The Mood of Chinese Pop Music: Representation and Recognition

Xiao Hu
Faculty of Education, University of Hong Kong, Hong Kong, China. E-mail: xiaoxhu@hku.hk

Yi-Hsuan Yang
Research Center for IT Innovation (CITI), Academia Sinica, Taipei, Taiwan. E-mail: yang@citi.sinica.edu.tw

Music mood recognition (MMR) has attracted much attention in music information retrieval research, yet there are few MMR studies that focus on non-Western music. In addition, little has been done on connecting the 2 most adopted music mood representation models: categorical and dimensional. To bridge these gaps, we constructed a new data set consisting of 818 Chinese Pop (C-Pop) songs, 3 complete sets of mood annotations in both representations, as well as audio features corresponding to 5 distinct categories of musical characteristics. The mood space of C-Pop songs was analyzed and compared to that of Western Pop songs. We also explored the relationship between categorical and dimensional annotations and the results revealed that one set of annotations could be reliably predicted by the other. Classification and regression experiments were conducted on the data set, providing benchmarks for future research on MMR of non-Western music. Based on these analyses, we reflect and discuss the implications of the findings to MMR research.

Introduction

Music mood, as an important metadata type, has not only attracted attention in the research field of music information retrieval (MIR) but has also been utilized in many music websites and services such as AllMusicGuide[1] and Spotify. However, it has been recognized that the field is dominated by studies on Western music. Researchers thus have started investigating cross-cultural issues on music mood such as mood descriptors applied to music in different regions (Lartillot & Toiviainen, 2007), mood perceptions of non-Western listeners (Lee & Hu, 2012), and generalizability of mood recognition models (Hu & Yang, 2016; Yang & Hu, 2012). Meanwhile, the subjective nature of music mood increases the difficulty in building data sets for research tasks related to mood recognition (Hu, Downie, Laurier, & Ehmann, 2008; Hu, 2010; Trohidis, Tsoumakas, Kalliris, & Vlahavas, 2008). In response to the fast growth in research in these areas, more shareable data sets with non-Western music and/or those annotated by non-Western listeners are much needed.

In representing music mood, there are two major types of models: categorical and dimensional ones. The former uses a set of natural language terms to represent different moods, such as "happy" or "angry"; the latter represents music moods with continuous values in a low-dimensional space (Kim et al., 2010; Yang & Chen, 2012). Both models have their own advantages but few studies have explored the relationships between them (Wang, Yang, Wang, & Jeng, 2012), perhaps due to the lack of data sets annotated with both models.

To bridge the gaps, we built a new data set of Chinese Pop (C-Pop) music, named CH818, for the purpose of investigating the mood of C-Pop. C-Pop broadly refers to popular music made by artists from the Greater China region and/or sung in a Chinese language (Liu & Mason, 2010). C-Pop songs are influenced by both contemporary Western Pop music and Chinese oriental music traditions. They are also found to be the most popular type of music in one of the largest digital music markets in the world (The International Federation of the Phonographic Industry, 2014). The CH818 data set consists of 818 C-Pop songs, each with three complete sets of annotations from three Chinese music experts in both categorical and dimensional representations, as well as five types of features (loudness, rhythm, pitch, timbre, and harmony) extracted from the music audio. To date, CH818 is the largest data set of C-Pop songs with comprehensive mood annotations and audio features. All metadata,

Received March 14, 2016; revised July 29, 2016; accepted September 6, 2016

©2017 ASIS&T • Published online 13 June 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23813

[1] http://www.allmusic.com/

annotations, and extracted audio features will be publicly available for research purposes.[2]

With CH818, we investigated a series of research questions on the representation and automated recognition of C-Pop music:

> RQ1: What is the mood space of C-Pop music? How is it compared to the mood space of Western Pop music?
> RQ2: What is the relationship between the two main representation models (i.e., categorical and dimensional models) on C-Pop?
> RQ3: How well can mood recognition techniques designed for Western music be applied to C-Pop?

The first two questions are on mood representation of C-Pop, whereas the last one is on automated mood recognition. The answers to RQ1 can enhance our understanding of the mood space of C-Pop and how it is similar or different from that of Western Pop music. RQ2 is to bridge the two major kinds of mood representation models (see Related Work) that have been widely used in both MIR and music psychology. Despite the popularity of the two models, the relationship between them has rarely been studied empirically. As the field has been dominated by studies on Western music, little is known about the applicability of mood recognition techniques on non-Western music. RQ3 is to fill the gap. By answering these questions, this study aims to make contributions to MIR research, particularly on mood representation and recognition in the cross-cultural context. In addition, this study also constructs a substantial data set and demonstrates detailed analyses on it. Openly accessible data sets are especially desirable, as they could be used for benchmarking across different labs and researchers. However, it is challenging to build shareable data sets, especially in MIR, partially due to intellectual property laws (Chen, Wang, Yang, & Chen, 2014; Hu, Lee, Choi, & Downie, 2014). With the analysis and experiments, hopefully this study will inspire further research on mood representation and recognition of music in different cultures.

Related Work

Music Mood Representation

Both categorical and dimensional models have been used widely in music mood recognition. A categorical model uses a set of finite discrete terms (usually adjectives) to represent moods in music. Hevner's "adjective circle" was an early influential categorical model for music mood (Hevner, 1936), where eight groups of adjectives collectively defined eight main categories of moods in (classical) music, such as "vigorous," "merry," and "dreamy." In categorical models, a song is labeled by one or more mood categories, without indications on the extent to which the song is associated with these moods. In MIR research, a widely used categorical model is the five-cluster model used in the audio mood classification (AMC) task in Music information Retrieval Evaluation eXchange (MIREX), a community-wide campaign for MIR evaluation (Hu et al., 2008). The model contains five mood clusters derived from the editorial mood labels on a widely used online music repository, AllMuscGuide (Hu & Downie, 2007). Reprinted in Table 1, the five clusters consist of 29 mood-related terms, representing five different categories of moods in (mostly Pop) music. This model has been used in various research tasks in MIR, including automated mood classification (Bischoff, Firan, Nejdl, & Paiu, 2009; Hu et al., 2008; Laurier, Grivolla, & Herrera, 2008), user mood perceptions (Hu & Lee, 2012; Hu & Lee, 2016; Lee & Hu, 2014), cross-cultural mood annotations (Hu et al., 2014; Singhi & Brown, 2014), and crowdsourcing for music mood recognition (Lee & Hu, 2012). The categorical model in the current study is developed from this five-cluster model by extending it with new terms suitable for the characteristics of Chinese songs.

TABLE 1. Five mood clusters used in MIREX (Hu et al., 2008).

Cluster	Mood labels
MIREX_C_1	Passionate, rousing, confident, boisterous, rowdy
MIREX_C_2	Rollicking, cheerful, fun, sweet, amiable/good natured
MIREX_C_3	Literate, poignant, wistful, bittersweet, autumnal, brooding
MIREX_C_4	Humorous, silly, campy, quirky, whimsical, witty, wry
MIREX_C_5	Aggressive, fiery, tense/anxious, intense, volatile, visceral

In contrast, dimensional models represent music moods with continuous values in a low-dimensional space. The most adopted dimensional model in MIR is Russell's model (Barthet, Fazekas, & Sandler, 2013; Kim et al., 2010; Yang & Chen, 2012), which consists of two dimensions, valence (measuring the level of pleasure) and arousal (measuring the level of activity) (Russell, 1980). This study also adopts this model, as reprinted in Figure 1.

