On the Interrelation between Listener Characteristics and the Perception of Emotions in Classical Orchestra Music

Markus Schedl, Emilia Gómez, Erika S. Trent,
Marko Tkalčič, Hamid Eghbal-Zadeh, and Agustín Martorell

Abstract—This study deals with the strong relationship between emotions and music, investigating three main research questions: (RQ1) Are there differences in human music perception (e.g., emotions, tempo, instrumentation, and complexity), according to musical education, experience, demographics, and personality traits?; (RQ2) Do certain perceived music characteristics correlate (e.g., tension and sadness), irrespective of a particular listener's background or personality?; (RQ3) Does human perception of music characteristics, such as emotions and tempo, correlate with descriptors extracted from music audio signals?
To investigate our research questions, we conducted two user studies focusing on different groups of subjects. We used web-based surveys to collect information about demographics, listening habits, musical education, personality, and perceptual ratings with respect to perceived emotions, tempo, complexity, and instrumentation for 15 segments of Beethoven's 3rd symphony, "Eroica". Our experiments showed that all three research questions can be affirmed, at least partly. We found strong support for RQ2 and RQ3, while RQ1 could be confirmed only for some perceptual aspects and user groups.

Index Terms—emotion perception in music, classical music, audio analysis, personality, user study, agreement and correlation in music perception

✦

1 INTRODUCTION

Music has always been closely related to human emotions. It can express emotions and humans can perceive and experience emotions when listening to music, as music cognition research has demonstrated over the last recent decades, e.g., [1], [2], [3]. Often, emotions, which are being constantly triggered by stimuli, may disrupt our daily lives. Hence, we are being constantly engaged in emotion-regulation activities. Music is considered an important factor to regulate emotions, not least used in therapeutic settings. Londsdale and North [4] even identified emotion regulation as the main reason why people actively listen to music. In order to be able to support affect regulation through tools such as music recommender systems or playlist generators, there is a need to better understand the relationship between the characteristics of the listener, the characteristics of the music pieces, and the emotional influence that these music pieces have on the listener.

In this paper, we present two studies that help gaining such a deeper understanding of the multifaceted ways different people perceive emotions and other related aspects in music, focusing on classical orchestra music. To this end, we characterize listeners by demographics, music knowledge and experience, and personality. The music material under investigation consists on 15 excerpts of *Beethoven's 3rd*

Symphony, "Eroica", a prototypical piece of diverse musical resources and potential to induce varied emotions. This material is described by a large set of music audio features (related to tempo, rhythm, harmony, and timbre). Responses of the listeners to the music are recorded via ratings of perceived emotions, tempo, complexity, and instrumentation. In this basic setting, we formulate and investigate the following research questions:

RQ1: Are there differences in human music perception (e.g., emotions, tempo, instrumentation, and complexity), according to musical education, experience, demographics, and personality traits?

RQ2: Do certain perceived music characteristics correlate (e.g., tension and sadness), irrespective of a particular listener's background or personality?

RQ3: Does human perception of music characteristics, such as emotions and tempo, correlate with descriptors extracted from music audio signals?

The remainder of the paper is organized as follows. In Section 2, we review related work that studies human music perception in terms of emotions and other musical characteristics, work that automatically infers emotions from music via acoustic and context features — a task known as music emotion recognition (MER) —, and personality studies in music. Section 3 subsequently describes in detail the experimental setup and execution of the two user studies we conducted to investigate our research questions. Section 4 provides a thorough analysis and discussion of the results and sheds light on the answers to our research questions. We round off by giving concluding remarks and indicating directions for future work in Section 5.

- *Markus Schedl and Hamid Eghbal-Zadeh are with the Department of Computational Perception, Johannes Kepler University Linz, Austria.
  Emilia Gómez and Agustín Martorell are with the Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain.
  Erika S. Trent is with the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA.
  Marko Tkalčič is with the Free University of Bozen-Bolzano, Italy.*

2 RELATED WORK

Previous literature can be categorized into *studies on affective perception of music*, *automated emotion recognition in music*, and *personality studies in music*.

2.1 Studies on Affective Perception of Music

Before engaging in a discussion of the relationship between affect and music, it is important to recognize the distinctions between certain terms that are often considered synonyms in their everyday uses, but entail critical differences in this context. The first distinction to be made is between the terms "affect", "mood", and "emotion". *Affect* is a general term that refers to the positive or negative valence of an emotional experience [5]. A *mood* is a long-lasting experience that is without an identifiable stimulus event, whereas an *emotion* is a briefly lasting experience with an identifiable stimulus event. Some suggest that emotions, unlike moods, are associated with facial expressions, e.g., Ekman [6]. The present study focuses on brief affective experiences triggered by short musical excerpts, thus *emotions* perceived when listening to music.

The second important distinction, as made by Juslin et al. [7], distinguishes between three types of emotions related to music: expressed, perceived, and induced. *Expressed emotions* are the ones that the composer or performer wants to express. A study cited in [7] shows that composers are able to compose music pieces that, when performed by a synthesizer, are perceived as the composer wanted. *Perceived emotions* are how a listener perceives (but not necessarily feels) the meant/expressed emotion of a music piece. Studies cited in [7] show that listeners show a high inter-rater agreement on broad emotional categories, but show less agreement on emotional nuances. The *induced emotion* is the one truly felt by the listener. There is a slight difference between perceived and induced emotion. A listener may perceive a piece of music as sad, but at the same time be in a happy emotional state. The difference is well summed up by the question that Juslin et al. use in their questionnaire: "If you perceive that the music expresses a certain emotion, do you also feel that emotion?". The present study focuses on *perceived* emotions.

Researchers have taken one of three main approaches to conceptualize emotions in relation to music: the categorical approach, the dimensional approach, and the prototype approach [8]. The *categorical approach* suggests that people experience emotions as one of several distinct and recognizable categories. Early categorical studies such as Hevner's [1] initially adopted this approach with a list of adjectives, which has been revised since then [9]. One of the most recent categorical models of music-induced emotions is the *Geneva Emotional Music Scale* (GEMS) [10]. It is a 9-factorial model of music-induced emotions developed as the result of four interrelated studies that compiled and categorized music-relevant emotion terms generated by hundreds of listeners. The model is domain-specific to music; Zentner et al.'s [10] fourth study demonstrated that the GEMS accounts for music-induced emotions better than other basic dimensional models. The 9 main GEMS emotions are "transcendence", "wonder", "joyful activation", "power", "tension", "sadness", "tenderness", "nostalgia", and "peacefulness" [10]. A subset of these 9 emotions
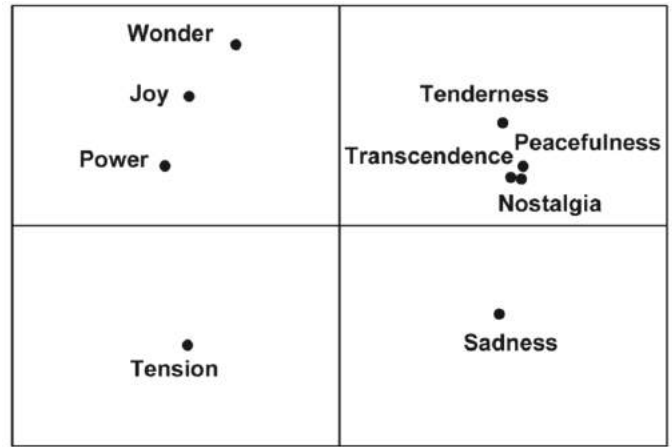

Fig. 1: Mapping between the GEMS emotional categories and the dimensions excited-calm (horizontal axis from left to right) and positive-negative (vertical axis from top to bottom). Source: [13].

was used for the present studies. The *dimensional approach* identifies emotions according to their location on a 1-dimensional to 3-dimensional scale of dimensions such as valence, arousal, activity, dominance, and potency. While there is some debate as to what the third dimension should be in the 3-dimensional case, this approach has enabled listeners to track their changing emotional responses to music in real-time, in a continuous manner (continuous measurement is discussed in the next subsection). The most widely adapted valence-arousal model has been confirmed to be a valid metric in several studies [11], [12]. The relation between the categorical model GEMS and the dimensional valence-arousal model has been studied [13] and we report the mapping in Figure 1. Finally, the *prototype approach* is built on the idea that emotions are categorical, but have hierarchical relationships to one another [14].

Previous studies have suggested that certain musical parameters especially influence the content of emotional responses, notably timbre, orchestration, acoustics, rhythm, melody, harmony, and structure [7]. Several studies have created mappings between musical descriptors and emotion categories [5], but these emotion categories are limited to the five emotions "happiness", "sadness", "anger", "fear", and "tenderness" [15]. Rentfrow et al. identified five genre-free latent factors that reflect the affective response of users to music [16]. They named them "mellow", "urban", "sophisticated", "intense", and "campestral" music preference factors, yielding the acronym *MUSIC*.

Secondary emotions were studied by Ronda et al. [17]. The authors explored the relationships between low-level features in a set of classical music pieces and secondary emotions, such as potency, tension and energy. They found that especially potency is correlated with features, such as roughness, loudness, and spectral flux.

Except for our studies at hand, not many others have focused on how listeners of different demographic and musical background experience different emotional interpretations of the same music. While there do exist several cross-cultural studies on music and perceived emotions [18], [19], these studies tend to focus on greatly different cultures,

rather than on more subtle differences such as age, gender, and musical experience or exposure.

2.2 Automated Emotion Recognition in Music

Based on the various approaches to conceptualize emotions, researchers have developed methods to quantify and measure emotions perceived through music, and subsequently create algorithms that learn relationships between music descriptors and emotions. This task is known as music emotion recognition (MER) and has recently been receiving considerable attention [20], [21]. This is not only evidenced by a wealth of publications on the topic, including [22], [23], [24], [25], [26], [27], [28], [29], [30], but also by the music mood classification tasks that have been running annually since 2007 as part of the "Music Information Retrieval Evaluation eXchange" (MIREX),[1] the most important annually run benchmarking activity for music information retrieval.

One common distinction between MER methods is whether the measurement of emotion is discrete or continuous. In *continuous* measurements, listeners continuously adjust their emotional response in real-time while listening to the music. Data acquisition is in this case, for instance, realized by moving a knob vertically (in the case of a 1-dimensional approach, such as rating the strength of one emotion from "weak" to "strong"), or dragging a cursor on a computer screen (in the case of a 2-dimensional approach, that employs, e.g., the valence-arousal model). A *discrete* measurement, on the other hand, requires listeners to give their response to a musical excerpt as a whole, without the time factor. Some researchers have argued that adopting a continuous (or dynamic) approach is important for emotion recognition in music, because music itself continuously changes through time [31].

Being a *classification* or *regression* task, respectively, for discrete or continuous emotion measurement, MER approaches commonly employ k-nearest neighbor classification [26], [32], support vector machines with different kernels [26], [28], [33], or random forest classifiers [26], in the case of discrete emotions. Frequently used regression methods for continuous MER include logistic regression [26], multiple linear regression [22], [27], and support vector regression [24].

Computational features exploited in MER range from descriptors obtained from audio signals (related to loudness, tempo, rhythm, spectrum, harmony, or melody), features extracted from the score, from lyrics, or contextual data sources, including music-related web pages or collaborative tags. Most works use either audio or lyrics features, or a combination of both. Laurier and Herrera [34] provide a list of the most frequent musical features (ex: tempo, pitch, loudness, timbre, vibrato) mapped with emotion categories. Spectral features, such as Mel frequency cepstral coefficients (MFCCs), spectral centroid, and spectral flux, have been found to be simple to extract and perform fairly well for the task [28], [32]. Features extracted from lyrics include TF·IDF representations of generic n-grams, affective terms (e.g., taken from psycholinguistic dictionaries), and stylistic features (e.g., number of unique words, number of words per minute, or ratio of repeated lines) [35]. Among lyrics-based features, TF·IDF weights and affective terms seem to

outperform stylistic descriptors [33]. However, a combination of content-based timbral, rhythmic, tonal, and temporal audio descriptors with TF·IDF weights of lyrics was shown to substantially outperform single source methods [26].

