What Does Music Mood Mean for Real Users?

Jin Ha Lee
University of Washington
Mary Gates Hall, Suite 370
Seattle, WA 98195
+1 206.685.0153
jinhalee@uw.edu

Trent Hill
University of Washington
Mary Gates Hall, Suite 370
Seattle, WA 98195
+1 206.685.0153
tghill62@uw.edu

Lauren Work
University of Washington
Mary Gates Hall, Suite 370
Seattle, WA 98195
+1 206.685.0153
lawork@uw.edu

ABSTRACT
Mood has recently received increasing attention as an interesting approach for organizing and accessing music. Our understanding of how users determine and describe music mood, however, is not fully developed. In this exploratory study, we investigate the concept of music mood from the end-user's perspective. In particular, we want to see how users describe music mood in their own terms as they react to different musical features. We investigate this by asking users to provide mood tags for various cover versions of the same song. The findings suggest that users rely on a small limited set of mood terms, although they do use a wide variety of terms. Typically, certain moods seem to carry over multiple cover versions despite differences in musical features. Along with lyrics, tempo, instrumentation, and delivery, factors like sources of mood, genre, musical expectancy, cultural context, etc. also seem to affect how people feel about music.

Categories and Subject Descriptors
H.1.2. [**Information Systems**]: User/Machine Systems – h*uman factors.*

General Terms
Human Factors, Theory.

Keywords
Music information retrieval, Music, Mood, Emotion, Mechanical Turk, Tagging, User.

1. INTRODUCTION
Music mood offers an interesting opportunity for meaningful organization and access of music collections, as well as generating recommendations to users. In recent years, music mood has received increasing attention in the music information retrieval literature. A number of music information retrieval (MIR) researchers have explored music mood with regards to lyrics [5], [10], [29], social tags, [3], [4], [14], [17], and collaborative games [1], [12], [15], [20], [26]. Emotional aspects of music, however, have been explored and studied in previous literature including music theory and psychology. Many of these studies attempt to organize mood in some meaningful way: either by representing them in a low dimensional space or categorizing them into different mood clusters [8]. The dimensional models for mood descriptors suggested in the seminal work by Thayer and Russell have proven useful as a foundation for many studies that attempt to categorize different moods into a low dimensional space. Another seminal work by Hevner [7] which dates back to the 1930s is still relevant for understanding how to categorize mood labels in different mood clusters. Many of these studies have provided us with some understanding of the concept of music mood. Hu [8] summarizes what we know about music mood based on the findings from various psychology studies as follows:

1. Mood effect does exist in music.
2. Not all moods are equally likely to be aroused by listening to music.
3. Uniform mood effects among different people do exist.
4. Not all types of moods have the same level of agreement among listeners.
5. There is some correspondence between listeners' judgments on mood and musical parameters such as tempo, dynamics, rhythm, timber, articulation, pitch, mode, tone attacks and harmony.

We are, however, still in the process of improving our understanding of music mood in information science [8]. In particular, few studies have explored what comprises music mood for real users. In other words, what factors are key for users as they determine and describe the mood of a song? As Hu [8] stated, we know some musical parameters affect how listeners feel about music, but are there other factors which also playing a role here?

In the MIR community, previous efforts in determining music mood have mostly focused on musical features. For instance, the Music Information Research Evaluation eXchange (MIREX) [1] carries out several music-related evaluation tasks including an Audio Music Mood Classification task. The objective of this task is to accurately categorize song based on their mood into one of the five mood clusters [9]. Researchers submit different

iConference 2012, February 7-10, 2012, Toronto, ON, Canada
ACM 978-1-4503-0782-6/12/02

[1] Music Information Retrieval Evaluation eXhange (MIREX) is the annual evaluation campaign for various music information retrieval algorithms hosted by the International Music Information Retrieval Systems Evaluation Lab (IMIRSEL) at the University of Illinois at Urbana-Champaign.

algorithms to be tested against the ground truth generated by human evaluators who listen to given songs and categorize the song into the most appropriate mood cluster. In the past, when generating the ground truth for this task, the human evaluators were asked to ignore the lyrics and mainly focus on the audio representation of the music. While this may be helpful for the purpose of evaluation, we believe that this does not accurately reflect how real users perceive the mood of music in real life.

Our motivation for this study is to further our understanding of music mood, focusing on how real users describe music moods in their own terms as well as how different factors can affect users' emotional reaction to music. This is part of a larger user study in which we explore different aspects of music mood such as the agreement between users' and music experts' mood tags, how users resolve conflicts between lyrics and melody, etc. Due to space limits, we only present part of our study in this paper.

2. RESEARCH QUESTIONS AND STUDY DESIGN

2.1 Research Questions

We explore two main research questions in this paper:

I) What terms do people use when they describe the mood of a song?