Categorical and dimensional models have their complementary advantages. Categorical models are recognized as easy for users to understand since the terms are from natural languages. On the other hand, dimensional models are advantageous for visualizing mood distributions of music and representing various intensity levels of moods. There is a recent trend in studies on music and emotion that strives to integrate both representations (Eerola & Vuoskoski, 2013). Data sets annotated with both models are precious in providing empirical evidence and support for theoretical advance on integrated representation frameworks. In MIR, it is also desirable to combine their advantages to enhance mood-based MIR systems. Multiple studies on music mood attempted to visualize the proximity of various mood categories in two-dimensional spaces (e.g., Laurier, Sordo, Serra, & Herrera, 2009; Yang & Hu, 2012) using dimension reduction techniques. However, the resultant dimensions may have little semantic meaning, marking it difficult for users to understand. It was then proposed to integrate the two representations by visualizing mood categories in a category model in the dimensional space of a dimensional model

[2] The data set is accessible at http://ccmir.cite.hku.hk/data/CH818.zip

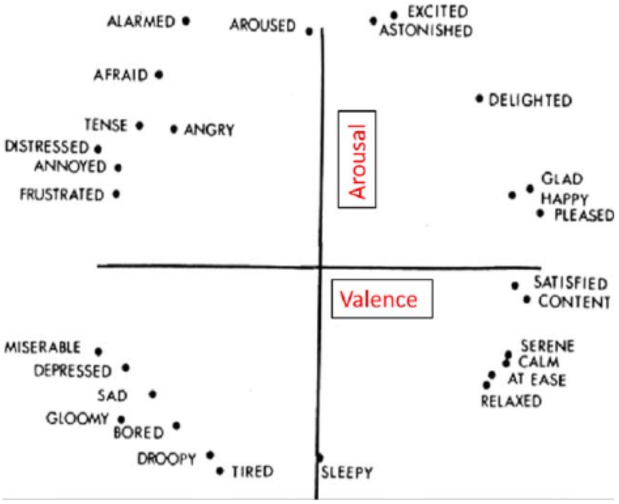


FIG. 1. Russell's valence-arousal model (Russell, 1980, p. 1168, annotation added). [Color figure can be viewed at wileyonlinelibrary.com]

(Wang et al., 2012). In this study, we also explore the mapping between the two representation models using CH818.

Studies on music mood recognition (MMR) revealed that, for dimensional models, music valence was consistently more difficult to predict than arousal (Barthet, Fazekas, & Sandler, 2013; Yang & Chen, 2012). For categorical models, it is also found that performances vary across mood categories and the differences might be related to the valence and arousal values corresponding to the categories (Hu, Choi, & Downie, 2016). Towards gaining more understanding on automated mood prediction, we conducted experiments with both representations: *classification* experiments for categorical annotations and *regression* experiments for dimensional annotations. CH818 also provides a good testbed for further studies on this line of research.

Mood Recognition on Non-Western Music

MMR studies propose algorithms and systems to automatically classify or predict music mood from musical materials, including audio, lyrics, and/or social tags (Hu et al., 2016; Kim et al., 2010; Yang & Chen, 2012). Similar to other tasks in MIR, there is a predominance of Western music and Western listeners in music mood studies. Most experiments were conducted on data sets consisting of Western music and annotated by Western listeners. As music-seeking and consumption have transcended the cultural boundary and become increasingly global, researchers have realized the importance of studying non-Western music and users as well as the possibility of generalizing research findings cross-culturally (Hu et al., 2014; Serra, 2011; Lee et al., 2013).

A few recent studies have begun evaluating cross-cultural generalizability of computational models. For instance, Yang and Hu (2012) evaluated to what extent classification models trained with Western songs can be applied to a set of Chinese songs. The last round of MIREX also started a new task on cross-cultural mood classification on a data set of K-Pop songs (Hu et al., 2014). In the front of mood regression, there are even fewer studies involving non-Western music, partially due to the lack of data sets with annotations in dimensional models. Yang, Lin, Su, and Chen (2008) conducted mood regression experiments on a data set of 195 Western, Chinese, and Japanese songs. Later, they evaluated a new approach on a set of 1,240 C-Pop songs (Yang & Chen, 2011b). To the best of our knowledge, this is the largest non-Western music data set for MMR but it only contains annotations on valence dimension. A recent study (Hu & Yang, 2016) investigated the cross-cultural and cross-data set generalizability of mood regression models, using the dimensional annotation of CH818 as one of the data sets. Notwithstanding the contributions of these previous studies, there is a sore lack of MMR studies based on *both* categorical *and* dimensional models. The next subsection will summarize existing available data sets for MMR, further illustrating the research gaps.

Existing Data Sets in MMR

Through the development of MMR, a few data sets have been made available for the research community. Table 2 summarizes key characteristics of them.

TABLE 2. Community accessible data sets in MMR studies.

Data set	Source	No. of songs	Music	Annotation
MIREX AMC data set	(Hu et al., 2008)	600	Western	Categorical: five mood clusters (Table 1)
CAL500	(Turnbull, Barrington, Torres, & Lanckriet, 2008)	500	Western	Categorical: 18 mood-related social tags
MIREX ATC[a] data set	(Hu, Downie, & Ehmann, 2009)	3,469	Western	Categorical: 18 groups of mood-related social tag
MER60	(Yang & Chen, 2011a)	60	Western	Dimensional: valence, arousal
MoodSwing	(Speck, Schmidt, Morton, & Kim, 2011)	240	Western	Dimensional: valence, arousal
NTUMIR-1240	(Yang & Chen, 2011b)	1,240	C-Pop	Dimensional: valence
AMG1608	(Chen et al., 2014)	1,608	Western	Dimensional: valence, arousal Categorical: mood labels on allmusic.com
MIREX K-POP data set	(Hu et al., 2014)	1,892	Korean	Categorical: five mood cluster (Table 1) 18 mood tag groups[b] Dimensional: valence, arousal

Note. [a]ATC stands for Audio Tag Classification task in MIREX.

[b]The same tag groups as those in the MIREX ATC data set.

All the aforementioned data sets are available to MIR researchers either through direct downloading of audio features extracted from the music files or by participation in MIREX which runs in an "algorithm to data" paradigm (Downie et al., 2014). All but the NTUMIR-1240 and MIREX K-POP data sets consist of Western music. In addition, only the last two data sets contain both categorical and dimensional annotations. However, in both data sets annotations on different music pieces may be provided by different annotators. This may limit their utility in some research topics such as personalized recommendation or classification. Compared to existing data sets, the CH818 data sets distinguishes itself in the following aspects: i) it consists of a significant amount of C-Pop music; ii) it is annotated with both categorical and dimensional models; iii) it has three sets of annotations given by the same three annotators throughout the data set; iv) it comes with a comprehensive set of audio features extracting with well-known MIR tools; and v) it has a maintenance plan in place for long-term access.

The CH818 Data Set

Data Collection

The CH818 data set contains 818 C-pop songs released in Taiwan, Hong Kong, and Mainland China from 1987 to 2010. Attempting to reflect the current trend of C-Pop, we collected C-Pop albums released in recent years and listed in hit boards (and thus were influential). To diversify songs in the data set, we randomly selected one song from each of the 818 albums collected. As the mood of a song can change during its course, and the topic of mood variation across time is beyond the scope of this study, a 30-second excerpt from each song was extracted, for the purpose of mitigating the effect of changing moods in a single song (Hu et al., 2014) as well as alleviating the cognitive load of annotators. For the purposes of MMR, the segment with the strongest emotion from each song was automatically detected and extracted (using the sum of square of predicted valence and arousal values). The detection algorithm was based on a regression model built on an external data set of Western Pop music, as there were no existing data sets of Chinese songs with proper valence and arousal annotations that could be used to train the regression model.

The annotations were collected from three music experts who were masters students or senior undergraduates in a music school at the time of annotation. All were born and raised in Mainland China and thus are regarded as having a Chinese cultural background. After a training session, each of them annotated all 818 songs independently, in both categorical and dimensional representation models.