For our studies, we extracted a variety of descriptors from the audio signal (cf. Section 3.1), in particular to address research question RQ3. However, since most classical orchestra music, including that used in our study, has no lyrics, we have to refrain from exploiting this data source. While in this paper we do not directly perform MER since we do not apply machine learning techniques to predict emotions, we study whether different listeners agree on emotions perceived in classical orchestra music (RQ1). In other words, we analyze whether user-generated emotion tags can serve to build a ground truth for MER. We also study the relationship between perceptive music qualities, including emotions and music content descriptors (RQ3). Thus, instead of *predicting* emotions from classical music content, we analyze the *correlation* between the two.

2.3 Personality Studies in Music

Personality accounts for the most important ways in which we differ in our enduring emotional, interpersonal, experiential, attitudinal, and motivational styles [36]. There are several models of personality, one of the most commonly used is the Five Factor Model (FFM), which is composed of the factors *Openness, Conscientiousness, Extraversion, Agreeableness*, and *Neuroticism* [36]. Personality has been related to music preferences in a number of studies. Rentfrow and Gosling showed that personality traits are related to four music preference dimensions: Reflective and Complex, Intense and Rebellious, Upbeat and Conventional, and Energetic and Rhythmic [37]. Furthermore, personality-based stereotypes are strongly correlated with music genre preferences [38]. We already mentioned in the introduction the study that showed that people use music to regulate their emotions [4]. A further investigation revealed that personality accounts for differences in usage of music. According to a study by Chamorro-Premuzic and Furnham, people who score high on *Openness* tend to consume music in a more rational way while people who score high on *Neuroticism* and those who score low on *Extraversion* and *Conscientiousness* tend to consume music to regulate their emotions [39]. Similarly, Ferwerda et al. showed that personality accounts for individual differences in mood regulation [40]. Personality has also been linked to how users tend to perceive and organize music [41]. In the domain of classical music, a number of correlations between personality and preferences for supporting multimedia material in the form of digital program notes for concerts has been identified [42].

The usage of personality for predicting various music-related aspects has been neglected in the past due to the lenghty process of personality acquisition through questionnaires, such as the Big Five Inventory [43] or the International Personality Item Pool [44]. However, in recent years, personality computing, as defined in [45], has taken off yielding many computational models for the automatic detection of personality from user digital traces, such as Facebook [46], Twitter [47], YouTube [48], Instagram [49], audio signals [50] and mobile phone usage [51].

---

1. http://www.music-ir.org/mirex

2.4 Contributions

By addressing the research questions, we provide the following contributions:

- Study human perception of music in different parts of the same piece by gathering and analyzing user ratings on emotions, tempo, complexity, and instrumentation when they listen to different audio excerpts.
- Identify the influence of personality, demographics, and musical background of listeners on how they perceive and agree on certain perceptual music characteristics (emotions, tempo, complexity, and instrumentation).
- Investigate if there exist universal correlations between perceived music characteristics, irrespective of the listeners' background.
- Analyze whether human perception of these music characteristics correlate with certain qualities extracted from the audio.

In order to investigate the research questions, we conducted two user studies on different groups of participants. First, we ran an initial small-scale study focusing on basic demographics and six emotions. Based on our findings and participants' feedback, we designed and executed a second study, including a more detailed user-centric questionnaire, e.g., adding questions about listening habits, concert attendance, and personality traits, as well as an additional five basic human emotions to characterize participants' perception of music. The number of participants in the studies was 26 and 241, respectively.

Addressing the research questions, we generated the contributions stated above, which have several practical applications, e.g., improved music recommendation services or playlist generators that take the analyzed descriptors into account, applications in receptive music therapy, and improved automatic music categorization. The findings could also be integrated into existing mood estimators. Furthermore, research on demographic and cultural influences on recognized emotions could open the gateway to further research on individual differences in emotional responses to the same music.

3 DATA ACQUISITION AND USER STUDIES

In the following, we first report on how we selected the music material for the studies and which content descriptors we extracted from the audio. Subsequently, we detail the studies' setup and recruitment of participants, before providing and discussing basic statistics of the participants.

3.1 Music Material and Descriptors

For the purpose of the studies, we focused on one particular music genre, namely classical orchestra music, and one particular piece, *Beethoven's 3rd symphony, "Eroica"*. There are several reasons for that. First, we can analyze the influence of listeners' background and expertise in classical (symphonic) music, on their perception. Second, as the "Eroica" is a well-known repertoire piece, also known to many who are not much into classical music, we can analyze the influence of listeners' familiarity with this particular piece.

Third, by studying listener responses to different excerpts of the same recording of the piece, we may accurately analyze their variations even within a single musical work, where features such as recording conditions and composition style are consistent throughout the piece. Fourth, the "Eroica" has a distribution of perceived emotions (i.e. more emotions with positive valence than those with negative valence), which resembles the distribution of emotions in music in general as well as in classical music, as reported in [52]. This approach is different from previous emotion studies on mainstream popular music (cf. [5], p. 57), where excerpts from different pieces are selected, having a great range of acoustic and musical properties. In addition, the present study complements the work by Rodà et al. [17] in restricting variations in music stimuli, as they studied the relationship between secondary emotions and audio features in classical music constraining modality and tempo in their stimuli (e.g., first experiment with all excerpts in major mode, second one with all excerpts in major mode and 104 bpm).

Beethoven's "Eroica" is generally agreed on as a key composition of the symphonic repertoire. It is recognized as a paradigm of formal complexity, it contains varied musical resources with potential for inducing varied emotions, and its length is comparable to that of the mainstream symphonic repertoire. This symphony was further selected because it is already well studied in terms of its musical content and is one of the key pieces of the PHENICX project.[2] It is rich and complex enough, yet analyzable and understandable. In this sense, the "Eroica" constitutes a useful prototypical piece of study, with potential results generalization to wider orchestral repertoires. Furthermore, scores, high quality audio recordings, and aligned MIDI files of the "Eroica" are available to us. We considered a performance by the Royal Concertgebouw Orchestra (RCO), Amsterdam.

The excerpts we used in the studies were carefully selected by the authors (trained in music theory and performance) and reviewed by a musicologist. The procedure is summarized as follows. Every section of the "Eroica" was analyzed and labeled with one of the 9 GEMS emotions, judged based on their relation with musical elements (tempo, rhythm, harmony, melody, and orchestration) as identified in the literature [3]. We then identified the six emotions that most frequently appeared: transcendence, peacefulness, power, joyful activation, tension, and sadness. Subsequently, we selected a set of excerpts following three main criteria: they should contain a variety of musical characteristics, last the duration of a complete musical phrase, and strongly represent one of the above six emotions. As a result, we finally selected a total of 15 excerpts of lengths between 10 and 27 seconds: three excerpts each for peacefulness, power, and tension, and two excerpts each for transcendence, joyful activation, and sadness. Given the length of the excerpts and cognitive complexity of the study, a number of 15 was deemed appropriate for the subjects to remain engaged throughout the study. Table 17 in the appendix summarizes the excerpts' musical characteristics, showing the musical richness of the selected material in terms of orchestration, rhythm, dynamics and tonality. For

---

2. http://phenicx.upf.edu

the sake of reproducibility, interested readers can download the excerpts from a dedicated web page,[3] with kind permission of the Royal Concertgebouw Orchestra (RCO).

In order to address research question RQ3, i.e., whether perceived emotions and descriptors extracted from audio signals correlate, we extracted a variety of audio content descriptors from the 15 audio excerpts by means of two well-known software packages. First, we used the MIR Toolbox[4] to select and extract 26 representative music content descriptors. These 26 descriptors were selected to cover different musical facets: loudness, brightness, noisiness, tempo/rhythm, harmony, and timbre as summarized in Table 1. Second, we complemented this feature set with a larger set of 1,728 audio descriptors using the Essentia[5] open-source C++ library for audio analysis and audio-based music information retrieval [53]. We used all possible features to model the music as comprehensively as possible. Audio descriptors were computed with default parameters, and we computed basic statistics and first and second derivatives of instantaneous (frame-based) descriptors. Detailed documentation of the descriptors is provided on the web[6] and summarized in Table 2. Please note that, although some features are repeated in the two considered toolboxes (e.g. Mel-Frequency Cepstrum Coefficients), there are some differences in the way they are implemented. Exact values can be downloaded from our dedicated web page.[3] We additionally considered manual annotations of key and mode.

3.2 Overview of User Studies

We divided the acquisition process of perceptual music ratings, music expertise and knowledge, personality, and demographics into two user studies. These were carried out on two different groups of listeners. In the first study, from now on referred to as the UPF/MTG study, we gained valuable feedback from participants, which led to a considerably extended follow-up study, from now on referred to as JKU/CP study. We detail the studies in the following.

The first study, designed at the Music Technology Group (MTG) at Universitat Pompeu Fabre (UPF), Barcelona, Spain, acquired emotion ratings for our particular music material along with basic user properties (*demographics* and *listening experience*), for a rather small number of participants ($n = 26$), and analyzed *verbal descriptions* and *emotion ratings*. For this study, participants were recruited by posting to the MTG's mailing list, a Barcelona-based gospel choir's mailing list, and social networking sites, i.e., the Facebook and Twitter accounts of the authors. This way, users from Spain, USA, and Japan were recruited. Participants belong to different cultural backgrounds, various ages, and different levels of musical knowledge. The questionnaire (see below) was available in Spanish and English, so subjects of either language proficiency were able to participate.[7] However, due to the way of recruiting, the composition

of participants was slightly biased towards people more knowledgeable in music than the average listener might be.

This fact, together with the feedback we received from participants and findings we made from the analysis of results, gave rise to a follow-up study. In this second study, designed and conducted at the Department of Computational Perception (CP) at Johannes Kepler University (JKU), Linz, Austria, we involved a larger number of participants ($n = 241$), included questions on *personality traits* and recorded more specific details regarding *listening habits* and *experience with classical music*. We further incorporated aspects of music perception with respect to additional musical properties, such as *tempo*, *complexity*, and *instruments*. Participants were recruited by mass mail to all students of JKU and by posting to several research mailing lists. Announcements were also made on various social media platforms the authors are active on. Participants in geographical reach of JKU were further incentivized by offering a compact disc of Beethoven's "Eroica", performed and recorded by RCO. Details on the distribution of participants are given in Section 3.4 for both studies.

3.3 Experimental Setup

Both studies were conducted as online survey, accessible via a web interface. We asked participants a range of questions, split into several categories: *general personal information* related to demographics, music education and experience, inclination to music and to classical music in particular, and familiarity with Beethoven's "Eroica"; *personality* information in the case of the JKU/CP study; and the actual *ratings of perceptual qualities* (emotions, tempo, complexity, and instrumentation). We further asked participants to enter a textual description of each excerpt after providing the numeric ratings. The personality questionnaire we used is the standardized *Ten Item Personality Instrument* (TIPI) [54]. The online interface did not provide any back buttons in order to discourage re-answering questions because we wanted to capture the subjects' initial answers. Screenshots and details of the formulation of questions for the UPF/MTG and the JKU/CP survey are provided in the appendix in Figures 2 and 3, respectively. The options available to participants for each answer, as well as their numeric coding for the following analysis, are provided in Table 3. As can be seen in the table, the JKU/CP survey included more specific details regarding participants' listening habits and experience with classical music, as well as a personality questionnaire, since we also wanted to investigate relationships between music perception and personality traits.