II) What factors affect users' determination of the mood of a song?

To address these questions, we created an online survey which asked participants to listen to a number of short music clips and to assign self-generated tags describing the mood of each song. The survey was hosted by Amazon Mechanical Turk (MTurk). Mechanical Turk is a web platform for posting tasks that require human input (provided by "Turkers" employed by Amazon) to complete. The task poster pays a fee. Previous research has shown that MTurk works well for music related tasks such as rating music similarity [16], [27] and providing tags [19]. Especially in the studies that compared the MTurk results with the results obtained from music experts for MIREX Audio Music Similarity task [16], and Symbolic Music Similarity task [27], they were able to obtain comparable results from MTurk with the ones from the music experts.

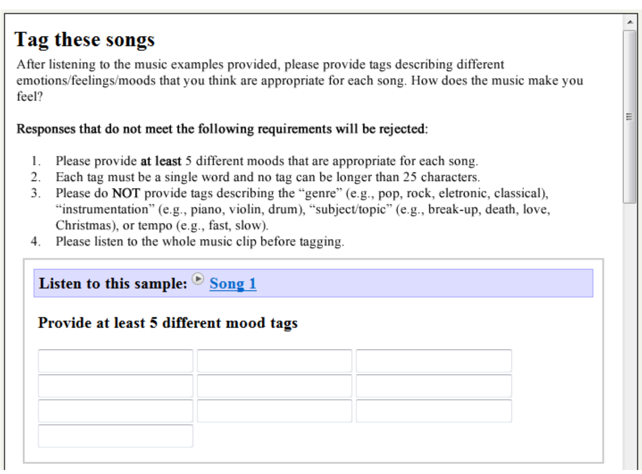

Figure 1. Screenshot of the MTurk HIT

Previous music information retrieval research on social tags [3], [4], [14], [17] mainly focused on automatically classifying music moods based on the user tags. In this study, we also attempt to provide a multidimensional and categorical model for representing the mood space; however, we conducted additional qualitative analyses of the mood tags in order to identify evidence of the effects of different factors of the music on how real users feel about music and mood.

2.2 Task Design

Users were instructed to listen to several 30 second music clips and assign a minimum of 5 unique tags to describe the mood of each song. Users provided tags for 68 songs in total, of which 39 were cover or karaoke versions of 12 select songs. The rest of the songs were included to test other aspects of music mood. Each song was tagged by 5 different users. Each "HIT" (how MTurk refers to a user task) consisted of 10 different songs. This meant that we needed 34 complete HITs in order to collect all the tags (i.e., (68*5)/10). Users were only given one version of each song; in other words, none of the users were assigned multiple versions of the same song. We also checked the worker ID in order to ensure that each HIT was done by separate workers, thus 34 in total. Instruction for the task was given as shown in Figure 1.

On MTurk, task requesters are allowed to review the submitted HITs before they accept them and reject any responses that do not meet the requirements set by the requester. Previous studies that used MTurk for music related tasks reported a high proportion of noise in the responses. For instance, Lee [16] reported that almost half of the responses needed to be filtered out by inserting verification questions in the task (e.g., inserting the same question twice to check the consistency of the answers). In our study, we specified several sets of rules: 1) users were not allowed to provide tags describing the genre (e.g., pop, rock, electronic, classical), 2) instrumentation (e.g., piano, violin, drum), 3) subject/topic (e.g., break-up, death, love, Christmas), or 4) tempo (e.g., fast, slow). Of the 46 submitted HITs, we rejected 12 responses (26%) that violated these rules.

2.3 Test Collection

We tested multiple cover versions of four Western songs and four Korean songs as shown in Table 1. Three to five cover versions with different instrumentation, genre and delivery were selected for each song.

Table 1. List of cover songs tested

Song Title	Artist	Genre or Instrument
Für Elise	AM Orchestra	Piano
	Rick Fogel	Hammered dulcimer
	Tim Lake	Banjo
Fly me to the moon	4 to the Bar	French pop
	Azz Izz Band	Ska
	Vox P	A cappella
	Roger Mason	Cont. classical guitar
	Seattle Womens Jazz Orchestra	Jazz
California dreamin'	The Mamas & The Papas	Original
	The Flashbulb	Techno
	Clare Teal	Piano ballad

	John Philips	Contemporary folk
Jingle bells	Andrew Burchett	Piano (instrumental)
	Margaret Tobolowska	Cello (instrumental)
	Jack Convery	Banjo (instrumental)
Jingle bells	Concino Children's Choir	Children's choir
	The Hidden Stars	Country
핑계 *(Excuse)*	Kim Kunmo	Original
	Kim Kunmo	House
	Various Artists	Trot
비처럼 음악처럼 *(Like rain, like music)*	Kim Hyunshik	Original (Male vocal)
	Koonta & Nuoliunce	Reggae
	Moony	Female vocal
	Takaoka Kenji	Saxophone (inst.)
그리움만 쌓이네 *(Yearning)*	Yeojin	Original
	Lazy Bone	Rock
	Various Artists	Dance
	Violet F	Piano (instrumental)
붉은 노을 *(Red sunset)*	Lee Munsae	Original
	Glassbox	Rock
	Big Bang	Dance