For categorical annotation, a set of 36 mood labels were used. Out of these, 29 labels were adapted from the five mood clusters in the MIREX AMC data set (Table 1). Seven additional labels deemed to be representative in C-Pop were also adopted (Yang & Hu, 2012), including "tender," "soothing," "calm/peaceful," "relaxed," "dreamy," "nostalgic," and "encouraging" (Chow & de Kloet, 2010). Each song could be annotated with one or more mood labels, to reflect the realistic situations where each individual song could express multiple moods (Lee & Hu, 2014; Yang & Chen, 2012). All labels were translated into Chinese, with the English originals presented alongside, to avoid possible misunderstanding introduced in translation.

For dimensional annotations, we followed the literature and adopted Russell's model of valence and arousal dimensions. The values ranged from −10 to 10 and were annotated using two separate slide-bars. The instructions explained what valence and arousal meant and additional notes were presented at both ends and the middle (neutral) point of the slide-bars. Figure 2 illustrates the annotation interface of one clip. On average each annotator spent 25.67 hours and was paid about 15 U.S. dollars per hour.

Song Characteristics

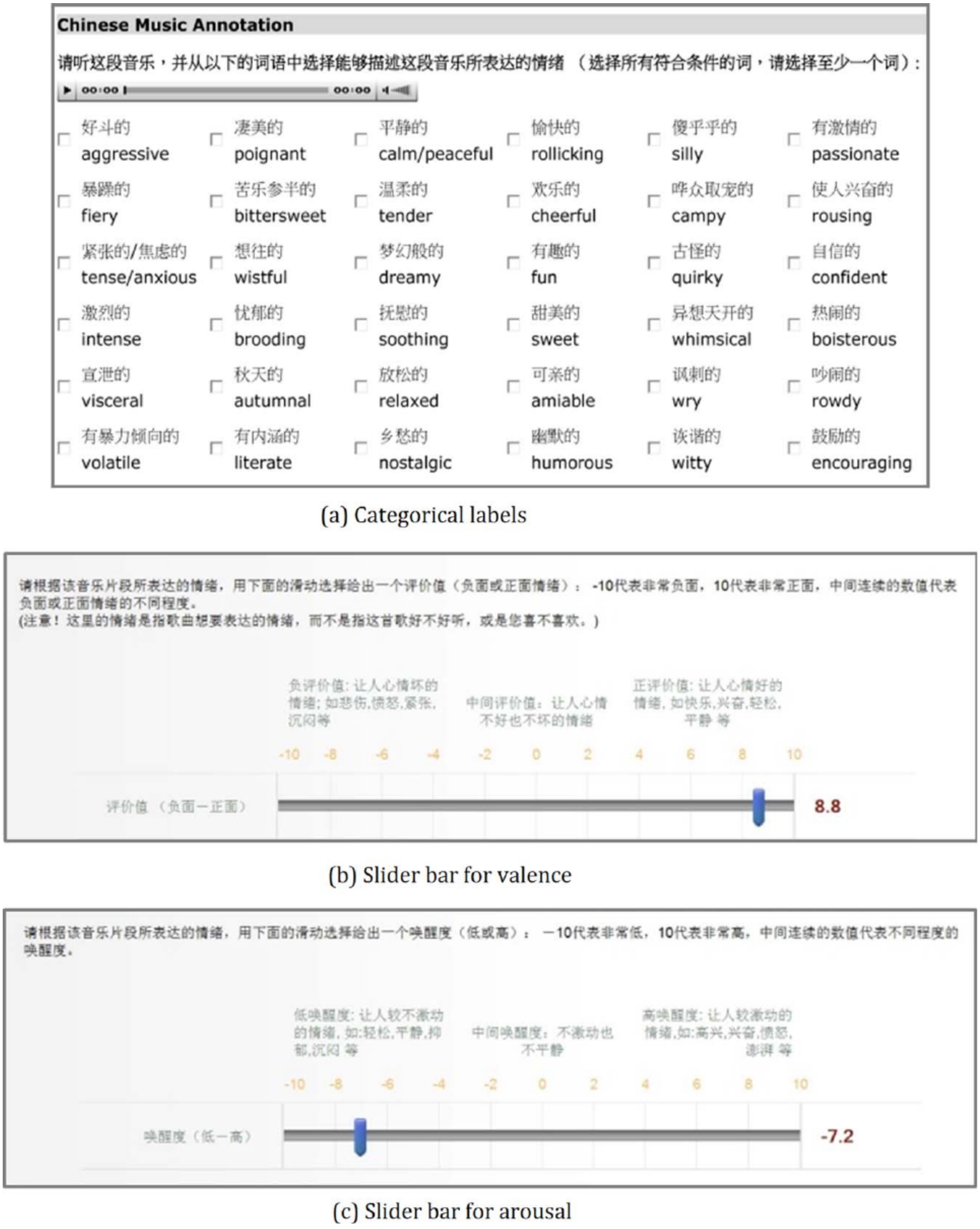
All but 12 songs in CH818 were released in the first decade of this century, with around 150 songs each year from 2005 to 2009. There are ~390 songs with male and female voices, as well as 39 songs with mixed voices (produced by bands or duets). Most of the songs were in Mandarin, while 17 were in Minnanese (Hokkien) and 15 were in Cantonese. It is acknowledged that, despite our best efforts in collecting and selecting the songs, the data set may not be comprehensive enough to represent all C-Pop music. Nevertheless, the recency and popularity of songs makes it highly relevant for research and practical purposes.

Categorical Annotations and Reliability

On average, each song received 10.20 labels, with a standard deviation of 2.67. The most popular mood labels applied were "tender" (1,027 times), "dreamy" (658), and "relaxed" (647), followed by "rousing" (645) and "passionate" (629). The least applied labels include "volatile" (18), "autumnal" (59), "nostalgic" (60), "fiery" (67), and "aggressive" (76).

To quantify the interrater reliability across annotators, we calculated the agreement ratio between each pair of annotators. In CH818, each song could be annotated with multiple mood labels by each annotator. The agreement ratio between two sets of annotations on a song is calculated as the number of identically tagged labels divided by the number of all labels applied to the song (Hu & Lee, 2012; Nowak & Rüger, 2010). The agreement ratios of all songs were then averaged to show an aggregated value. The averaged agreement ratio in CH818 is 0.37 (Table 3, first row). It is comparable to the agreement ratio of 0.35 between Chinese listeners of Western songs in Lee and Hu (2014) and higher than the ratio of 0.15 between Korean listeners on K-Pop songs (Hu et al., 2014).

To improve the reliability of the annotations, we calculated the majority voted annotations of each song. That is, only mood labels applied two or three times to a song were counted. In this way, the average number of labels for each



Chinese Music Annotation

请听这段音乐，并从以下的词语中选择能够描述这段音乐所表达的情绪 （选择所有符合条件的词，请选择至少一个词）：

□ 好斗的 aggressive	□ 凄美的 poignant	□ 平静的 calm/peaceful	□ 愉快的 rollicking	□ 傻乎乎的 silly	□ 有激情的 passionate
□ 暴躁的 fiery	□ 苦乐参半的 bittersweet	□ 温柔的 tender	□ 欢乐的 cheerful	□ 哗众取宠的 campy	□ 使人兴奋的 rousing
□ 紧张的/焦虑的 tense/anxious	□ 想往的 wistful	□ 梦幻般的 dreamy	□ 有趣的 fun	□ 古怪的 quirky	□ 自信的 confident
□ 激烈的 intense	□ 忧郁的 brooding	□ 抚慰的 soothing	□ 甜美的 sweet	□ 异想天开的 whimsical	□ 热闹的 boisterous
□ 宣泄的 visceral	□ 秋天的 autumnal	□ 放松的 relaxed	□ 可亲的 amiable	□ 讽刺的 wry	□ 吵闹的 rowdy
□ 有暴力倾向的 volatile	□ 有内涵的 literate	□ 乡愁的 nostalgic	□ 幽默的 humorous	□ 诙谐的 witty	□ 鼓励的 encouraging