After having provided their general personal and personality information, participants were presented a questionnaire, which they had to fill in for the 15 selected segments, cf. Figures 2 and 3. The **emotion** descriptors for the UPF/MTG study were taken from the *Geneva Emotion Music Scale* (GEMS) [10] and selected as described in Section 3.1: transcendence, peacefulness, power, joyful activation, tension, and sadness. Five basic human emotions identified in psychological literature [55], [56] were added for the JKU/CP study: anger, disgust, fear, surprise, and tenderness. They are illustrated in Figure 2 and question 1 of Figure 3. In the JKU/CP study, we also added neutral ("neither agree nor disagree") and "don't know" options

---

3. http://mtg.upf.edu/download/datasets/phenicx-emotion

4. https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

5. http://essentia.upf.edu

6. http://essentia.upf.edu/documentation/streaming_extractor_music.html

7. All participants from Japan were proficient in English.

TABLE 1: Musical descriptors extracted from excerpts using MIR Toolbox, organized by category.

Category	MIR Toolbox Features	Description
Loudness	RMS mean (μ) and standard deviation (σ)	Statistics of Root Mean Square of audio signal
	Low energy rate	Percentage of frames showing less-than-average energy
Brightness	Brightness	Amount of energy above $1.5 KHz$
	Spectral centroid	Centroid of the magnitude spectrum
Noisiness	Zero crossing rate	Rate of sign-changes of the audio signal
Tempo/Rhythm	Tempo mean (μ), standard deviation (σ)	Beats per minute (bpm)
	Number of note onsets per second	
Harmony	Modality	Numerical probability estimation of major (+1) to minor (-1) key
	Major/Minor	Major (+1) or Minor (-1) scale, according to the sign of Modality descriptor
	Key clarity	Correlation (i.e. strength) value for the estimated key
Timbre	MFCCs	Mel-Frequency Cepstrum Coefficients (13),
		Compact representation of spectrum.
	Attack time	Duration of the attack of the fastest event in the signal

TABLE 2: Musical descriptors extracted from excerpts using Essentia, organized by category.

Category	Essentia Features
Loudness	Loudness, spectral energy, RMS, energy in 27 Bark bands, energy in 40 ERB bands, energy in 40 mel bands, silence rate, temporal centroid.
Timbre, brightness and noisiness	Sensory dissonance, spectral characteristics (centroid, spread, skewnewss, kurtosis, complexity, crest, entropy, decrease, flatness, flux, rolloff, strong peak), zero Crossing rate, first 13 mel frequency cepstrum coefficients (MFCC), first 13 gammatone feature cepstrum coefficients (GFCC), spectral contrast features.
Tempo/Rhythm	Number of detected beats, spectral energy computed on beats segments of audio across the whole spectrum, ratios of energy in 6 frequency bands, bpm value according to detected beats, descriptors characterizing highest and second highest peak of the bpm probability histogram, onset rate (number of detected onsets per second).
Harmony	Pitch, pitch salience and confidence, estimated chord, key and mode, key strength (i.e. clarity), harmonic pitch class profiles (chroma features).

to describe the perceived emotion, as these were frequently requested in initial dry runs among colleagues of the authors. In the JKU/CP study, we further asked participants to indicate the perceived **tempo**, the perceived **complexity**, and the number of **kinds of instruments** for each segment (cf. Figure 3, questions 2, 3, and 4, respectively). Asking for instrumentation rather than individual instruments was motivated by the fact that it seemed too hard — even for experts — to identify, for instance, whether two or three flutes are playing at the same time. Eventually, participants could optionally give an additional description of the segment.

3.4 Statistics of Participation

Table 4 summarizes the distribution of participants in both studies, in terms of demographics and music experience. The preliminary study ran by UPF/MTG had 26 participants complete the survey (12 male and 14 female); each participant took 15 to 20 minutes. Over half of the participants were between ages 30 and 50 ($\mu = 36.8$, $\sigma = 5.2$), the youngest being 15 and the oldest being 58. Participants came from a variety of cultural backgrounds; 13 (50%) participants were of Spanish origin, 5 (19%) from Japan, 4 (15%) from the USA, 2 (8%) from France, and 1 (4%) from each India and South Korea. Most participants reported having some form of musical experience such as playing an instrument or studying music ($\mu = 6.4$ years), and half reported having more than 10 years of musical experience. The majority of participants listen to classical music as often as a few times a month (5 or 19% almost every day, 10 or 38% a few times a month, 3 or 12% a few times a year, and 4 or 15% almost never). Most participants were either completely unfamiliar with the "Eroica" or were somewhat familiar (both 11 or 42%); only 4 (15%) reported being very familiar with the piece.

In the JKU/CP study, 241 participants completed the survey, which took around 40 minutes per participant. We had 123 male and 118 female participants. The vast majority of 217 participants were Austrians, followed by 3 Germans, 3 Italians, 2 Russians, 2 Englishmen, and 2 Spaniards. For all other countries, the number of participants was at most 1. Participation was slightly biased towards younger people, the median age of participants being 25 years, the arithmetic mean $\mu = 27.4$, and the standard deviation $\sigma = 8.5$ years. However, the youngest participants were only 16, while the eldest one was 67. This distribution can be explained by the way we recruited participants (cf. Section 3.2). As for participants' music taste and listening frequency, on average subjects listen to classical music 2.6 hours per week, and to other genres 11 hours per week. Interestingly, the median for listening to classical music (1 hour per week), is much lower than the median of listening to other genres (8 hours per week). It thus seems that participants either love classical music and devote a lot of time to it, or do not listen to it at all. Less than half of the participants play an instrument (median of 0 hours per week), but most had some form of musical education, on average 6.8 years. Participants attend on average 2 classical and 4 non-classical concerts per year, but the median values are again smaller (1 and 2 concerts, respectively). Many participants do not attend concerts at all: 39% do not attend a single classical concert, 22% do not attend a single concert of another genre per year. Most participants were not (72 or 30%) or somewhat (137 or 57%) familiar with Beethoven's "Eroica". Only 32 (14%) indicated to be very familiar with the piece. Analyzing the personality traits we acquired from the JKU/CP participants (cf. Table 5), we observe that subjects tend to regard themselves as open to new experiences, sympathetic, calm, but also dependable (average and median ratings are at least "agree

TABLE 3: Options available to participants for the questions in both surveys (UPF/MTG and JKU/CP), and their numerical encoding for analysis. "—" means not asked for or not available.

Aspect	Options (UPF/MTG)	Options (JKU/CP)	Numeric encoding of answers
Name/Alias	free-form	—	—
Age	free-form	free-form	years
Gender	free-form	male or female	—
Country	free-form	list selection from 193 countries	—
Listening classical	almost never, a few times/year, a few times/month, almost every day	free-form	0–3 (UPF/MTG); hours per week (JKU/CP)
Listening non-classical	—	free-form	hours per week
General musical experience	free-form	—	years
Playing instrument	—	free-form	hours per week
Musical education	—	free-form	years
Concerts classical	—	free-form	attendances per year
Concerts non-classical	—	free-form	attendances per year
Familiar with "Eroica"	unfamiliar, somewhat familiar, very familiar	unfamiliar, somewhat familiar, very familiar	0–2
All personality traits	—	strongly disagree–strongly agree	1–7
All emotions	not at all, a little, somewhat, strongly	strongly disagree, disagree, neither agree nor disagree, agree, strongly agree, don't know	0–3 (UPF/MTG); 0–4, -1 (JKU/CP)
Perceived tempo	—	slow, fast, don't know	0, 1, -1
Perceived complexity	—	very low–very high, don't know	0–4, -1
Kinds of instruments	—	1, 2, 3, 4, more, don't know	1, 2, 3, 4, 5, -1
Description of the excerpt	free-form	free-form	—

a little"). On the other hand, they negate being disorganized, conventional, and anxious (average and median ratings are at most "disagree a little").

Summarizing the major differences between the two groups of subjects in the UPF/MTG and the JKU/CP study, we observe that participants in the former are slightly older and their age distribution is centered around 36 years, with a standard deviation considerably smaller than that of the JKU/CP participants. While the number of participants in the JKU/CP study is more than 9 times as high as in the UPF/MTG study, the former is biased towards Austrians (217 or 90%). However, given the long tradition of classical music in Western and Central Europe in general, and in Austria in particular, we believe that participants represent an adequate sample to study music perception, from an admittedly "Western" perspective. The UPF/MTG study includes more nationalities, still with a focus on Spaniards (13 or 50%). Studying a larger cultural variety will be part of future work.

4 ANALYSIS AND DISCUSSION

In the following, we give a detailed analysis of the study results, with regard to our research questions RQ1–RQ3.

4.1 Agreement Between Listeners

To assess research question RQ1, i.e., whether there exist differences in human music perception according to musical education, experience, personality, and demographics, we first group participants according to their provided general personal characteristics (cf. Section 3.3), taking into account the different nature of participants in the two studies. Subsequently, we compute the agreement on all perceptive aspects under investigation, in terms of standard deviation and Krippendorff's α score for inter-rater agreement [57], where $\alpha \in [0.00 - 0.20], [0.21 - 0.40], [0.41 - 0.60]$ indicate no, slight, and moderate agreement, respectively [58]. We do so for the entire set of participants in each study, as well as for

the participant splits according to personal characteristics. In the latter case, we report the sign and level of difference between the agreement among all participants $\alpha_{all\_subjects}$ and the agreement α_i within each group i sharing similar personal characteristics. If α_i differs substantially from the entire sample (i.e., $|\alpha_i - \alpha_{all\_subjects}| < t$), we highlight two levels of difference: $t = 0.01$ and $t = 0.05$. Results and analysis for all participants are given in the subsequent section, while category-specific discussions are provided in Sections 4.1.2 and 4.1.3, respectively, for general personal characteristics (demographics and music experience) and for personality traits, the latter being only available for the JKU/CP study.

4.1.1 Overall Agreement

Table 6 shows the overall mean ratings, standard deviations, medians, and agreement scores among participants for each investigated aspect, macro-averaged over all segments, for both studies. As measure of agreement we use *Krippendorff's α* [57], computed on the ratings given by participants for each segment separately and subsequently averaged. We excluded from the calculations the "don't know" answers, i.e., treated them as missing values.

We observe that participants of both studies give highest average ratings (columns μ in Table 6) to the aspects of power and tension, followed by transcendence and joyful activation. Lowest ratings are given to sadness and peacefulness. Here, both studies agree almost perfectly. For the additional aspects asked for in the JKU/CP study, we observe highest ratings for surprise and tenderness, followed by fear, anger, and — much below — disgust. Overall, it seems that the aspects ranging in the lower arousal range (sadness, peacefulness, etc.) are perceived to a smaller degree in the material under consideration. Tempo is, on average, neither perceived as particularly low nor high. So is complexity. As for instrumentation, overall, most participants could distinguish 4 kinds of instruments.

TABLE 4: Basic statistics of the participants. JKU/CP participant data is in normal type, UPF/MTG data is italicized.

Aspect	μ	σ	Median	Min.	Max.
Age	27.35, 36.77	8.47, 5.15	25, 36	16, 15	67, 58
Listening classical (JKU/CP scale: hours per week)	2.56	5.20	1	0	40
Listening classical (UPF/MTG scale: almost never (0) – almost every day (3))	*1.62*	*0.98*	*2*	*0*	*3*
Listening non-classical (hours per week)	11.16	11.86	8	0	70
General musical experience (years)	*6.35*	*5.15*	*5*	*0*	*15*
Playing instrument (hours per week)	1.93	4.23	0	0	40
Musical education (years)	6.77	6.39	5	0	33
Concerts classical (attendances per year)	2.43	5.28	1	0	40
Concerts non-classical (attendances per year)	3.93	6.70	2	0	70
Familiar with "Eroica" (unfamiliar (0), somewhat (1), very (2))	0.83, 0.73	0.64, 0.72	1, 1	0, 0	2, 2

TABLE 5: Personality statistics of the JKU/CP participants (only those were asked to provide personality information).

