some of the most relevant audio features for emotion detection in music files. These features were divided into three groups: timbre, rhythm, and tonal. Their meaning is presented and an analysis

of the distribution of their values for audio excerpts labeled using four basic emotions was carried out.

In Chap. 7, I conducted experiments for detecting emotions in audio files using the categorical approach. I built classifiers for different combinations of feature sets, enabling distinguishing the most useful ones for individual emotions. The result of emotion tracking in music files is emotion maps, which visualize the distribution of four emotions over time.

Chapter 8 proposes a system for the analysis of emotions contained within radio broadcasts, which is a practical application of the categorical approach for emotion detection in audio files from the previous chapter. The obtained results provide a new, interesting view of the emotional content of radio station broadcasts.

Chapter 9 focuses on building emotion maps of musical compositions using the dimensional approach. Emotion recognition was treated as a regression problem, and a two-dimensional valence-arousal model was used to measure emotions. I also examined the influence of different audio feature sets—low-level, rhythm, tonal, and their combination—on arousal and valence prediction. On the basis of the created emotion maps, I propose selected features to analyze and compare musical compositions taking into account changes in arousal and valence over time.

Chapter 10 describes the final, most complex system for comparative analysis of musical performances by using emotion tracking. It is an example of applying the dimensional approach for emotion detection in audio files. Here, we discover which performances of the same composition are more similar and which are quite distant in terms of the shaping of arousal and valence over time.