(a) Categorical labels

请根据该音乐片段所表达的情绪，用下面的滑动选择给出一个评价值（负面或正面情绪）： -10代表非常负面，10代表非常正面，中间连续的数值代表负面或正面情绪的不同程度。
(注意！这里的情绪是指歌曲想要表达的情绪，而不是指这首歌好不好听，或是您喜不喜欢。)

负评价值：让人心情坏的情绪：如悲伤,愤怒,紧张,沉闷等
中间评价值：让人心情不好也不坏的情绪
正评价值：让人心情好的情绪，如快乐,兴奋,轻松,平静 等
-10
-8
-6
-4
-2
0
2
4
6
8
10

评价值 （负面一正面） 8.8

(b) Slider bar for valence

请根据该音乐片段所表达的情绪，用下面的滑动选择给出一个唤醒度（低或高）： 一10代表非常低，10代表非常高，中间连续的数值代表不同程度的唤醒度。

低唤醒度：让人较不激动的情绪，如:轻松,平静,抑郁,沉闷 等
中间唤醒度：不激动也不平静
高唤醒度：让人较激动的情绪,如:高兴,兴奋,愤怒,澎湃 等
-10
-8
-6
-4
-2
0
2
4
6
8
10

唤醒度（低一高） -7.2

(c) Slider bar for arousal

FIG. 2. Screenshots of the annotation interface of one clip. [Color figure can be viewed at wileyonlinelibrary.com]

song was 4.03, with a standard deviation of 1.96. The number of songs with each mood label is presented in Table 5, and the agreement ratios are shown in Table 3 (second row), which are much higher than those between the annotators (first row). In the subsequent analysis, we use the majority voted categorical annotations unless otherwise specified.

TABLE 3. Agreement ratio between each pair of annotations (A1, A2, A3 represents the three annotators).

Original	A1 vs. A2	A2 vs. A3	A3 vs. A1	Average
	0.46	0.39	0.33	0.37
With majority vote	A1 vs. Majority vote	A2 vs. Majority vote	A3 vs. Majority vote	Average
	0.63	0.73	0.63	0.66

Dimensional Annotations and Reliability

Figure 3 illustrates the scatterplot of the annotations in the valence–arousal space. There are more values in the first quadrant than others, indicating songs with positive valence and positive arousal annotations are the most popular in the data set. There are no songs in the bottom right corner of the space (near [10, −10], indicating high valence and low arousal values). These observations are consistent with the MER60 (Yang & Chen, 2011a) and AMG1608 (Chen, Yang, Wang, & Chen, 2015) data sets, which both consist of Western Pop songs.

Following existing research (Hu et al., 2014; Yang & Chen, 2011a), we measured agreement across annotators on dimensional annotations using Krippendorff's alpha (Gwet, 2010). The results were 0.49 and 0.62 for valence (V) and arousal (A), respectively, which fall in "fair" and "moderate" agreement (Gwet, 2010). Similar to the findings in previous studies (Hu et al., 2014; Kim et al., 2010; Yang & Chen, 2012), valence annotations are harder to reach agreement than arousal. The alpha values of CH818 are higher than those of AMG1608 (V: 0.31, A: 0.46) (Chen et al., 2014) and the MIREX K-POP mood data set (V: 0.28, A: 0.54) (Hu et al., 2014),[3] while comparable to those of MER60 (V: 0.39, A: 0.70).

It is noteworthy that due to the subjectivity of music mood, interannotator agreement for music mood is usually moderate. We have thus split CH818 into three subsets with controlled annotation reliability levels. An initial experiment has shown better prediction performances on subsets with higher reliability levels, which is in accordance with the results in Hu and Yang (2016). The subsets are also released with CH818 to facilitate further studies.

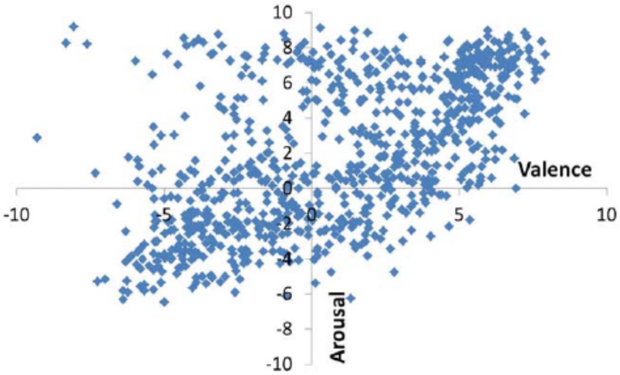


FIG. 3. Scatterplot of valence and arousal values given by annotators. [Color figure can be viewed at wileyonlinelibrary.com]

Audio Features

To capture a range of musical characteristics that are recognized to be related to music mood (Juslin, 2000), five categories of acoustic features were extracted from the songs in CH818 using tools that are well recognized in the MIR field. These features can be broadly categorized into loudness, pitch, rhythm, timbre, and harmony. Table 4 lists the features, their dimensionality, and the music audio processing tools that were used to extract them. For specific descriptions of the features, please refer to Hu and Yang (2016).

Maintenance of the Data Set

The metadata, three sets of annotations in both categorical and dimensional models, and audio features presented will be publicly available for research purposes. Due to copyright restrictions, the audio clips cannot be shared. To compensate for this disadvantage, we plan to continue extracting new audio features, and add them to the data set. For this purpose, we welcome MIR researchers to share their feature extraction programs with us so that we can run the programs against the audio files of CH818 and share the newly extracted features by adding them to the webpage of the data set.[1]

It is our plan to maintain the data sets for long-term availability and values (Donnelly, 2014). While preserving research data sets in MIR is beyond the scope of this paper, we are seeking opportunities to deposit stable versions of the CH818 data set to a data or institutional repository for the sake of long-term sustainability.

RQ1: Mood Spaces of Chinese and Western Pop Music

The mood space of C-Pop is constructed using the categorical annotations in CH818. It is then compared to that of Western Pop music (as represented in Yang & Hu, 2012) in three aspects: i) distributions of mood labels; ii) relative distance among mood labels; and iii) mood clusters.

Mood Space of CH818

To visualize the mood space of CH818, we projected the mood labels based on their distances to a 2D space using multidimensional scaling (MDS), the same method used in Yang and Hu (2012). The distance between a pair of mood categories was calculated based on the common songs shared by them in the majority voted categorical annotations. As shown

[3] The MIREX K-Pop mood data set reported intraclass correlation (ICC) with the one-way random model, which is equivalent to Krippendorff's alpha (Gwet, 2010).

TABLE 4. Extracted audio features (Hu & Yang, 2016).

Category	Features	Dimensions	Tool
Loudness	RMS Energy	2	M
	Loudness	18	P
Pitch	Pitches	88	C
	Chroma-Pitch	24	C
	Chroma-Log-Pitch	24	C
	Chroma Energy Normalized Statistics	24	C
	Chroma Discrete Cosine Transform-Reduced log Pitch	24	C
Rhythm	Fourier-based cyclic tempogram	80	T
	Autocorrelation-based cyclic tempogram	80	T
	Rhythm strength, regularity, clarity, average onset frequency, average tempo	5	M
Timbre	Mel-frequency cepstral coefficients	120	M
	Spectrum characteristics	28	M
	Dissonance	8	P
Harmony	Key clarity, mode, and harmonic change detection function	6	M
	Tonalness, multiplicity, and chord change likelihood	8	P

Note. M = MIR toolbox (Lartillot & Toiviainen, 2007); P = PsySound (Cabrera, 1999); T = Tempogram toolbox (Grosche & Müuller, 2011); C = Chroma toolbox (Müller & Ewert, 2011).

in Figure 4, the proximity of labels reflects their semantic closeness in most cases (e.g., "aggressive" is close to "volatile" and away from "dreamy").