Personality trait	μ	σ	Median	Min.	Max.
Extraverted	4.27	1.88	5	1	7
Critical	4.54	1.68	5	1	7
Dependable	5.27	1.43	6	1	7
Anxious	3.17	1.64	3	1	7
Open to new experiences	5.59	1.27	6	2	7
Reserved	4.41	1.81	5	1	7
Sympathetic	5.39	1.32	6	1	7
Disorganized	2.83	1.69	2	1	7
Calm	5.01	1.56	6	1	7
Conventional	2.84	1.63	2	1	7

TABLE 6: Means (μ), standard deviations (σ), medians, and agreement scores (Krippendorff's α) for investigated aspects of music perception. For emotions, higher than average ratings in column μ are depicted in bold face, lower than average in italics. Highest median values are highlighted in bold face. For Krippendorff's agreement scores, italic font is used to indicate slight agreement (0.21–0.40) and bold face is used to denote moderate agreement (0.41–0.60).

Aspect	JKU/CP					UPF/MTG				
	Scale	μ	σ	Median	α	Scale	μ	σ	Median	α
Transcendence	0–4	**2.215**	1.095	2.0	0.005	0–3	1.072	0.997	1.0	-0.017
Peacefulness	0–4	*1.812*	0.986	2.0	**0.427**	0–3	0.874	1.003	0.0	0.096
Power	0–4	**2.477**	0.937	**3.0**	**0.432**	0–3	1.508	1.055	1.5	0.040
Joyful activation	0–4	**2.048**	1.059	2.0	*0.303*	0–3	1.118	0.961	1.0	0.121
Tension	0–4	**2.318**	1.121	**3.0**	*0.206*	0–3	1.228	0.985	1.0	0.123
Sadness	0–4	*1.233*	0.979	1.0	*0.284*	0–3	0.559	0.813	0.0	0.045
Anger	0–4	*1.204*	1.008	1.5	*0.280*					
Disgust	0–4	*0.808*	0.941	1.0	0.106					
Fear	0–4	*1.292*	1.084	1.5	*0.259*					
Surprise	0–4	*1.790*	1.162	2.0	0.049					
Tenderness	0–4	*1.687*	1.046	1.5	*0.304*					
Tempo	0–1	0.460	0.337	1.0	**0.419**					
Complexity	0–4	2.240	0.864	2.0	0.103					
Instrument kinds	1–5	3.899	0.980	4.0	0.051					

In the UPF/MTG study, we could not find substantial agreement (column α). The JKU/CP study, on the other hand, evidences a low to moderate agreement for most aspects, according to Krippendorff's α. Participants do not (0.00–0.20) or at most slightly (0.21–0.40) agree on most concepts. The values indicating moderate agreement (0.41–0.60) according to [58] are printed in bold, whereas slight agreement is indicated by italics. Highest agreement among the emotion aspects is found for peacefulness and power, while tempo shows the highest agreement among the other investigated aspects. Slight agreement can be observed for joyful activation, tension, sadness, anger, fear, and tenderness. No relevant agreement is observed for transcendence, disgust, surprise, and perceived complexity. There is also a discrepancy between listeners with regard to their ability to distinguish different instrumentations. Whether this is due to different music knowledge and expertise levels will be analyzed in the next sections, where we discuss the influence of general personal characteristics (demographics and music knowledge) and personality traits on the agreement on perceptual aspects.

4.1.2 Influence of Demographics and Music Experience

We investigate the influence of basic user characteristics (demographics and music experience) on the perceptual aspects (emotions, tempo, complexity, instrumentation) by two means: (i) *correlation analysis* between the numerically encoded factors and (ii) *agreement analysis* within groups defined from the user characteristics. We measured (i) using Pearson's correlation coefficient between all numerically encoded user aspects and perceptual ratings and further computed respective two-sided p values for a hypothesis test whose null hypothesis is that the variables are uncorrelated. We investigated (ii) using Krippendorff's α measure for inter-rater agreement.

Correlation Analysis: Tables 7 and 8 summarize the results of the correlation analysis, respectively, for the UPF/MTG and the JKU/CP study. Significant results are printed in bold face. The results of the UPF/MTG study reveal a significant correlation only between familiarity with the piece and transcendence as well as peacefulness. In

the JKU/CP study, on the other hand, we found several significant correlations. In particular the emotional ratings of transcendence and power correlate with some basic user characteristics. Interestingly, these correlations exist almost exclusively for characteristics directly related to classical music (listening classical, concerts classical, familiarity with the piece). Among the non-emotion ratings, the ability to distinguish certain instrument groups seem to depend strongly on the intensity of music education (highest observed correlation of $r = 0.418$) and the amount of time spent playing an instrument ($r = 0.259$).

Agreement Analysis: Tables 9 and 10 show the results of the agreement analysis between participants belonging to different groups, respectively, for the UPF/MTG and the JKU/CP study. For each aspect, we split participants at the median value (cf. Table 4) into two groups, except for the aspect of familiarity with "Eroica", where we kept all three categories: "unfamiliar", "somewhat familiar", and "very familiar", as well as for country, where we investigated only the most frequent ones. We performed this median-split to obtain approximately equally sized groups of listeners, although, as can be seen in the tables, groups are still of different sizes because many participants indicated exactly the median value for some aspects. In this case, those with median values were put into the group with lower values. Only group scores substantially different from scores among all subjects are included and the sign of difference is shown.[8] Their level of difference is illustrated by either '*' or '**', indicating a difference of 0.01 and 0.05 Krippendorff's α.

Summarizing the results for the UPF/MTG study (Table 9), we observe an interesting difference in agreement between younger and older participants. In particular, participants aged 36 or younger show a substantially higher agreement on five out of the six emotions than participants of higher age, who agree to a much smaller extent and disagree to a much higher. This can be especially well seen for the aspects joyful activation and tension ($t \geq 0.05$ higher agreement for the younger vs. $t \geq 0.05$ lower agreement for the older group), but also holds for tension and sadness. Only peacefulness and power is agreed on in both groups. No substantial differences in agreement can be identified between participants of different gender and of different general musical experience. However, infrequent listeners of classical music tend to agree more on four out of six aspects, while classical aficionados tend to disagree much more on the same aspects. A similar observation can be made with regard to familiarity with "Eroica". Knowing the piece seems to lead to less agreement on perceived emotions, which is particularly substantiated for joyful activation and tension. This lower agreement may be explained by a deeper individual interpretation by participants familiar with the piece. A general preference or prejudice with respect to composer, performer, or the piece itself may have an influence too. We also assume that listeners who are familiar with the piece associate their own subjective interpretations or memories while listening.

The main findings from analyzing the ratings in the JKU/CP study (Table 10) mostly agree with those of the

UPF/MTG study. However, unlike the UPF/MTG study, the JKU/CP study does not indicate any substantial influence of age on perceptual agreement. Indeed, only slight differences in 2 out of 14 aspects are found between differently aged participants. This may be explained by the quite different median-split value between the two studies (36 versus 25 years). Indeed, using a split value of 36 years in the JKU/CP data also leads to pronounced agreement differences, in particular for the older participants, but at the same time reduces the number of participants in the older group to 34. In contrast, more pronounced differences are found between genders. While female listeners show more consensus for a range of aspects (joyful activation, tension, anger, fear, and perceived tempo), their male counterparts rather disagree on tension, anger, fear, and tempo. In terms of participants' music preferences, the findings in the JKU/CP study are in line with those of the UPF/MTG study. While infrequent listeners of classical music tend to agree more on perceived emotions (particularly on joyful activation, tension, and tenderness), frequent listeners disagree on the same aspects. Analogously, participants who listen a lot to non-classical music rather tend to agree on emotions, while those not listening to it tend to disagree. We also observe a tendency that those commonly listening to classical music or attending classical concerts, playing an instrument, or having spent more than five years on musical education, rather agree on perceived tempo, while laymen do not. Listeners who attend classical concerts, on the other hand, agree more on peacefulness, power, joyful activation, tension, and sadness, while their counterparts tend to disagree on the same emotions. The former also agree more on perceived tempo and complexity than all subjects. In terms of familiarity, again, listeners who know the "Eroica" well tend to disagree more and to a higher extent on perceived emotions than those less familiar with the piece.

4.1.3 Influence of Personality

Analogous to the investigations for demographics and music experience, we analyze whether personality has an influence on music perception using *correlation analysis* between personality traits and perceptive aspects, and *analysis of agreement* among participants with similar personality traits. Results are summarized in Tables 11 and 12, respectively.

Correlation Analysis: Taking a closer look at the correlation values between personality traits and perceived emotions, we see that correlations exist especially for the traits openness to new experiences, disorganization, and calmness. People who rank high on openness tend to perceive positive aspects such as transcendence, peacefulness, joyful activation, and tenderness more strongly. Listeners who are rather disorganized, on the other hand, perceive negative aspects like sadness, anger, and disgust more strongly. Remarkably, we observe negative correlations between conventionalism and transcendence as well as tension. Hence, highly conventional people perceive the music as less tense, but also less transcendent. As for perceptual aspects other than emotions, we can observe a positive correlation between tempo on the one hand and open, sympathetic, and disorganized personality on the other. As for complexity, there exists only a significant positive correlation with calmness. The ability to distinguish instrument groups seems to be positively influenced by openness, sympatheticness, and

8. For the sake of readability, we report only the sign here. To foster reproducibility, we provide the actual numbers as part of an online appendix available at http://mtg.upf.edu/download/datasets/phenicx-emotion.

TABLE 7: Pearson's correlation coefficients between *demographics and music expertise* and aspects of music perception, for the **UPF/MTG** study. Significant results are depicted in bold face. Significance at $p < 0.05$ and $p < 0.01$ is illustrated by '*' and '**', respectively.

	Trans.	Peace.	Power	Joyful.	Tension	Sadness
Age	-0.013	0.124	0.189	0.339	0.105	0.240
General musical experience	0.018	-0.066	-0.158	0.000	0.039	0.367
Listening classical	0.223	0.034	-0.013	0.176	0.187	0.118
Familiar with "Eroica"	**0.477****	**0.571****	-0.094	-0.014	0.154	0.195

TABLE 8: Pearson's correlation coefficients between *demographics and music expertise* and aspects of music perception, for the **JKU/CP** study. study. Significant results are depicted in bold face. Significance at $p < 0.05$ and $p < 0.01$ is illustrated by '*' and '**', respectively.

