Emotion–Based Music Retrieval

A

C.H.C. Leung
Department of Computer Science, Hong Kong Baptist University, Hong Kong

J. Deng
Department of Computer Science, Hong Kong Baptist University, Hong Kong

INTRODUCTION

With the astounding growth of digital music, music retrieval is extremely important for discovering music that matches listeners taste or preferences. Music is a complex acoustic and physical product, which encompasses mind, feeling, emotion, culture and other aspects of human beings. Therefore, music plays a prominent role in people's daily lives, not only in relieving stress, but also in cultivating sentiment.

Currently there are numerous digital music services on the Internet. Many music service websites (e.g. Yahoo Music, MySpace) provide music retrieval by meta-data information such as music title, genre, album, lyrics and biography, which are not able to analyze music content and retrieve music by content. However, a few online music providers (e.g. Pandora.com, Musipedia) attempt to retrieve music relying on melody, rhythm, timbre, or harmony, which greatly improves music retrieval results.

In many research studies, many people believe that music can induce emotion, and some psychologists have done experiments using physiological measurements such as heart rate and skin conductance to prove this view (Zentner et al., 2008; Scherer, 2005). In this article, we view music as an art form and soul of language which can engender a feeling or evoke emotions. Naturally, emotional expression in music is the key factor to analyze music emotional content. However, most current music services ignore the emotion and sentiment influence or simply utilize tags to represent some general emotions conveyed in music. Considering that emotions induced by music are significant for deeper analyses of music, this article introduces a method of emotion-based music retrieval, which provides a more natural and humanized way to better experience music.

The aim of emotion-based music retrieval systems is to efficiently retrieve music from a music database by emotional similarity. Therefore, the first task is to define the expression of emotion induced by music. Currently there are many views on emotion models. For example, some researchers view that emotion should be expressed by discrete basic human emotions such as joy, sadness, anger or grief, while others believe that emotion should be depicted in a psychological dimensional space, though there are no consensus on how many dimensions there are. In this article, we review different emotion models and propose to represent emotion by combining discrete emotion model and dimensional emotion model. The second task is to find the relationship between acoustic features and their emotional impacts. We describe music attributes such as pitch, timbre, rhythm, melody, harmony, and then point out their emotional impacts on our applied emotion model. The final task is to retrieve music based on their emotions. We suggest three query methods: query-by-music, query-by-tag, and hybrid. In addition, we also apply some ranking algorithms to return an optimal retrieval list.

The rest of this article is organized as follows: the next section will review some significant emotion models and approaches of emotion-based music retrieval. Then we shall define a hybrid music emotion model combining discrete and dimensional representations. And then the relationship between acoustic features and their emotional impact based on the utilized emotion model will also be described. After that a unified

DOI: 10.4018/978-1-4666-5888-2.ch015

framework for music retrieval by three query methods is presented. Furthermore, an effective ranking algorithm applied to emotion-based music retrieval system is proposed. Finally, some future potential directions and trends for future research are pointed out.

BACKGROUND

Emotion Models

As emotion is a complex psychological and physiological human subjective experience, many researchers have explored many different emotion models in their specific domains. Commonly, these emotion models are rooted in two widely established emotion theories: discrete emotion theory and dimensional emotion theory (Scherer, 2004). Discrete emotion theory utilizes a number of emotional descriptors or adjectives to express basic human emotions (e.g. joy, sadness, anger, contempt, happiness, etc.). Hevner investigated the relation between music and listeners' emotion and developed an adjective circle consisting of eight clusters totalling 67 emotional terms (Hevner, 1936). Ortony, Clore, and Collins proposed an emotion cognitive model named OCC model (Ortonly et al. 1988) to hierarchically describe 22 emotion type specifications. However, for the sake of simplicity, many researchers in the audio music area only use general descriptions to represent emotions induced by music, such as (soft, neutral, aggression) or (happy, neutral, sad) (Pohle et al., 2005).

In addition, the research community of music information retrieval evaluation exchange (MIREX) has classified music into five categories by clustering different emotion labels (Hu & Downie, 2007). However, there still exist some issues in the discrete emotion theory. Firstly, the small number of primary emotions is not able to adaptively describe the numerous emotional effects of music. Secondly, the distinction between emotional effects in music is not always apparent, so sometimes it is difficult for listeners to accurately express their emotions. Thirdly, the emotions induced by music are unlike the discrete psychological response in reality.

Dimensional emotion theory believes that emotion should be depicted in a psychological dimensional space. It helps to represent a wide range of emotions not necessarily depicted by a particular emotion descriptor. However, currently there is no consensus on how many dimensions of emotion should there be. Russell proposed to represent emotion by linear combinations of two independent dimensions: arousal and valence (Russell, 1980). Arousal represents the level of activation in stimuli, with range from sleepy to aroused. Valence accounts for pleasantness-unpleasantness. Furthermore, Thayer rotated the arousal-valence axes 45 degrees to produce two separate arousal dimensions: energetic arousal and tense arousal (Thayer, 1989). However, some researchers argued that energetic arousal and tense arousal are mixtures of a single activation dimension and valence (Yik et al., 1999), while the opposite view pointed out that they are not mixtures of valence and activation (Schimmack & Reisenzein, 2002). In addition, Fontaines et al. indicate that emotion is not two-dimensional through their experimental stimuli by three different languages (Fontaine et al., 2007).

The pleasure-arousal-dominance (PAD) model (Mehrabian, 1980) applies three numerical dimensions to represent emotions. The dimension of dominance represents the controlling and dominant nature of the emotion. As anger is a dominant emotion, while fear is a submissive emotion, compared with arousal-valence model, the PAD model is able to distinguish particular emotions like anger and fear. Gebhard proposed a layered model which mapped emotions represented by the OCC model into PAD space (Gebhard, 2005). Some researchers applied PAD model to measure the emotions induced by music (MacDorman & Ho, 2007). From their results, the arousal and valence show high validity for music-induced emotions, while the dominance seems to have problems that many pieces of music are neither dominant nor submissive. Fortunately, Bigand et al. (2005) proposed a three-dimensional model to represent emotions invoked by music, while they suggested that the third dimension seemed to have an emotional character measured by continuity-discontinuity or melodic-harmonic contrast.

Current Approaches to Emotion-Based Music Retrieval

There are many music annotation and retrieval works in the literature (Turnbull et al., 2008; Fu et al., 2011; Weston et al., 2011), which are relevant to emotion-based music retrieval. Turnbull et al. utilized a hier-

Related Content