Inspired by the MIREX five-cluster mood model (Table 1) where individual labels were merged into clusters, we also constructed a set of mood clusters for CH818, using the same agglomerative clustering technique used on the MIREX AMC data set (Hu et al., 2008). The resultant dendrogram showed six distinct clusters, but two labels, "autumnal" and "nostalgic," were only merged in very late iterations (i.e., they are not very close to any clusters). Therefore, these two labels were excluded from the resultant clusters. This exclusion did not result in any songs being excluded because few songs were annotated with them and all those songs also had other labels. As shown in Table 5, overall labels in each of the clusters are semantically consistent. Among the clusters, C_6 contains five of the seven labels added in this study. The numbers of songs in the clusters are quite balanced except for C_5 and C_6, which may reflect that there are more calm and tender C-Pop songs than aggressive ones (Hu & Lee, 2016; Lee & Hu, 2014; Liu & Mason, 2010).

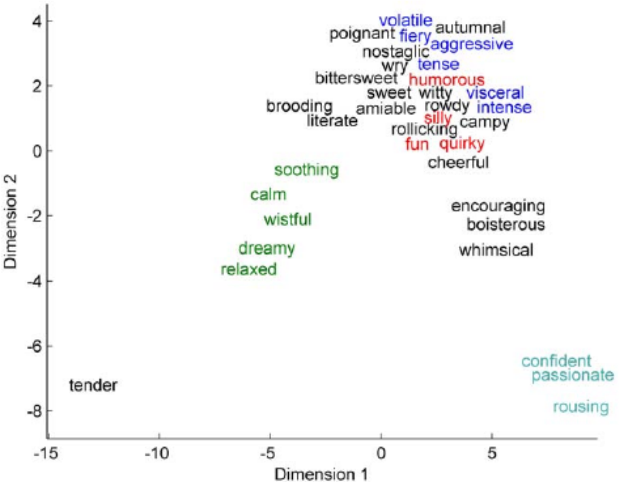


FIG. 4. Mood space of CH818. [Color figure can be viewed at wileyonlinelibrary.com]

Cross-Cultural Comparison

With a substantial number of C-Pop songs and thorough mood annotations, CH818 provides a representative case for exploring the mood space of C-Pop. The space can then be compared to that of music from other cultures, providing systematic insights on how music moods are similar or different across cultures. As the mood labels in this study are originally from AllMusicGuide, the Western song data set in Yang and Hu (2012) is used for comparison, which consists of 1,520 "Top Songs" associated with mood labels on AllMuiscGuide. Figure 5 shows the number of times each mood label is applied to both data sets. It should be noted that in AllMusicGuide, each mood label can at most have 100 top songs. Nonetheless, some patterns of differences are clear to see: "dreamy," "relaxed," and "calm" are popular in CH818 but not in the Western data set, whereas "wry," "aggressive," and "fiery" are the opposite. This observation may be attributed to the songs. Influenced by the Chinese culture which values modesty and introversion, Chinese music may favor music elements that sound mellow, mellifluous, and light-spirited (Hu & Lee, 2016; Liu & Mason, 2010).

The mood space of CH818 shown in Figure 4 can be compared to that of the Western songs in Yang and Hu (2012; reprinted in Figure 6). In both spaces, labels sharing similar semantics are gathered together, such as the group of "aggressive," "fiery," and "volatile" and that of "confidence," "passionate," and "rousing." The two spaces also differ in several ways. The group of "aggressive" and that of "confident" are close in the Western space but far apart in the CH818 space. This is possibly related to Chinese people's tendency of disliking radical moods (e.g.,

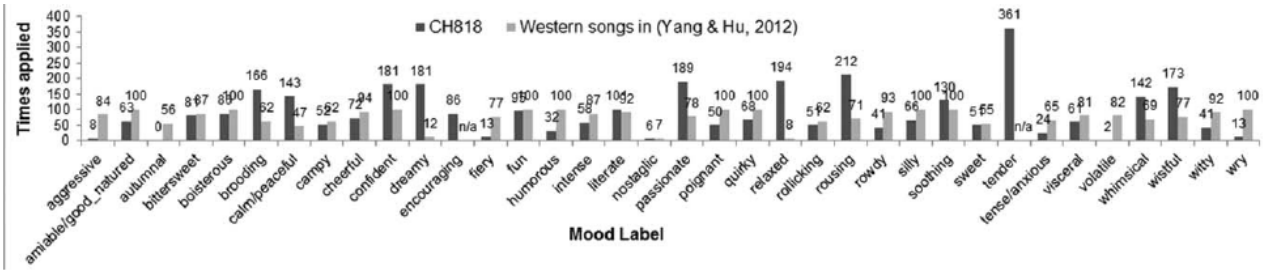


FIG. 5. Distribution of mood labels ("n/a" indicates the label was not in the Western song data set).

"aggressive"). In addition, the middle parts of both spaces are crowded by many labels, but the group of "relaxed," "dreamy," "calm," and that of "humorous," "silly," "quirky" are separated from the middle in the two spaces, respectively, indicating that they represent quite distinct moods in each culture, but not the other.

We also compare the mood clusters of CH818 (Table 5) to that of the MIREX AMC five-cluster representation model (Table 1). The two sets of clusters are comparable, as both were derived from expert annotations using similar methods. The comparison discloses some commonalities: i) the fifth clusters (C_5) in both sets completely match; ii) all labels in the first clusters (C_1) match except for "encouraging," which is new in CH818 and shares similar semantics with other labels in C_1; iii) C_2 and C_4 in both spaces are very similar, except two labels, "humorous" and "witty," are moved from C_4 of the MIREX model to C_2 of the CH818 model. In fact, a previous study (Lee & Hu, 2012) found that MIREX_C_2 and MIREX_C_4 were indeed the most confusing pairs for MIREX evaluators, which can at least patricianly explain the change of cluster membership of the two labels; iv) All three labels in CH818_C_3 are in MIREX_C_3, while two labels in the latter, "literate" and "wistful," move to CH818_C_6, which is unique to CH818. The biggest difference between the two sets of clusters is probably the new emergence of CH818_C_6. This suggests that calm and relaxed moods are more prominent in C-Pop than Western Pop songs.

RQ2: Relationship Between Categorical and Dimensional Mood Models

Mapping Between the Models

It is well recognized that both categorical and dimensional models of music mood have their own advantages and disadvantages, and thus it is desirable to have both annotations. Alternatively, as human annotations are expensive and time-consuming to obtain, in a more realistic sense it would be helpful if one set of annotations could be inferred from the other. Given that both kinds of models are widely used in the field, MIR researchers have been interested in the relationships between them (Wang et al., 2012). As the first C-Pop data set with annotations in both categorical and dimensional models, CH818 provides an excellent opportunity to investigate the mappings between them.