	Trans.	Peace.	Power	Joyful.	Tension	Sadness	Anger	Disgust	Fear	Surprise	Tender	Tempo	Compl.	Instr.
Age	**0.155***	0.040	0.102	**0.261****	0.075	-0.081	-0.110	-0.002	**-0.186****	-0.015	0.104	-0.031	-0.019	-0.026
Listening classical	**0.203****	0.112	**0.212****	0.078	0.019	-0.082	-0.090	-0.105	**-0.190****	-0.029	**0.148***	0.028	0.123	**0.192****
Listening non-classical	0.085	0.092	0.121	0.007	0.033	0.028	**0.139***	0.042	0.078	**0.149***	0.054	0.122	0.064	-0.036
Playing instrument	0.085	-0.016	**0.133***	0.010	**0.190****	0.077	0.113	0.073	0.050	0.042	0.014	0.061	0.012	**0.259****
Musical education	**0.140***	-0.073	**0.143***	0.007	**0.170****	0.029	0.101	0.085	0.008	-0.064	0.007	0.077	0.076	**0.418****
Concerts classical	**0.170****	0.065	**0.175****	0.108	**0.192****	-0.015	-0.033	-0.028	-0.065	-0.046	0.076	0.017	0.086	**0.243****
Concerts non-classical	0.114	-0.004	0.048	-0.008	0.099	0.080	0.079	0.061	0.091	0.069	-0.003	0.106	0.045	**0.153***
Familiar with "Eroica"	**0.141***	0.118	**0.211****	**0.184****	0.116	-0.045	0.057	0.026	-0.018	0.004	**0.149***	0.056	0.096	**0.242****

TABLE 9: Agreement scores (Krippendorff's α) for each perceptual aspect (columns) for user groups defined based on *demographics and music expertise* (rows), in the **UPF/MTG** study. The first row shows the agreement scores among all participants. The following rows indicate whether agreement was lower ('-') or higher ('+') in the respective user group. Only scores α_i substantially different from the entire sample (i.e., $|\alpha_i - \alpha_{all\_subjects}| < t$) are depicted. For $t = 0.01$ and $t = 0.05$ values are further highlighted by '*' and '**', respectively. Abbreviations: y = year. (We refrain from analyzing groups split according nationality of participants due to the low number of US-Americans and Japanese in the study.)

	n	Trans.	Peace.	Power	Joyful.	Tension	Sadness
All subjects	26	-0.017	0.096	0.040	0.121	0.123	0.045
Age ≤ 36	13		+**	+**	+**	+**	+**
Age > 36	13		-*	+**	+*	-**	-**
Gender: Male	12	-*		+*	-*	-*	
Gender: Female	14	-*	-*	-*			-*
General musical experience: ≤ 5 y	14	-*	-*	-*	+*		
General musical experience: > 5 y	12		-*	-*		-*	-*
Listening classical: Infrequent (few times/year or less)	11	-*	-*	+*	+*	+*	+*
Listening classical: Frequent (few times/month or more)	15		-*	-*		-*	-*
Familiar with "Eroica": No	11	-*	-**	-*	+**		+**
Familiar with "Eroica": Somewhat	11		+**	+*	-**	-**	-*
Familiar with "Eroica": Very	4	-**		-*	-**	-*	

TABLE 10: Agreement scores (Krippendorff's α) for each perceptual aspect (columns) for user groups defined based on *demographics and music expertise* (rows), in the **JKU/CP** study. The first row shows the agreement scores among all participants. The following rows indicate whether agreement was lower ('-') or higher ('+') in the respective user group. Only scores α_i substantially different from the entire sample (i.e., $|\alpha_i - \alpha_{all\_subjects}| < t$) are depicted. For $t = 0.01$ and $t = 0.05$ values are further highlighted by '*' and '**', respectively. Abbreviations: h = hour, w = week, y = year.

	n	Trans.	Peace.	Power	Joyful.	Tension	Sadness	Anger	Disgust	Fear	Surprise	Tender	Tempo	Compl.	Instr.
All subjects	241	0.005	0.427	0.432	0.303	0.206	0.284	0.280	0.106	0.259	0.049	0.304	0.419	0.103	0.051
Age ≤ 25	124					-*								+*	
Age > 25	117				+*						-*			-*	
Gender: Male	123					-*	-*			-*	+*		-*		-*
Gender: Female	118				+*	+*		+*		+*	-*		+*		
Listening classical: ≤ 1 h/w	153				+*	+*						+*	-*		
Listening classical: > 1 h/w	88		+*		-*	-*	+*					-*	+*		
Listening non-classical: ≤ 8 h/w	129		+*	-*		-*	-*						+*	+*	
Listening non-classical: > 8 h/w	112		-*	+*		+*	+*						-*	-*	
Playing instrument: 0 h/w	127		-*		-*	-*	-*					-*	-*		
Playing instrument: > 0 h/w	114		+*		+*	+*	+*	+*				+*	+*		
Musical education: ≤ 5 y	122		-*			-*	-*	-*	-*	-*		+*	-*		
Musical education: > 5 y	119		+*			+*	+*	+*	+*			-*	+*	-*	
Concerts classical: ≤ 1/y	152		-*	-*	-*	-*	-*					+*	-*		
Concerts classical: > 1/y	89		+*	+*	+*	+*	+*					-*	+*	+*	
Concerts non-classical: ≤ 2/y	137				-*				-*	-*		+*		+*	
Concerts non-classical: > 2/y	104		-*		+*	-*			+*			-*		-*	
Familiar with "Eroica": No	72					-*	-*				-*	-*	-*	-*	-*
Familiar with "Eroica": Somewhat	137		+*		+*	+*		+*	+*			+*	+*	+*	
Familiar with "Eroica": Very	32	-*	+**	-*	-**	-**	+*		-**	-*		-*		-*	-*

calmness, while negatively influenced by conventionalism. A reason might be that conventionalists tend to be cautious

and in turn rather underestimate the number of instrument groups.

Agreement Analysis: Table 12 summarizes the results of the agreement analysis for different user groups.[9] Analyzing agreement on emotions, the most pronounced differences in agreement between various groups are observed for power. Dependable and calm people agree to a substantially higher extent than the overall population, while critical, anxious, reserved, disorganized, and conventional listeners agree considerably less. On joyful activation, there is considerably more agreement among non-anxious, reserved, and disorganized people, while substantially less agreement among people showing inverse values for these personality traits. Looking at the relationship between these traits and joyful activation ratings (cf. Table 11), we observe negative correlations for both anxiousness and reservation. Hence, anxious and reserved people agree more and perceive less joyful activation when listening to the piece. Sympathetic, calm, and open people agree considerably more on whether a piece is peaceful and tend to give higher values to this emotional aspect. In general, the group of calm people agrees substantially more on emotion aspects than any other listener group. This is particularly pronounced for power, anger, and fear. With regard to perceptual aspects other than emotions (perceived tempo, complexity, and instrument groups), we observe that extraverts show a higher agreement on all three, while reserved and non-conventional listeners agree less on all three.

In summary, we believe that research question RQ1 can be affirmed for certain user groups and perceptual aspects (e.g., those familiar with the piece and with classical music in general, those open to new experiences; transcendence, power, and instrumentation), while there is no clear evidence for others (e.g., listening to non-classical music; peacefulness, sadness, disgust, tempo, and complexity).

4.2 Cross-correlation Between Perceptual Aspects

To assess research question RQ2, i.e., whether certain perceived music characteristics correlate, irrespective of a particular listener's background or personality, we compute Pearson's correlation coefficient between all participants' ratings, i.e., $26 \times 15 = 390$ and $241 \times 15 = 3,615$ in the UPF/MTG (cf. Table 13) and JKU/CP (cf. Table 14) data respectively. We do so for all pairs of perceptual aspects. Significant results at a level of $p < 0.01$ are printed in bold face in the tables. From the results of the UPF/MTG study, which are in line with those of the JKU/CP study, we observe moderate negative correlations between peacefulness on the one hand and power (-0.51) as well as tension (-0.48) on the other. Furthermore, power and tension show a moderate positive correlation (0.41). Joyful activation and sadness, in contrast, are negatively correlated (-0.43). None of the studies evidences any notable correlation for transcendence.

In addition to the results of the UPF/MTG study, the JKU/CP study provides further insights. Notable correlations are found between all pairs of aspects anger, fear, and disgust (all between 0.50 and 0.56). Also sadness is correlated with the above three (correlations between 0.31 and

0.47). Tenderness is moderately positively correlated with peacefulness (0.56), but negatively with sadness and power (both -0.34). Peacefulness is, in addition, negatively correlated with anger (-0.43), power (-0.40), tension (-0.35), and fear (-0.32), and positively with joyful activation (0.33). Tension typically comes with anger (0.39), power (0.37), or fear (0.34). Transcendence and surprise do not show remarkable correlations with any other aspects.

Among the non-emotional perceptual aspects, a noteworthy correlation is found between perceived complexity and number of instrument groups (0.32). Hence, a larger number of different instruments seems to increase the perceived complexity of music. A similar, although less pronounced observation is made for complexity and tempo (0.27). Faster excerpts are therefore perceived more complex. While this makes sense for the genre and repertoire under investigation, it might not generalize to other genres, for instance, electronic music with fast and dominant recurring beats. Correlations between emotions and other perceptual aspects are observed in particular between power on the one hand, and tempo (0.28) and complexity (0.24) on the other hand. Hence, music perceived faster or more complex tends to be perceived more powerful, too. Sadness and tempo show a weak negative correlation (-0.20). All in all, we can state that our research question RQ2 is approved for a substantial number of perceptual aspect pairs.

4.3 Correlation Between Perceptual Aspects and Music Descriptors

In this section we investigate our third research question, i.e., whether human perception of music characteristics, such as emotions and tempo, correlate with music descriptors extracted from the music content.

In order to do that, we computed 1,754 descriptors for the 15 audio excerpts using two different toolboxes, the MATLAB MIR Toolbox and the Essentia C++ library, using default parameters. We included basic statistics and first and second derivatives of instantaneous (frame-based) Essentia descriptors. The obtained features are related to different musical facets such as loudness, timbre, rhythm and harmony, as shown in Tables 1 and 2. In addition, we incorporated the manual analysis of tonality (key and mode) shown in Table 17 as a way to contrast automatic computation.

We then computed Pearson's correlation coefficient between subject-generated emotional ratings in both studies and the values of these musical descriptors, and we found high correlation values between certain emotional ratings and music descriptors. In Table 15 we summarize the sign ('+' or '-') of absolute correlation values higher than 0.7 according to emotional rating, descriptor category, and specific descriptor. We chose 0.7 because it is often considered as a lower bound for highly correlated variables [59]. Detailed values are provided in the online appendix: http://mtg.upf.edu/download/datasets/phenicx-emotion.

We observe very high correlations between emotional ratings and descriptors related to different categories: loudness, timbre, rhythm and harmony, most of them agreeing with previous literature in automated emotion recognition in music [5], summarized in Section 2.2. For instance, loudness energy features are positively correlated with high

---

9. Again, for the sake of readability, we report only the sign of difference to the group comprising all subjects and the actual numbers in an online appendix: http://mtg.upf.edu/download/datasets/phenicx-emotion.

TABLE 11: Pearson's correlation coefficients between *personality traits* and aspects of music perception, for the **JKU/CP** study. Significant results are depicted in bold face. Significance at $p < 0.05$ and $p < 0.01$ is illustrated by '*' and '**', respectively.

	Trans.	Peace.	Power	Joyful.	Tension	Sadness	Anger	Disgust	Fear	Surprise	Tender	Tempo	Compl.	Instr.
Extraverted	0.045	0.024	0.120	0.065	0.022	0.031	-0.014	-0.027	0.007	0.041	**0.166****	0.112	0.059	0.065
Critical	0.010	0.031	0.094	0.081	0.049	0.037	-0.035	-0.041	-0.011	**-0.141***	0.043	0.066	0.075	0.052
Dependable	0.054	-0.098	-0.074	-0.098	0.009	-0.049	-0.065	-0.035	0.011	-0.018	0.007	-0.023	-0.075	0.030
Anxious	-0.084	-0.054	-0.108	-0.114	-0.108	-0.003	0.017	0.064	0.055	0.023	-0.089	-0.072	-0.054	-0.086
Open to new experiences	**0.159***	**0.139***	0.108	**0.181****	0.054	0.053	0.010	0.005	-0.003	0.009	**0.222****	**0.173****	0.006	**0.200****
Reserved	-0.049	0.033	-0.112	-0.057	-0.095	-0.038	-0.033	-0.014	-0.045	-0.042	-0.084	-0.026	-0.054	-0.062
Sympathetic	0.077	**0.147***	0.098	0.107	0.059	-0.031	-0.012	0.020	0.026	0.078	**0.166****	**0.148***	0.015	**0.132***
Disorganized	0.076	0.120	0.032	0.083	0.114	**0.167****	**0.157***	**0.146***	0.116	0.111	**0.129***	**0.130***	-0.014	-0.069
Calm	0.076	**0.142***	-0.002	**0.153***	-0.032	-0.023	-0.044	-0.060	0.031	-0.063	**0.132***	0.069	**0.153***	**0.134***
Conventional	**-0.145***	0.099	-0.048	0.012	**-0.135***	0.050	0.087	0.070	0.102	0.008	-0.058	-0.040	-0.002	**-0.129***

