



Registered Report

Computer-based music feature analysis mirrors human perception and can be used to measure individual music preference



Kai R. Fricke^{a,*}, David M. Greenberg^{b,d}, Peter J. Rentfrow^c, Philipp Yorck Herzberg^a

^aDepartment of Personality Psychology and Psychological Assessment, Helmut-Schmidt-University/University of the German Federal Armed Forces Hamburg, Hamburg, Germany

^bThe School of Performance, Anglia Ruskin University, Cambridge, United Kingdom

^cDepartment of Psychology, School of the Biological Sciences, University of Cambridge, Cambridge, United Kingdom

^dAutism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

ARTICLE INFO

Article history:

Received 28 February 2018

Revised 6 June 2018

Accepted 8 June 2018

Keywords:

Music preference

Music features

Music taste

Music perception

Music information retrieval

ABSTRACT

This paper explores the measurement of individual music feature preference using human- and computer-rated music excerpts. In the first of two studies, we correlated human ratings of song excerpts with computer-extracted music features and found good accordance, as well as similar criterion validity with preference for musical styles (the MUSIC model, mean $r = 0.88$). In a second online study ($N = 2118$), using PCA and Procrustes analysis, we found that measured music preference showed the same established three-component structure from previous research (Arousal, Valence, Depth), regardless of whether the music pieces were rated by humans or the ESSENTIA music analysis software. Our results suggest that computer-extracted music features can be used to assess individual music preference.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Research on musical genre and feature preference has established that humans show individual music preferences (e.g. Rentfrow, Goldberg, & Levitin, 2011), which in turn are related to various personality traits, such as Sensation Seeking (Little & Zuckerman, 1986), the Big Five personality dimensions (Rentfrow & Gosling, 2003), and their facets (Zweigenhaft, 2008; Greenberg et al., 2016). Besides its social implications, such as acting as a conveyor for social affiliation (North & Hargreaves, 1999), music preference thus can also be construed as an idiosyncratic personality characteristic. The assessment of music preferences currently often relies on audio based assessment, which in turn relies on a selection of pre-rated music pieces. Software-based rating of music pieces can be used to broaden the selection of stimuli for music preference assessment. Concretely, such automated ratings could be used to measure music preference from actual, individually selected or even user-provided songs (e.g. own music collections, playlists, and alike), rather than from a pre-selected pool of music. However, research has not yet examined whether these computer-extracted features are empirically equivalent to human perception. Specifically, while we can assume that computer-extracted

features partly reflect human perception, it is unclear if the results can also be used to infer idiosyncratic music preference. Therefore, in this paper we examine how the results from computational music feature extraction relate to human perception and how they can be utilized for the measurement of individual music preference.

In research, music preference is usually assessed in one of two ways: Participants can state their preference for musical genres (e.g. Rentfrow & Gosling, 2003; Bonneville-Roussy, Rentfrow, Xu, & Potter, 2013) or musical features (e.g. Fricke & Herzberg, 2017) in self-report, or participants can express their liking of certain selected music excerpts, which are used as a proxy measure to assess their music preference (e.g. Rentfrow et al., 2011; Langmeyer, Guglhör-Rudan, & Tarnai, 2012). Much of the past research on excerpt-based music preference assessment relied on a pool of neatly compiled music excerpts that have been rated by humans on various music features (e.g. Rentfrow et al., 2011; Rentfrow et al., 2012; Greenberg et al., 2016). These excerpts were selected from professionally produced, yet commercially unreleased music pieces (Rentfrow et al., 2011). Over all, 250 song excerpts have been introduced to this line of research, and they have been rated by humans on over 40 sonic and psychological music features. Separate participants would listen to the excerpts and express their preference, which would then be used to calculate their music preference profile for the music features.

* Corresponding author.

E-mail address: fricke@hsu-hh.de (K.R. Fricke).

The measurements were then used to examine relationships with cognitive styles (Greenberg, Baron-Cohen, Stillwell, Kosinski, & Rentfrow, 2015) and personality constructs (Greenberg et al., 2016).

These music excerpts are suitable for the fine assessment of music preference in overt assessment. But if we wanted to infer music preference from indirect data sources, such as actual listening behavior, we would need access to the music features of a large body of popular music. While today it is easy to get access to a large selection of music through online streaming services, gaining human ratings of thousands of songs on multiple features is a challenging and very time-consuming task if done manually. A way to overcome this limitation is to use automatic music feature extraction from music analysis software. This software can infer low-level music features directly from audio input and thus automatically annotate thousands of songs in relatively short time.

Low-level audio features derived from analysis software can be subjected to machine learning algorithms to classify songs on more high-level features, such as moods (e.g. *Danceable*, *Aggressive*) or genres (e.g. *Rock*, *Pop*, *Jazz*). These algorithms generalize learned features from ground truth datasets (i.e. manually annotated music pieces) to new audio inputs. The ground truth data is usually annotated by humans; still, it is sensible to confirm the validity of the computed feature output on a different dataset using human ratings. Also, audio classifiers usually use relatively broad categories, such as *Happy*, or *Sad*. Comparing the machine ratings with human ratings on more detailed music features can help us understand the facets of such categories and help in the interpretation of the music analysis results and music preference assessment.