Figure 7 illustrates the 36 categorical labels in the valence-arousal (VA) 2D space where the position of each label is decided by the mean valence and arousal values of all songs annotated with that label. The distribution of labels in Figure 7 is in accordance with song distribution plotted in

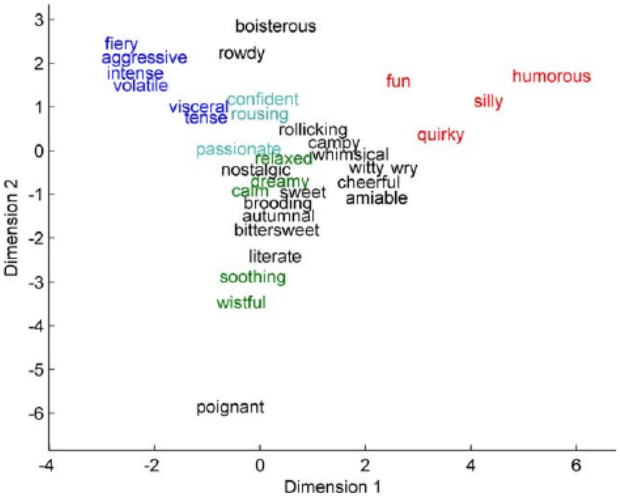


FIG. 6. Mood space of Western pop in (Yang & Hu, 2012, p. 22). [Color figure can be viewed at wileyonlinelibrary.com]

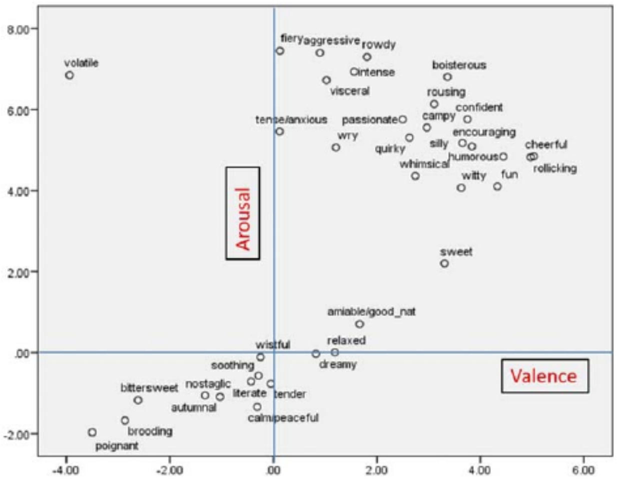


FIG. 7. CH818 mood labels plotted in valence-arousal space. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5. Clustering results of mood labels based on majority votes of categorical annotations.

Cluster (no. of songs)	Mood labels (no. of songs)
CH818_C_1 (293)	Passionate (189), rousing (212), confident (181), boisterous (86), encouraging (86), rowdy (41)
CH818_C_2 (219)	Rollicking (51), cheerful (72), fun (95), sweet (51), amiable/good natured (63), humorous (32), witty (41)
CH818_C_3 (206)	Poignant (50); brooding (166); bittersweet (81)
CH818_C_4 (194)	Silly (66), campy (52), quirky (68), whimsical (142), wry (13)
CH818_C_5 (75)	Aggressive (8), fiery (13), volatile (2), tense/anxious (24), intense (58), visceral (61)
CH818_C_6 (506)	Calm/peaceful (143), tender (361), relaxed (194), dreamy (181), soothing (130), literate (101), wistful (173)

Figure 3, in that most songs and labels fall in the first and third quadrants (+V, +A, and –V-A). In addition, the relative positions of the mood labels in the VA space generally match the cluster compositions shown in Table 5. These observations support that the two sets of annotations in this data set are well connected. For instance, “passionate,” “confident,” and “rousing” are placed together (C_1 in Table 5) in the first quadrant while “wistful,” “soothing,” and “literate” (C_6 in Table 5) cluster near the zero point (with slightly negative valence and arousal). It is noteworthy that the arousal axis in Figure 7 is slightly shifted to the left compared to Figure 1, which is based on Western music. Some mood categories such as “aggressive” are of negative valence in Russell’s model (Figure 1) but carry positive values here in Figure 7. This difference may be attributed to the sparseness of aggressive C-Pop songs and/or Chinese listeners’ low threshold for perceiving a piece as “aggressive.” Future studies are warranted to further verify the reasons.

Prediction Between Models

We conducted experiments to test the extent to which one set of annotations on CH818 could be predicted by the other. A linear regression model was constructed to predict valence or arousal annotations of the songs based on the binary variables of the categorical labels. The R^2 of the two models are 0.53 for valence and 0.70 for arousal, which are comparable or even higher than regression performances based on audio features (Kim et al., 2010; Yang & Chen, 2012). A closer examination of the significant predictor variables (i.e., the categorical labels) in both models reveals that those unique to each model do bear semantics indicating valence or arousal. For example, “volatile,” “cheerful,” and “encouraging,” which clearly indicate negative or positive sentiment, are unique to the valence prediction model. In contrast, labels with energy implications such as “fiery,” “boisterous,” “calm/peaceful,” and “dreamy” are unique to the arousal model.

TABLE 6. Classification and regression experiment results.

Classification	**Cluster**	**Accuracy**	**Kappa (κ)**
	CH818_C_1	0.91	0.81
	CH818_C_2	0.74	0.50
	CH818_C_3	0.82	0.64
	CH818_C_4	0.81	0.63
	CH818_C_5	0.90	0.80
	CH818_C_6	0.92	0.82
Regression	**Dimension**	**R^2**	**RMSE**
	Valence	0.25	3.54
	Arousal	0.79	2.31

For classification of mood labels, we combined the labels into clusters shown in Table 5, to reduce the number of models and to avoid problems caused by data sparseness. Six logistic regression models were built based on the valence and arousal values to classify whether or not a song belongs to the clusters. Compared to a baseline of random chance (accuracy = 50%), the results show that CH818_C_1, C_3, C_5, and C_6 are highly predictable (accuracy =89.73%, 88.14%, 93.64%, and 87.90%, respectively). C_2 and C_4 performed less ideal (accuracy = 74.82% and 75.31%, respectively), which again is probably related to the fact that the terms in these two clusters were somewhat confusing to human annotators (Lee & Hu, 2012). The results are comparable to audio-based classifiers (see next section) and those reported in the literature (Kim et al., 2010; Yang & Chen, 2012). It is particularly encouraging given the fact that there are only two predictor variables (i.e., valence and arousal).

The experiments show that the relationship between the two sets of annotations in CH818 is reliable. Annotations of one model could predict those of the other. This makes it one step further towards the goal of unified user interfaces that can benefit from *both* mood representation models. To this end, CH818 itself could serve as a ground truth on which various techniques can be evaluated. Furthermore, this could be in conjunction with cross-cultural and cross-data set experiments, to find out whether and to what extent the predictability between annotation models can transcend across such boundaries as culture, genre, etc.

RQ3: Music Mood Recognition (MMR) for C-Pop

With both categorical and dimensional annotations, as well as rich audio features extracted (Table 4), CH818 can serve as a benchmarking data set for evaluating the tasks of mood classification (predicting mood labels) and mood regression (predicting valence and arousal values) for C-Pop songs. To illustrate the utility of CH818 in these tasks and benchmark the performances, we conducted two sets of classification and regression experiments on this data set. The first was to evaluate the classification and regression performances on the combination of all extracted audio features. The second was to compare the performances of the five sets of features summarized in Table 4. Performances are compared to the state-of-the-art of MMR on Western music.

DOI: 10.1002/asi

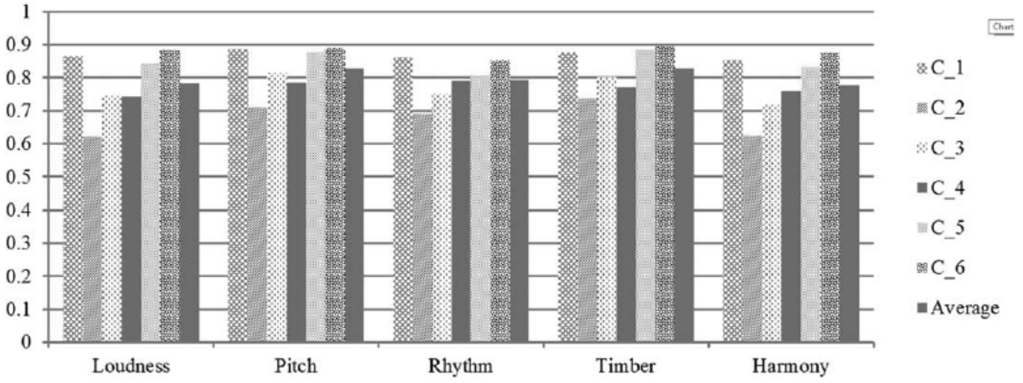


FIG. 8. Classification accuracies across clusters using different audio features sets.