TABLE 12: Agreement scores (Krippendorff's α) for each perceptual aspect (columns) for user groups defined based on *personality traits* (rows), for the **JKU/CP** study. The first row shows the agreement scores among all participants. The following rows indicate whether agreement was lower ('-') or higher ('+') in the respective user group. Only scores α_i substantially different from the entire sample (i.e., $|\alpha_i - \alpha_{all\_subjects}| < t$) are depicted. For $t = 0.01$ and $t = 0.05$ values are further highlighted by '*' and '**', respectively.

	n	Trans.	Peace.	Power	Joyful.	Tension	Sadness	Anger	Disgust	Fear	Surprise	Tender	Tempo	Compl.	Instr.
All subjects	241	0.005	0.427	0.432	0.303	0.206	0.284	0.280	0.106	0.259	0.049	0.304	0.419	0.103	0.051
Extraverted: ≤ 5	163											-*	-*		
Extraverted: > 5	78			+*	+*		+*		-*			+*	+*	+*	+*
Critical: ≤ 5	155			+*					+*	+*	-*	+*	+*		
Critical: > 5	86		-*	-*	-*			-*	-*	-*	+*	-*	-*	+*	
Dependable: ≤ 6	198			-*											
Dependable: > 6	43		+*	+**	-*	+*		-*	+*		-*		-*	-*	
Anxious: ≤ 3	146		+*	+*	+*		+*								
Anxious: > 3	95		-*	-*	-*		-*				+*	-*	-*		
Open to new exp.: ≤ 6	180											+*		-*	
Open to new exp.: > 6	61		+*				-*	+*		+*		-*	+*	+*	-*
Reserved: ≤ 5	156			+*	+*								+*	+*	
Reserved: > 5	85			-*	-*		-*		+*		+*	-*	-*		-*
Sympathetic: ≤ 6	195		-*												
Sympathetic: > 6	46		+*	-*		+*	-*	+*	-*				+*		+*
Disorganized: ≤ 2	130		+*	+*	-*	+*						+*	+*	+*	
Disorganized: > 2	111		-*	-*	+*	-*		+*				-*	-*	-*	
Calm: ≤ 6	207			-*			-*								
Calm: > 6	34		+*	+**		+*	+*	+**		+**	+*	-*			+*
Conventional: ≤ 2	125		+*	+*	+*	+*	+*			-*		+*	+*		+*
Conventional: > 2	116		-*	-*	-*	-*	-*			+*		-*	-*	-*	-*

TABLE 13: Cross-correlations (Pearson's correlation coefficients) between aspects of music perception, for the **UPF/MTG** study. Significant results ($p < 0.01$) are depicted in bold face. There are no correlations that are significant at $p < 0.05$, but not at $p < 0.01$. We therefore refrain from illustrating different levels of significance explicitly.

	Trans.	Peace.	Power	Joyful.	Tension	Sadness
Transcendence		0.085	**0.187**	-0.020	**0.160**	0.077
Peacefulness	0.085		**-0.507**	**0.147**	**-0.482**	-0.083
Power	**0.187**	**-0.507**		**0.212**	**0.411**	**-0.227**
Joyful activation	-0.020	**0.147**	**0.212**		**-0.134**	**-0.426**
Tension	**0.160**	**-0.482**	**0.411**	**-0.134**		**0.198**
Sadness	0.077	-0.083	**-0.227**	**-0.426**	**0.198**	

arousal emotions such as transcendence, power, tension, anger, disgust or fear, and negatively correlated with emotions with low arousal such as peacefulness or tenderness. In addition, rhythm descriptors (related to tempo in bpm) positively correlate with perceived tempo and with high arousal emotions such as power, tension, anger, and fear, and negatively correlate with high arousal ones such as peacefulness, disgust, tenderness and with perceived complexity. Moreover, the number of perceived instrument kinds is correlated with brightness, and pitch confidence is negatively correlated to complexity (in complex pieces it is difficult to estimate pitch information), but positively correlated with joyfulness.

We can summarize that research question RQ3 was tested for a wide range of descriptors and of perceptual aspects and was affirmed in both studies. However, while some high correlations were found in both studies (e.g., positive correlation between transcendence and energy in Bark bands, negative correlation between peacefulness and some MFCC coefficients), we also identified differences between the studies. As part of future work, we therefore plan to test these correlations on a larger music corpus.

5 CONCLUSIONS AND FUTURE WORK

We presented the outcomes of two user studies that investigated various aspects of the perception of classical orchestra music, in particular of Beethoven's 3rd symphony, "Eroica". Listeners were characterized by demographics, music knowledge, inclination, and experience, as well as

TABLE 14: Cross-correlations (Pearson's correlation coefficients) between aspects of music perception, for the **JKU/CP** study. Significant results ($p < 0.01$) are depicted in bold face. Only the correlations between peacefulness and complexity (-0.041) and between surprise and tenderness (0.042) are significant at $p < 0.05$, but not at $p < 0.01$. We therefore refrain from illustrating different levels of significance explicitly.

	Trans.	Peace.	Power	Joyful.	Tension	Sadness	Anger	Disgust	Fear	Surprise	Tender	Tempo	Compl.	Instr.
Transcendence		0.110	0.123	0.135	0.100	0.000	-0.016	-0.004	-0.030	0.120	0.172	-0.019	0.123	0.089
Peacefulness	0.110		-0.394	0.331	-0.347	-0.053	-0.434	-0.261	-0.315	-0.007	0.560	-0.176	-0.041	-0.009
Power	0.123	-0.394		0.088	0.367	-0.144	0.341	0.143	0.114	0.179	-0.335	0.278	0.240	0.150
Joyful activation	0.135	0.331	0.088		-0.148	-0.380	-0.304	-0.277	-0.367	0.221	0.305	0.205	0.158	0.077
Tension	0.100	-0.347	0.367	-0.148		0.178	0.386	0.253	0.341	0.141	-0.263	0.123	0.129	0.122
Sadness	0.000	-0.053	-0.144	-0.380	0.178		0.313	0.336	0.474	-0.109	0.003	-0.191	-0.101	-0.006
Anger	-0.016	-0.434	0.341	-0.304	0.386	0.313		0.562	0.539	0.085	-0.341	0.079	0.066	0.064
Disgust	-0.004	-0.261	0.143	-0.277	0.253	0.336	0.562		0.497	0.058	-0.182	0.010	-0.002	0.047
Fear	-0.030	-0.315	0.114	-0.367	0.341	0.474	0.539	0.497		0.080	-0.224	-0.022	0.020	0.029
Surprise	0.120	-0.007	0.179	0.221	0.141	-0.109	0.085	0.058	0.080		0.042	0.176	0.188	0.047
Tenderness	0.172	0.560	-0.335	0.305	-0.263	0.003	-0.341	-0.182	-0.224	0.042		-0.107	0.000	0.056
Tempo	-0.019	-0.176	0.278	0.205	0.123	-0.191	0.079	0.010	-0.022	0.176	-0.107		0.269	0.106
Complexity	0.123	-0.041	0.240	0.158	0.129	-0.101	0.066	-0.002	0.020	0.188	0.000	0.269		0.319
Instrument kinds	0.089	-0.009	0.150	0.077	0.122	-0.006	0.064	0.047	0.029	0.047	0.056	0.106	0.319	

personality traits. They rated 15 musicologically defined excerpts of the piece with respect to up to 10 emotions as well as perceived tempo, complexity, and instrumentation. Our research questions were: (RQ1) Are there differences in human music perception between people of different characteristics?; (RQ2) Do certain perceived music characteristics correlate, irrespective of listener characteristics?; (RQ3) Does human perception of music characteristics correlate with descriptors extracted from music audio signals?

Our main findings with respect to RQ1 are that (i) the aspects transcendence, power, and number of instrument groups are significantly correlated with most demographics and musical knowledge characteristics of participants and (ii) agreement on perceived emotions is substantially higher among listeners with a decent musical education and those playing an instrument; on the other hand, agreement is lower among frequent classical music listeners and listeners familiar with the piece under investigation. As for RQ2, there exist several significant cross-correlations between perceived emotions. Concerning RQ3, music audio descriptors, in particular those related to loudness, timbre, harmony, and rhythm show high correlations with perceived emotions.

In more detail, our findings indicate for RQ1 that

- in particular the perception of transcendence and power correlates significantly with basic user characteristics, most pronounced with aspects related to affinity to classical music,
- the ability to distinguish certain instrument groups depends on the intensity of music education and the amount of time spent playing an instrument,
- participants trained in classical music tend to disagree more on perceived emotions, but agree more on perceived tempo, than those not trained,
- the agreement among most perceptual aspects (majority of emotions, complexity, and instrument groups) decreases with increasing familiarity with the piece,
- people who are open to new experiences tend to have a stronger perception of positive aspects (transcendence, peacefulness, joyful activation, and tenderness), while disorganized listeners perceive stronger negative aspects (sadness, anger, and disgust), and

- calm listeners agree considerably more on almost all emotion aspects than others.

With regard to RQ2, we found that

- substantial correlations between all pairs of anger, fear, and disgust are identified,
- peacefulness is moderately negatively correlated with power and tension, but positively with tenderness,
- power is significantly correlated with tension, but also with anger,
- transcendence and surprise do not show notable correlations with any other aspects, and
- perceived complexity is moderately correlated with number of instrument groups and with tempo.

Investigating RQ3, we observed that

- music audio descriptors related to loudness, rhythm, timbre, and harmony show high correlations to emotional ratings,
- audio descriptors related to rhythm correlate substantially with perceived tempo,
- pitch confidence is negatively correlated with perceived complexity, and
- the number of perceived instrument groups is correlated with brightness.

In future work, we will extend the present study to a larger music corpus and to listeners from different cultural backgrounds, to take a cross-cultural perspective on the studied research questions. In particular, it will be interesting to investigate whether agreement in collectivist cultures is higher than in individualist cultures. Furthermore, we will perform a deeper analysis of the ratings on the level of the individual 15 excerpts. We also plan to investigate a wider repertoire of music and assess whether results generalize to other music material. However, acquiring the necessary comprehensive user information and ratings is costly and time-consuming, and therefore left for future work.

ACKNOWLEDGMENTS

TABLE 15: Sign of Pearson's correlation coefficient for values > 0.7 of correlation between perceptual ratings and music descriptors.)