Music preference research shows a robust five-factor (for musical genres; see Rentfrow et al., 2011) and three-factor (for musical features; see Greenberg et al., 2016; Fricke & Herzberg, 2017) structure. If computer-extracted music features proved to be a valid way to assess music feature preference, we would expect to find a similar factor structure here, as well.

In our first study we use rating data based on human perception of 150 songs collected by Rentfrow et al. (2011, 2012) to validate the feature extractions of the music analysis software ESSENTIA (Bogdanov et al., 2013). In our second study, we then examine and compare the component structures of music feature preference for human-rated and computer-extracted music features. Our study aims to show that computer-extracted music features can be used for the assessment of music preference, just as human-rated music features have shown in previous research. The overall goal of this paper is hence to provide a new tool for the assessment of individual music preference, which can be applied to any body of music, including individually selected or user-provided music pieces.

1.1. Music features and music preference

Music can be described and classified using various dimensions. Human-driven classification often revolves around categories, such as genres, moods, and sound characteristics. Machine-extracted features on the other hand are typically very technical, as they usually describe quantitative characteristics of the audio signal, such as amplitudes, energies, and spectral bands. These technical attributes are referred to as *low-level* features. Classifications derived from a combination of these low-level features often mimic human concepts and are referred to *high-level* features. We will first discuss human classification and rating of music attributes, and then proceed to illustrate machine-extracted low-level and high-level features.

Traditionally, music has been categorized into genres such as *Rock*, *Pop*, or *Jazz*. Psychological research suggests that preference for these genres can be modeled on the five factors of *Mellow*,

Unpretentious, *Sophisticated*, *Intense*, and *Contemporary* (the MUSIC-Model; Rentfrow et al., 2011). In the past years, however, research switched focus from broader music *genre* preference (e.g. “I like Jazz”) to music *feature* preference (e.g. “I like relaxing music”) (Fricke & Herzberg, 2017). Preference for music features in excerpt-based assessment has been shown to load on three factors: *Arousal*, *Valence*, and *Depth* (AVD) (Greenberg et al., 2016). This factor structure has been replicated for self-reported music preference assessment (Fricke & Herzberg, 2017).

The relationship of preference for musical styles (i.e., the MUSIC model dimensions) and for music features (i.e., the AVD model) has been examined in research (e.g. Rentfrow et al., 2012; Fricke & Herzberg, 2017). Furthermore, the factor structure of the MUSIC model has been confirmed in various studies, including a large cohort study with over 250,000 participants (Bonnevill-Roussy et al., 2013). The relationship of music feature preference with the MUSIC model is hence suitable to determine criterion validity of the computer-extracted features.

Regarding the relationship to personality, music preference showed robust relationships to the Big Five personality dimensions (Fricke & Herzberg, 2017; Langmeyer et al., 2012; Zweigenhaft, 2008; Rentfrow & Gosling, 2003). A recent meta analysis confirmed some of these findings, but noted that most relationships are rather small (Schäfer & Mehlhorn, 2017). In addition to their primary relationships with music preference, personality traits have been shown to moderate age trends in music preference (Bonnevill-Roussy et al., 2013). Further, some studies showed that music preference correlated with the same biological indicators as the Big Five personality traits; for instance, a higher testosterone level correlated negatively with preference for sophisticated music in males (Doi, Basadonne, Venuti, & Shinohara, 2018). All these results suggest that music preference itself is closely related to personality. As such, the exploration of novel assessment methods for music preference can be used to enable research and real-world applications to infer personality characteristics from music preference data, and use these insights to tailor their tasks and services to each user. Additionally, personality research suggested that digitally derived behavior data, such as Facebook likes, predicted personality better than judgments of human peers, and sometimes even self-ratings (Youyou, Kosinski, & Stillwell, 2015). It is hence conceivable that computer-based methods might provide increased validity, or at least increased objectivity over human rating-based assessments.

1.2. Automatic music feature extraction

The subfield of computer sciences that deals with extracting information from music data is called *Music Information Retrieval* (MIR). MIR systems seek to analyze music files in terms of pitch, tempo, harmony, and timbre, as well as editorial, textual, and bibliographic facets (Downie, 2003). The last two decades saw various research studies developing algorithms to extract these facets from music files, as well as the creation of several software suites implementing these algorithms.

Among these software suites are open-source solutions such as *librosa* (McFee et al., 2015), *jMIR* (McKay, 2010), *ESSENTIA* (Bogdanov et al., 2013), and many others. We chose to use *ESSENTIA* for our analyses, because (a) it is actively developed and maintained, (b) it implemented state-of-the-art MIR algorithms, and (c) it provides learned models for high-level features based on a large count of research papers.

ESSENTIA analyzes the digital audio data from songs and extracts various low-level parameters, such as *beats per minute*, *spectral complexity* and *MFCC* (Bogdanov et al., 2013). These low-level parameters can be administered to machine learning algorithms, which map certain low-level audio profiles to high-level features. *ESSENTIA* provides pre-trained Support Vector Machines

Download English Version:

<https://daneshyari.com/en/article/7326306>

Download Persian Version:

<https://daneshyari.com/article/7326306>

[Daneshyari.com](https://daneshyari.com)