MMR on All Extracted Features

As support vector machines (SVM) have been widely adopted in MMR studies and shown superior performance, we used SVM classification and regression (SVR) models. For classification, we used the clusters in Table 5 as the prediction target. As each song may be in multiple clusters, we constructed the problem into a binary one. That is, one classifier was built for each cluster, with positive examples being songs labeled with any terms contained in this cluster. An equal number of negative examples was then randomly selected from the remaining songs.[4] The sampling process was repeated 20 times and the averaged results are reported. The measures of accuracy and Cohen's kappa are used to gauge the performances. For binary classification with balanced data, a trivial prediction by chance would result in an accuracy of 50% and a kappa of 0. For regression, the valence and arousal values averaged across annotators were used as the ground truth. The measures of R^2 and root mean squared error (RMSE) are calculated. Both the classification and regression experiments used the RBF kernel with default parameter settings, and were conducted with 10-fold cross-validation. The results are shown in Table 6.

The classification performances on C_1, C_5, and C_6 are fairly good, not only with high accuracies of over 90%, but also with kappa values in the "very good agreement" level ($\kappa = 0.80$–1.00) with the ground truth (Gwet, 2010). The accuracy levels of these clusters are also close to the latest result of the MIREX Audio Tag Classification (ATC) task (90%),[5] where the task was also binary classification. The performances on C_3 and C_4 are in "good agreement" with the ground truth ($\kappa = 0.60$–0.79), while C_2 is only in the "moderate agreement" ($\kappa = 0.40$–0.59). The less satisfactory performance on C_2 may be due to the inherent difficulty of prediction of some moods from audio features (e.g., the "happy" group in the ATC task). The performances of regression are comparable to the literature, where R^2 is usually from 0.17 to 0.30 and 0.58 to 0.80 for valence and arousal, respectively (Guan, Chen, & Yang, 2012; Kim et al., 2010; Yang & Chen, 2012). These results verified the feasibility of evaluating music mood recognition with CH818 as well as the applicability of the audio features proposed from Western music on C-Pop songs.

Comparison of Feature Sets

To further investigate the relative advantages of the different feature sets on C-Pop songs, we conducted similar classification and regression experiments on each of the five feature sets: loudness, pitch, rhythm, timbre, and harmony. Figures 8 and 9 show the results. For classification, timbre and pitch features performed better than other features, and such advantages are more obvious on the worse-performing clusters, C_2 and C_3. Timbre features have been shown effective in MMR of Western music (Barthet et al., 2013; Yang & Chen, 2012). The fact that pitch features (designed from Western music) worked well also evidences that C-Pop is influenced by the 12-pitch class structure in Western music. For regression, timbre, pitch, and loudness features were good for arousal predictions, whereas rhythm and timbre features were good for valence prediction. These results are in accordance with those in Hu and Yang (2016) where similar features were also evaluated on Western music. The applicability of state-of-the-art audio features on C-Pop not only supports the evolutionary relationship between C-Pop and Western Pop (Liu & Mason, 2010), but also inspires further research on comparative studies of different feature sets on music of different cultural origins.

In the context of cross-cultural music retrieval, such evaluation can help determine which features that have been working for Western music are also applicable to non-Western music. In a broader sense, CH818 can work with other data sets for cross-data set mood recognition, where models are trained with one data set and evaluated with another (e.g., Hu & Yang, 2016). Depending on the composition of the data sets, such experiments could help assess the generalizability of mood recognition techniques across different boundaries such as music types, annotator backgrounds, annotation methods, listening situations, etc.

[4] The random sampling process is reversed for CH818_C_6, as there are more positive examples than negative ones.

[5] http://nema.lis.illinois.edu/nema_out/mirex2014/results/atg/subtask2_report/bin/summary.html

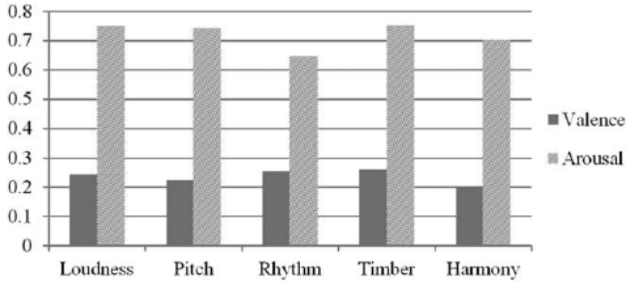


FIG. 9. Regression performances (R^2) on valence and arousal using different audio features sets.

Evaluations of these tasks are essential for developing MIR systems that cater to increasingly diversified listener groups.

Conclusions and Future Work

This study investigated a few longstanding questions on music mood recognition, using the CH818 data set of C-Pop music. The comparison of mood spaces between Chinese and Western Pop music demonstrated the high consistency between the two, as well as the distinct and salient cluster of “calm,” “soothing,” “relaxed” mood in C-Pop music. Such a comparison can also be made with music from other cultural origins. Collectively, the results will show a more complete picture of the mood spaces of music in the world and contribute to the goal of global access to music. The mapping between categorical and dimensional models verifies the internal consistency of the two kinds of models, which provides a theoretical base of computational work on automated prediction of one representation from the other.

The applicability of MMR techniques for both classification and regression on CH818 was verified by a set of experiments. The comparison of performances across different audio feature sets helps answer the question of which features are good for the prediction of different mood categories or dimensions for C-Pop.

It is noteworthy that the songs in CH818 span mainly across one decade and thus may not be sufficiently comprehensive to represent the entire landscape of C-Pop. Nonetheless, by exploring the raised research questions using CH818, we hope this study can stimulate further explorations and innovations related to MMR. We will further develop the CH818 data set by adding more dimensions of information into it, such as lyrics, metadata, and/or editorial/social tags, so that it can be used for multimodal MMR (Hu et al., 2016) and to help explore the question of the effects of lyrics and melodies on music mood (Ali & Peynircioğlu, 2006). As data sets are foundations of development of a field, it is also an intriguing and important direction to explore how an MIR data repository can be built up. This will be beneficial to the entire MIR community.

Acknowledgment

This study was partially supported by a Seed Fund for Basic Research from the University of Hong Kong and Grant NSC 102-2221-E-001-004-MY3 from the Ministry of Science and Technology of Taiwan.

References

Ali, S.O., & Peynircioğlu, Z.F. (2006). Songs and emotions: Are lyrics and melodies equal partners? Psychology of Music, 34, 511–534.

Barthet, M., Fazekas, G., & Sandler, M. (2013). Music emotion recognition: From content-to context based models. In M. Aramaki, M. Barthet, R. Kronland-Martinet, & S. Ystad (Eds.), From sounds to music and emotions (pp. 228–252). Berlin, Heidelberg: Springer.

Bischoff, K., Firan, C.S., Nejdl, W., & Paiu, R. (2009). How do you feel about “Dancing Queen”? Deriving mood and theme annotations from user tags. In Proceedings of Joint Conference on Digital Libraries (JCDL), Austin, Texas, (pp. 285–294). New York: ACM Press.

Cabrera, D. (1999). PSYSOUND: A computer program for psychoacoustical analysis. In Proceedings of the Australian Acoustical Society Conference, Melbourne, (Vol. 24, pp. 47–54). Darra, QLD: Australian Acoustical Society.

Chen, Y.A., Wang, J.C., Yang, Y.H., & Chen, H.H. (2014). Linear regression-based adaptation of music emotion recognition models for personalization. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2149–2153).