Study	Emotional rating	Descriptor category	Descriptor	Source	Sign
MTG	Transcendence	Loudness	Energy in bark bands and frequency bands (derivative)	Essentia	+
MTG	Peacefulness	Timbre	MFCC (coeff. 12)	MIRToolbox	-
MTG	Joyfulness	Timbre	GFCC, MFCC	Essentia	+
		Harmony	Rate of chord change, chord histogram (coeff. 21)		-
MTG	Tension	Timbre	MFCC (coeff. 12)	MIRToolbox	-
MTG	Sadness	Harmony	Chord histogram (coeff. 21)	Essentia	+
JKU	Transcendence	Loudness	Energy in bark bands and frequency bands (derivative)	Essentia	+
		Timbre	GFCC (coeff. 3) and std of derivative (coeff. 5), MFCC (coeff. 8), spectral contrast (maximum value)	Essentia	-
JKU	Peacefulness	Loudness	Energy in barkbands, ERB, RMS (statistics)	Essentia	-
			Silence rate		+
		Timbre	Spectral decrease, skewness, GFCC, MFCC, spectral contrast, spectral valleys.		+
			Spectral Complexity		-
		Rhythm	Beat loudness	Essentia	-
		Harmony	Chord histogram (coeff. 2)	Essentia	+
JKU	Power	Loudness	RMS (mean and std)	MIRToolbox	+
			Silence rate, energy (barkbands, ERB), RMS and statistics	Essentia	+
		Timbre	Dissonance (mean), HFC, spectral complexity, decrease, contrast, strong peak, GFCC, MFCC		+
					+
			Dissonance (var), spectral crest, skewness, MFCC		-
		Rhythm	Beat loudness		+
		Harmony	Pitch, pitch salience and confidence		-
JKU	Joyfulness	Timbre	GFCC, MFCC (statistics)		+
		Harmony	Pitch confidence and HPCP derivates		+
JKU	Tension	Loudness	Mean RMS	MIRToolbox	+
			MFCC (coeff.1)		-
			RMS (mean, max, derivative), energy in bands	Essentia	+
			Silence rate		-
		Timbre	Dissonance, HFC, spectral flux, kurtosis,		+
		Rhythm	Beats loudness (mean, min, ratio, ratio derivatives)		+
		Harmony	HPCP (chroma) variance (coeff. 17), chords histogram (coeff. 2)		-
JKU	Sadness	Timbre	GFCC, MFCC, spectral contrast, temporal kurtosis, tristimulus		-
JKU	Anger	Loudness	Mean RMS	MIRToolbox	-
			MFCC (coeff.1)		+
			RMS (mean, max), Energy in bands, ERB bands	Essentia	+
			Silence rate		-
		Timbre	Spectral flux, complexity, HFC, MFCC		+
			Dissonance (derivate), spectral skewness, GFCC		-
		Rhythm	Beats histogram		+
		Harmony	HPCP (chroma) max, chords histogram		-
JKU	Disgust	Loudness	Energy in bark bands, frequency bands, ERB	Essentia	+
		Timbre	Dissonance derivate, GFCC		-
			Spectral flux min		+
		Rhythm	Beat loudness min		-
		Harmony	HPCP max		-
JKU	Fear	Loudness	Energy in bark bands, frequency bands	Essentia	+
		Timbre	Dissonance derivate, GFCC derivate, MFCC derivate		-
		Rhythm	Beats loudness mean		+
		Harmony	HPCP var, derivate		-
JKU	Surprise	Timbre	GFCC (min, var), MFCC (Var), energy ratio of odd/even harmonics	Essentia	-
			MFCC (variance of derivate), spectral contrast		+
JKU	Tenderness	Loudness	RMS (mean and std)	MIRToolbox	-
			Silence rate	Essentia	+
			Energy (derivate and statistics), in frequency bands, bark bands, ERB.		-
		Timbre	Dissonance derivate, spectral decrease, kurtosis, GFCC, MFCC		+
			Spectral complexity (statistics), flux, GFCC, MFCC, HFC, spectral contrast		-
		Rhythm	Beat loudness		-
		Harmony	Tonal chord histogram (coeff. 2)		+
JKU	Tempo	Timbre	Spectral entropy, rolloff (derivate), zero crossin rate (derivate)	Essentia	+
			MFCC (derivate), spectral contrast		+
			Spectral flatness, GFCC (Var)		-
		Rhythm	Beat loudness band ratio		+
		Harmony	Pitch confidence derivate		+
JKU	Complexity	Loudness	Silence rate	Essentia	-
		Timbre	Bark band skewness, kurtosis, spectral crest, entropy, flatness, spread		-
			Spectral complexity derivate, MFCC, spectral contrast		+
		Rhythm	Beat loudness band ratio		-
		Harmony	Pitch confidence and salience, HPCP (chroma) derivate		-
			HPCP (chroma) max		+
JKU	Instrument kinds	Timbre	Brightness	MIRToolbox	+

APPENDIX
SCREENSHOTS OF THE WEB INTERFACE

In the following, screenshots of the questionnaires used to assess perceptive qualities of music are provided. Figure 2 shows the one used in the UPF/MTG study, while Figure 3 shows the one used in the JKU/CP study. The extended JKU/CP questionnaire provides in question blocks 1 to 4 explanations of the emotions, perceived tempo, perceived complexity, and kinds of instruments, which are shown to participants on mouse hovering. These are detailed in Table 16. In addition, question 4 lists all instruments playing in the piece and provides links to their respective Wikipedia pages.

APPENDIX
MUSICAL DESCRIPTION OF THE EXCERPTS

Table 17 provides a summary of the musical characteristics of the 15 segments investigated in the user studies.

Fig. 2: Music tag questionnaire of the UPF/MTG survey.

Fig. 3: Music tag questionnaire of the JKU/CP survey.

REFERENCES

[1] K. Hevner, "Expression in Music: A Discussion of Experimental Studies and Theories," *Psychological Review*, vol. 42, March 1935.

[2] A. Pike, "A phenomenological analysis of emotional experience in music," *Journal of Research in Music Education*, vol. 20, pp. 262–267, 1972.

[3] J. Sloboda and P. Juslin, *Music and Emotion: Theory and Research*. Oxford University Press, 2001.

[4] A. J. Lonsdale and A. C. North, "Why do we listen to music? A uses and gratifications analysis." *British Journal of Psychology*, vol. 102, no. 1, pp. 108–134, February 2011.

[5] C. Laurier, "Automatic classification of music mood by content-based analysis," Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona, Spain, 2011.

[6] P. Ekman, *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Pergamon, 1972.

[7] P. Juslin and P. Laukka, "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *Journal of New Music Research*, vol. 33, no. 2, pp. 217–238, 2004.

[8] P. Juslin and J. Sloboda, *Psychological perspectives on music and emotion*. Oxford University Press, 2001, pp. 71–104.

[9] E. Schubert, "Update of the hevner adjective checklist," *Perceptual and Motor Skills*, vol. 96, pp. 1117–1122, 2003.

[10] M. Zentner, D. Grandjean, and K. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement," *Emotion*, vol. 8, no. 4, pp. 494–521, August 2008.

[11] J. Russel, "A circumspect model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

TABLE 16: Explanation of emotions, tempo, complexity, and kinds of instruments shown to participants of the study when hovering the mouse over the respective text.

Aspect	Description
Transcendence	I perceive the music as fascinating, overwhelming, inspiring, spiritual.
Peacefulness	I perceive the music as serene, calm, soothed, meditative, relaxed.
Power	I perceive the music as strong, energetic, triumphant, fiery, heroic.
Joyful activation	I perceive the music as animating, bouncy, joyful, dancing, amusing, stimulating.
Tension	I perceive the music as tense, agitating, irritating, nervous, impatient.
Sadness	I perceive the music as sad, tearful, sorrowful.
Anger	I perceive the music as angry, annoying, frustrating.
Disgust	I perceive the music as disgusting, repulsive, loathing.
Fear	I perceive the music as alarming, making me anxious or afraid.
Surprise	I perceive the music as astonishing, exciting, surprising.
Tenderness	I perceive the music as tender, serene, or sympathetic.
Perceived tempo	What do you think about the tempo of the audio excerpt? Does it have slow tempo or fast tempo?
Perceived complexity	Select the level of complexity of the excerpts.
Kinds of instruments	Select the number of kinds of instruments that are playing.

TABLE 17: Musical description of the excerpts. Clip: location of the excerpt in the score (movement and bars) and duration. Composition: place/role of the excerpt within the piece, and style/writing features. Orchestration: instrumental/timbral aspects. Rhythm: meter, tempo and other rhythmic features. Dynamics: energy-related character and expressive resources. Tonal: mode of the key.

ID	Clip	Composition	Orchestration	Rhythm	Dynamics	Tonal
1	Mov I 83-94 16"	Exposition, second subject More harmony than melody	Phrase 1: woods, strings Bridge: strings, woods Phrase 2: woods, horns	Mostly homophonic beatwise Bridge adds shorter figures	Piano crescendo Sforzando apoggiaturas Closing: diminuendo	Major
2	Mov I 189-197 10"	Development 3 motives combined	Motive's head: lower strings Galloping motive: strings Fanfare calls: woods, brass	Motive's head: long figures Galloping: asymmetric patterns Fanfare: onbeat, long figures	Fortissimo	Minor
3	Mov I 287-302 21"	Development, "new theme" 3-part homophonic, melodic Ascending-descending lines	Phrase 1: oboes, strings Phrase 2: violins, bassoons	Concatenated varied rhythmic patterns	Piano Mild sforzando accents	Minor
4	Mov I 685-695 14"	Coda Movement's closing cadence	Tutti Orchestral accelerando	Begin: syncopated accents Rhythmic stabilization End: homophonic chords	Forte to fortissimo	Major
5	Mov II 96-101 19"	Closing section Slow march Dominant-tonic alternation	Tutti (incl. timpani)	Mostly homophonic (winds) Figured tremolo (strings) End: forte-piano (subito)	Forte and fortissimo Sforzando anacruses	Major
6	Mov II 146-150 16"	Climax of fugal section Ascending-descending Suspensions, broken cadence	Woods, strings Figured tremolo: strings	Homophonic melodic lines Suspended values Slow figured tremolo	Forte to fortissimo Subito silence after broken cadence	Minor
7	Mov II 159-166 27"	Codetta of fugal section Slow fanfare, harmonic progression, suspensions	Fanfare: brass, woods Figured tremolo: strings	Fanfare: long notes 16th triplets (strings)	Fortissimo	Minor
8	Mov II 238-242 23"	Coda Fragmented main theme Unexpected interruptions	Main theme: 1st violins Pizzicato: low strings Oboes, horns	Fragmented, modified main theme's rhythm	Pianissimo (sottovoce)	Minor
9	Mov III 39-69 16"	Main theme in the dominant	Theme: flutes, 1st violins Dialogued accents: strings and woods	Mostly to the beat (3/4) Dialogued downbeat accents	Pianissimo	Major
10	Mov IV 1-11 12"	Introduction Fast opening, followed defining chords	Fast opening: strings Answer: tutti chords	Mostly homophonic Opening: 16th-note stream Chords mostly to the beat	Fortissimo Sforzando chords	Major
11	Mov IV 229-256 26"	Variation in march style Lively march, accompanied by varied rhythms	Main theme: 1st violins and woods	March, swinged subdivision Varied patterns, including beatwise triplets (bass)	Forte Sforzando punctuations	Minor
12	Mov IV 279-300 20"	Variation in fugue style Counterpoint interplay	Mostly strings, flutes	3 rhythms interplay Long notes, 16th-note stream and syncopations	Pianissimo to piano Sforzando accents at some syncopations	Major
13	Mov IV 367-374 23"	Variation of the theme Melody and harmonization	Melody: 1st violins, oboes Harmony: clarinets (arpeggiated)	Binary vs. 16th-note triplets Offbeat (mid strings)	Piano Crescendo towards the cadence	Major
14	Mov IV 398-403 18"	Variation of the theme Phrase 1: harmonized melody Phrase 2: descending progression	Melody: strings, flutes Harmony: clarinets (arpeggiated)	Binary vs. ternary Offbeat (high and mid strings)	Phrase 1: piano Phrase 2: crescendo	Major
15	Mov IV 416-422 20"	Chordal variation Mostly homophonic Offbeat accompaniment	Begin: almost tutti Offbeat: mid strings Harmony: 1st violins, basses	Homophonic long figures Offbeat (mid strings) Orchestral accelerando	Forte to fortissimo	Minor

[12] C. Laurier, M. Sordo, J. Serra, and P. Herrera, "Music mood representation from social tags," in *Proceedings of the 10th International*

Society for Music Information Retrieval Conference (ISMIR), Kobe, Japan, October 2009.

[13] W. Trost, T. Ethofer, M. Zentner, and P. Vuilleumier, "Mapping aesthetic musical emotions in the brain," *Cerebral Cortex*, vol. 22, no. 12, pp. 2769–2783, 2012.

[14] J. Russel, "In defense of a prototype approach to emotion concepts," *Journal of Personality and Social Psychology*, vol. 60, no. 1, pp. 37–47, 1991.

[15] T. Eerola and J. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music*, vol. 39, no. 1, pp. 18–49, 2011.

[16] P. J. Rentfrow, L. R. Goldberg, and D. J. Levitin, "The structure of musical preferences: A five-factor model." *Journal of Personality and Social Psychology*, vol. 100, no. 6, pp. 1139–1157, 2011.

[17] A. Rodà, S. Canazza, and G. D. Poli, "Clustering Affective Qualities of Classical Music: Beyond the Valence-Arousal Plane," *IEEE Transactions on Affective Computing*, vol. 5, no. 4, pp. 364–376, October–December 2014.