Chen, Y.A., Yang, Y.H., Wang, J.C., & Chen, H.H. (2015). The AMG1608 dataset for music emotion recognition, In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brisbane, Australia, (pp. 693–697). Piscataway, NJ: IEEE.

Chow, Y.F., & de Kloet, J. (2010). Blowing in the China Wind: engagements with Chineseness in Hong Kong’s zhongguofeng music videos. Visual Anthropology, 24(1–2), 59–76.

Donnelly, M. (2014). Review: Research data management: Practical strategies for information professionals. International Journal of Digital Curation, 9, 1–5.

Downie, J.S., Hu, X., Lee, J.H., Choi, K., Cunningham, S.J., Hao, Y., & Bainbridge, D. (2014). Ten years of MIREX (Music Information Retrieval Evaluation eXchange): Reflections, challenges and opportunities. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Taipei, Taiwan, Canada: ISMIR.

Eerola, T., & Vuoskoski, J.K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. Music Perception: An Interdisciplinary Journal, 30, 307–340.

Grosche, P., & Müuller, M. (2011). Extracting predominant local pulse information from music recordings. IEEE Transactions on Audio Speech and Language Processing, 19, 1688–1701.

Guan, D., Chen, X., & Yang, D. (2012). Music emotion regression based on multi-modal features. In Proceedings of the International Symposium on Computer Music Modeling and Recognition, London, UK, (pp. 70–77).

Gwet, K.L. (2010). Handbook of inter-rater reliability. Gaithersburg, MD: Advanced Analytics.

Hevner, K. (1936). Experimental studies of the elements of expression in music. The American Journal of Psychology, 48, 246–268.

Hu, X. (2010). Music and mood: Where theory and reality meet. In iConference, Champaign, IL (pp. 1–8).

Hu, X., Choi, K., & Downie, J.S. (2016). A framework for evaluating multimodal music mood classification. Journal of the Association for Information Science and Technology. doi:10.1002/asi.23649

Hu, X., & Downie, J.S. (2007). Exploring mood metadata: Relationships with genre, artist and usage metadata. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Vienna, Austria, (pp. 462–467). Canada: ISMIR.

Hu, X., Downie, J.S., & Ehmann, A. (2009). Lyric text mining in music mood classification. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Kobe, Japan, (pp. 441–446). Canada: ISMIR.

Hu, X., Downie, J.S., Laurier, C., & Ehmann, M.B.A.F. (2008). The 2007 MIREX audio mood classification task: Lessons learned. In

Proceedings of the International Symposium on Music Information Retrieval (ISMIR) (pp. 462–467).

Hu, X., & Lee, J.H. (2012). A cross-cultural study of music mood perception between American and Chinese listeners. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Porto, Portugal, (pp. 535–540). Canada: ISMIR.

Hu, X., & Lee, J.H. (2016). Towards global music digital libraries: A cross-cultural comparison on the mood of Chinese music. Journal of Documentation, 72(5), 858–877.

Hu, X., Lee, J.H., Choi, K., & Downie, J.S. (2014). A cross-cultural study on the mood of K-POP songs. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Taipei, Taiwan, (pp. 385–390). Canada: ISMIR.

Hu, X., & Yang, Y.H. (2016). Cross-dataset and cross-cultural music mood prediction: A case on western and Chinese pop songs. IEEE Transactions on Affective Computing. doi:10.1109/TAFFC.2016.2523503 [epub ahead of print].

Juslin, P.N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. Journal of Experimental Psychology: Human Perception and Performance, 16, 1797–1813.

Kim, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., . . . Turnbull, D. (2010). Music emotion recognition: A state of the art review. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Utrecht, Netherlands, (pp. 255–266). Canada: ISMIR.

Lartillot, O., & Toiviainen, P. (2007). MIR in Matlab (II): A toolbox for musical feature extraction from audio. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Vienna, Austra, (pp. 127–130). Canada: ISMIR.

Laurier, C., Grivolla, J., & Herrera, P. (2008). Multimodal music mood classification using audio and lyrics. In Proceedings of the International Conference on Machine Learning and Applications, San Diego, CA, (pp. 688–693). Piscataway, NJ: IEEE.

Laurier, C., Sordo, M., Serra, J., & Herrera, P. (2009). Music mood representations from social tags. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Kobe, Japan, (pp. 381–386). Canada: ISMIR.

Lee, J.H., Choi, K., Hu, X., & Downie, J.S. (2013). K-Pop genres: A cross-cultural exploration. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Curitiba, Brazil, (pp. 529–534). Canada: ISMIR.

Lee, J.H., & Hu, X. (2012). Generating ground truth for mood classification in music digital libraries using mechanical turk. In Proceedings of the IEEE-ACM Joint Conference on Digital Libraries, Washington, DC, (pp. 129–138).

Lee, J.H., & Hu, X. (2014). Cross-cultural similarities and differences in music mood perception. In Proceedings of the iConference. 2014, Berlin, Germany.

Liu, J., & Mason, C. (2010). A critical history of new music in China. Hong Kong: Chinese University Press.

Müller, M., & Ewert, S. (2011). Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In Proceedings of the International Conference on Music Information Retrieval (ISMIR) (pp. 215–220).

Nowak, S., & Rüger, S. (2010). How reliable are annotations via crowdsourcing: A study about inter-annotator agreement for multi-label image annotation. In Proceedings of the ACM International Conference on Multimedia Information Retrieval, Philadelphia, Pennsylvania, (pp. 557–566). New York: ACM.

Russell, J.A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39, 1161–1178.

Serra, X. (2011). A multicultural approach in music information research. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Miami, Florida, (pp. 151–156). Canada: ISMIR.

Singhi, A., & Brown, D.G. (2014). On cultural, textual and experiential aspects of music mood. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Taipei, Taiwan, (pp. 3–8). Canada: ISMIR.

Speck, J.A., Schmidt, E.M., Morton, B.G., & Kim, Y.E. (2011). A comparative study of collaborative vs. traditional musical mood annotation. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Miami, Florida, (pp. 549–554). Canada: ISMIR.

The International Federation of the Phonographic Industry. (2014). *IFPI digital music report*. Retrieved on from http://www.ifpi.org/downloads/Digital-Music-Report-2014.pdf

Trohidis, K., Tsoumakas, G., Kalliris, G., & Vlahavas, I.P. (2008). Multi-label classification of music into emotions. In Proceedings of the International Conference on Music Information Retrieval (ISMIR) (pp. 325–330).

Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G. (2008). Semantic annotation and retrieval of music and sound effects. IEEE Transactions on Audio Speech and Language Processing, 16, 467–476.

Wang, J.C., Yang, Y.H., Wang, H.M., & Jeng, S.K. (2012). The acoustic emotion Gaussians model for emotion-based music annotation and retrieval. In Proceedings of the ACM International Conference on Multimedia, Nara, Japan, (pp. 89–98). New York: ACM.

Yang, Y.H., & Chen, H.H. (2011a). Prediction of the distribution of perceived music emotions using discrete samples. IEEE Transactions on Audio Speech and Language Processing, 19, 2184–2196.

Yang, Y.H., & Chen, H.H. (2011b). Ranking-based emotion recognition for music organization and retrieval. IEEE Transactions on Audio Speech and Language Processing, 19, 762–774.

Yang, Y.H., & Chen, H.H. (2012). Machine recognition of music emotion. A review. ACM Transactions on Intelligent Systems and Technology (TIST), 3, 40.

Yang, Y.H., & Hu, X. (2012). Cross-cultural music mood classification: A comparison on English and Chinese songs. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Porto, Portugal, (pp. 19–24).

Yang, Y.H., Lin, Y.C., Su, Y.F., & Chen, H.H. (2008). A regression approach to music emotion recognition. IEEE Transactions on Audio, Speech, and Language Processing, 16, 448–457.