[18] L. Balkwill and W. Thompson, "A cross-cultural investigation of the perception of emotion in music," *Music Perception*, vol. 17, no. 1, pp. 43–64, 1999.

[19] T. Fritz, S. Jentschke, N. Gosselin, D. Sammler, I. Peretz, R. Turner, A. Friederici, and S. Koelsch, "Universal recognition of three basic emotions in music," *Current Biology*, vol. 19, no. 7, pp. 573–576, 2009.

[20] Y.-H. Yang and H. H. Chen, *Music Emotion Recognition*. CRC Press, 2011.

[21] ——, "Machine recognition of music emotion: A review," *Transactions on Intelligent Systems and Technology*, vol. 3, no. 3, May 2013.

[22] T. Eerola, O. Lartillot, and P. Toiviainen, "Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models," in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, October 2009.

[23] Y.-H. Yang and H. H. Chen, "Ranking-Based Emotion Recognition for Music Organization and Retrieval," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 762–774, May 2011.

[24] A. Huq, J. Bello, and R. Rowe, "Automated Music Emotion Recognition: A Systematic Evaluation," *Journal of New Music Research*, vol. 39, no. 3, pp. 227–244, November 2010.

[25] X. Hu, J. S. Downie, and A. F. Ehmann, "Lyric Text Mining in Music Mood Classification," in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, October 2009.

[26] C. Laurier, J. Grivolla, and P. Herrera, "Multimodal Music Mood Classification Using Audio and Lyrics," in *Proceedings of the 7th International Conference on Machine Learning and Applications (ICMLA)*, San Diego, California, USA, December 2008, pp. 688–693.

[27] E. M. Schmidt and Y. E. Kim, "Projection of acoustic features to continuous valence-arousal mood labels via regression," in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, October 2009.

[28] Y. Song, S. Dixon, and M. Pearce, "Evaluation of Musical Features for Emotion Classification," in *Proceedings of the 13rd International Society for Music Information Retrieval Conference (ISMIR)*, Porto, Portugal, October 2012.

[29] D. Liu, L. Lu, and H.-J. Zhang, "Automatic mood detection from acoustic music data," in *Proceedings of the 4th International Society for Music Information Retrieval Conference (ISMIR)*, Baltimore, Maryland, USA, October 2003.

[30] T. Li and M. Ogihara, "Detecting emotion in music," in *Proceedings of the 4th International Society for Music Information Retrieval Conference (ISMIR)*, Baltimore, Maryland, USA, October 2003.

[31] Y. Kim, E. Schmidt, R. Migneco, B. Morton, P. Richardson, J. Scott, J. Speck, and D. Turnbull, "Music emotion recognition: A state of the art review," in *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, Netherlands, August 2010.

[32] X. Hu, J. S. Downie, C. Laurier, M. Bay, and A. F. Ehmann, "The 2007 MIREX Audio Mood Classification Task: Lessons Learned," in *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, Philadelphia, USA, September 2008.

[33] X. Hu and J. S. Downie, "When Lyrics Outperform Audio for Music Mood Classification: A Feature Analysis," in *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, the Netherlands, August 2010.

[34] C. Laurier and P. Herrera, *Automatic Detection of Emotion in Music: Interaction with Emotionally Sensitive Machines.* IGI Global, 2009, ch. 2, pp. 9–32.

[35] X. Hu, Z. Chen, and D. Yang, "Lyric-Based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method," in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, October 2009.

[36] R. R. McCrae and O. P. John, "An Introduction to the Five-Factor Model and its Applications." *Journal of Personality*, vol. 60, no. 2, pp. 175–215, 1992.

[37] P. J. Rentfrow and S. D. Gosling, "The do re mi's of everyday life: The structure and personality correlates of music preferences." *Journal of Personality and Social Psychology*, vol. 84, no. 6, pp. 1236–1256, 2003.

[38] ——, "The content and validity of music-genre stereotypes among college students," *Psychology of Music*, vol. 35, no. 2, pp. 306–326, February 2007.

[39] T. Chamorro-Premuzic and A. Furnham, "Personality and music: can traits explain how people use music in everyday life?" *British Journal of Psychology*, vol. 98, pp. 175–185, May 2007.

[40] B. Ferwerda, M. Schedl, and M. Tkalcic, "Personality & Emotional States : Understanding Users Music Listening Needs," in *UMAP, Extended Proceedings*, Dublin, Ireland, June-July 2015.

[41] B. Ferwerda, E. Yang, M. Schedl, and M. Tkalčič, "Personality Traits Predict Music Taxonomy Preferences," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, Seoul, Korea, April 2015, pp. 2241–2246.

[42] M. Tkalčič, B. Ferwerda, D. Hauger, and M. Schedl, "Personality Correlates for Digital Concert Program Notes," in *UMAP, Lecture Notes On Computer Science*, Dublin, Ireland, June-July 2015, vol. 9146, pp. 364–369.

[43] O. John and S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives," in *Handbook of personality: Theory and research*, 2nd ed. New York, USA: Guilford Press, 1999, vol. 2, no. 510, pp. 102–138.

[44] L. Goldberg, J. Johnson, H. Eber, R. Hogan, M. Ashton, C. Cloninger, and H. Gough, "The international personality item pool and the future of public-domain personality measures," *Journal of Research in Personality*, vol. 40, no. 1, pp. 84–96, 2006.

[45] A. Vinciarelli and G. Mohammadi, "A Survey of Personality Computing," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 273–291, July–September 2014.

[46] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 15, pp. 5802–5805, March 2013.

[47] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our Twitter Profiles, our selves: Predicting personality with Twitter," in *Proceedings of 3rd IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing (PASSAT) and 3rd IEEE Inernational Conference on Social Computing (SocialCom)*. IEEE, October 2011, pp. 180–185.

[48] G. Farnadi, G. Sitaraman, S. Sushmita, F. Celli, M. Kosinski, D. Stillwell, S. Davalos, M.-F. Moens, and M. De Cock, "Computational Personality Recognition in Social Media," *User Modeling and User-Adapted Interaction — Special Issue on Personality in Personalized Systems*, pp. 1–34, February 2016.

[49] M. Skowron, M. Tkalčič, B. Ferwerda, and M. Schedl, "Fusing social media cues: personality prediction from twitter and instagram," in *Proceedings of the 25th International Conference Companion on World Wide Web (WWW)*, Montreal, Canada, April 2016, pp. 107–108.

[50] G. Mohammadi and A. Vinciarelli, "Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features," *Transactions on Affective Computing*, vol. 3, no. 3, pp. 273–284, July–September 2012.

[51] B. Lepri, J. Staiano, E. Shmueli, F. Pianesi, and A. Pentland, "The Role of Personality in Shaping Social Networks and Mediating Behavioral Change," *User Modeling and User-Adapted Interaction — Special Issue on Personality in Personalized Systems*, pp. 1–33, February 2016.

[52] M. Zenter, D. Grandjean, and K. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement," *Emotion*, vol. 8, p. 494, 2008.

[53] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, "Essentia: an open source library for audio analysis," *ACM SIGMM Records*, vol. 6, 2014.

[54] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the Big-Five personality domains," *Journal of Research in Personality*, vol. 37, no. 6, pp. 504–528, December 2003.

[55] P. Ekman, *Basic Emotions*. New York, NY, USA: John Wiley & Sons Ltd., 1999, pp. 45–60.

[56] R. Plutchik, "The Nature of Emotions," *American Scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[57] K. Krippendorff, *Content Analysis – An Introduction to Its Methodology*, 3rd ed. SAGE, 2013.

[58] J. Landis and G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159–174, 1977.

[59] B. Ratner, "The correlation coefficient: Its values range between +1/1, or do they?" *Journal of Targeting, Measurement and Analysis for Marketing*, vol. 17, pp. 139–142, May 2009.

Erika S. Trent is a graduate student at Boston University, Massachusetts. She received her dual Bachelor's degree in Brain and Cognitive Sciences and Music from the Massachusetts Institute of Technology in 2015. During the summer of 2014, she was funded by the MIT International Science and Technology Initiatives to conduct research on the emotional characterization of classical music with Dr. Gómez at the Music Technology Group, Barcelona, which was presented at the 9th Triennial Conference of the European Society for the Cognitive Sciences of Music. She is currently pursuing a Master's degree in mental health counseling, while contributing to NIH-funded clinical psychology research at the Center for Anxiety and Related Disorders at Boston University.

Markus Schedl is an Associate Professor at the Johannes Kepler University Linz, Austria / Department of Computational Perception. He graduated in Computer Science from the Vienna University of Technology and earned his Ph.D. from the Johannes Kepler University Linz. Markus further studied International Business Administration at the Vienna University of Economics and Business Administration as well as at the Handelshögskolan of the University of Gothenburg, Sweden, which led to a Master's degree. His main research interests include web and social media mining, information retrieval, multimedia, and music information research. He (co-)authored more than 120 refereed conference papers and journal articles, among others in ACM Multimedia, SIGIR, ECIR, IEEE Visualization; Journal of Machine Learning Research, ACM Transactions on Information Systems, Springer Information Retrieval, IEEE Multimedia. Furthermore, he is associate editor of the Springer International Journal of Multimedia Information Retrieval.

Marko Tkalčič is an Assistant Professor at the Faculty of Computer Science at the Free University of Bolzano in Italy. He received his PhD degree from the University of Ljubljana (Slovenia) in 2011. From 2013 to 2015 he was a postdoctoral researcher at the Department of Computational Perception at the Johannes Kepler University in Linz (Austria). His research interests lie in the exploitation of psychological constructs (such as emotions and personality) in user modeling and personalized systems, such as recommender systems. He has published in leading journals, such as *User Modeling and User-adapted Interaction*, *Elsevier Information Sciences*, *IEEE Transactions on Multimedia*, *Springer Multimedia Tools and Applications* and *Interacting with Computers*. Marko has been the lead organizer of the EMPIRE workshop series with editions in 2013 (UMAP), 2014 (UMAP) and 2015 (RecSys) as well as the lead editor of the upcoming edited volume *Springer Emotions and Personality in Personalized Systems*.

Hamid Eghbal-zadeh is a PhD student at Johannes Kepler University Linz, Austria. He received his Master's degree in Computer Engineering from the Shiraz State University, Iran in 2012. From 2012 to 2014, he worked as a researcher in the area of speech processing at Özyegin University in Istanbul, Turkey. In 2014, he joined the Department of Computational Perception at Johannes Kepler University. During the following years, he worked as a researcher on content-based music information retrieval (MIR). He is currently pursuing his PhD and his focus of research is on factor analysis for multiple tasks in MIR.

Emilia Gómez is an Associate Professor (Serra-Húnter and ICREA Fellow) at the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, where she leads the Music Information Research Lab at the Music Technology Group. She graduated as a Telecommunication Engineer at Universidad de Sevilla (1999) and studied piano performance at Seville Conservatoire of Music. She then received a DEA in Acoustics, Signal Processing and Computer Science applied to Music at IRCAM, Paris (2000) and a PhD in Computer Science at the UPF (2006, awarded by EPSON foundation). She has been lecturer at the Escola Superior de Música de Catalunya and visiting researcher at the Royal Institute of Technology, Stockholm (Marie Curie Fellow), McGill University, Montreal, and Queen Mary University of London. She has co-authored more than a 100 peer-reviewed publications and software libraries, and contributed to more than 15 projects, mostly funded by the European Commission. She is currently president-elect of the International Society for Music Information Retrieval.

Agustín Martorell is Assistant Professor and postdoc researcher at the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona. He holds degrees in Electronics Engineering and Musicology, and both MSc and PhD degrees in the field of computational musicology. As a practical musician he plays violin and performs regularly as a chamber choir singer. His main interests are related with music theory and analysis, combining theoretical, psychological and computational methods, with particular interest in interactive and educational tools for music understanding.