In addition, we also tested the original and karaoke versions (which consist of backing tracks only) of four additional songs to understand the effect of lyrics. The songs tested were: 1) *Paparazzi* – Lady GaGa, 2) *He'll have to go* – Jim Reeves, 3) *He ain't heavy, he's my brother* – The Hollies, and 4) *Any dream will do* – Jason Donovan.

3. DATA AND DISCUSSION

A total of 1778 tags were assigned for the 68 songs we tested by 46 users. After correcting typographical errors, there were a total of 551 unique tags. Users relied heavily on a small number of tags, resulting in a long tail distribution: of the 551 tags, only 36 tags appeared 10 times or more. This also meant that there were a large number of unique tags: 305 tags (55%) appeared only once.

Table 2. Tags that appeared 10 or more times

Tag	Count	Tag (continued)	Count
happy	79	*lonely*	16
energetic	41	*optimistic*	16
sad	40	*excited*	15
hopeful	31	*mellow*	15
relaxed	30	*nostalgic*	15
calm	29	*thoughtful*	14
cheerful	22	*bouncy*	12
joyful	21	*confused*	12
relaxing	21	*dark*	12
lively	20	*depressed*	12
romantic	20	*dreamy*	12
upbeat	20	*playful*	12
angry	19	*warm*	12
melancholy	19	*moving*	11
peaceful	18	*sleepy*	11
bored	17	*uplifting*	11
soothing	17	*vibrant*	11
Fun	16	*tired*	10

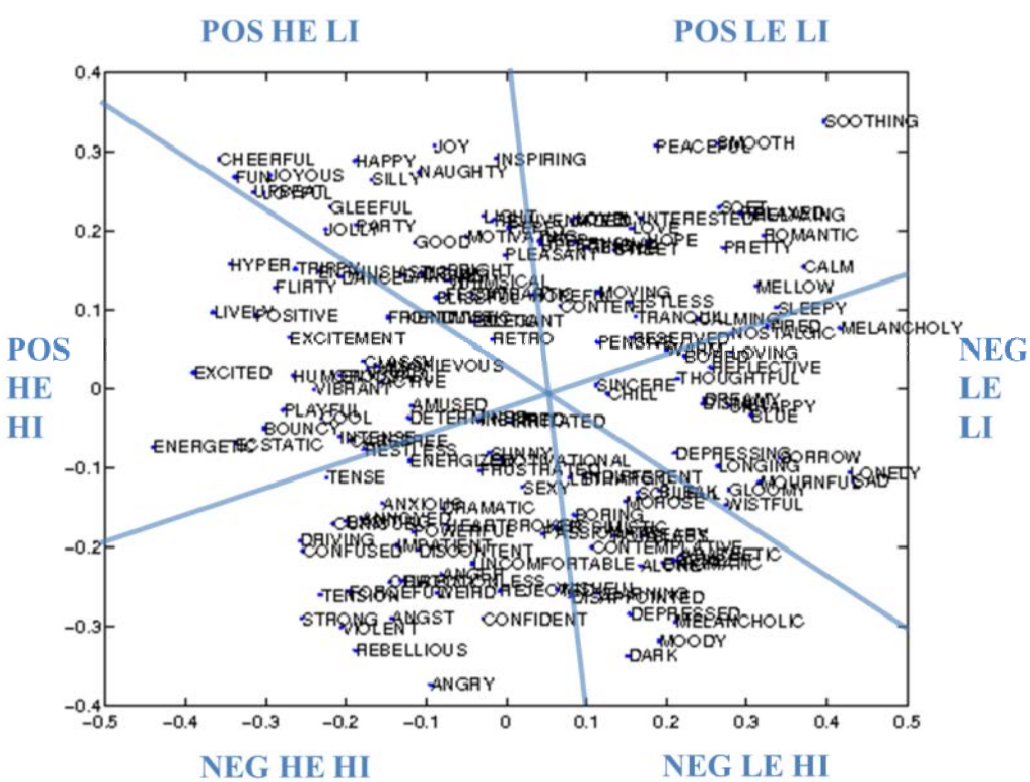

Figure 2. Distances of the moods based on co-occurrences

When we compared these user tags to an extensive set of 156 mood tags generated by "music experts" (i.e., editors and expert contributors with extensive knowledge on various musical styles) on allmusic.com, only 80 terms overlapped, meaning only about half of the expert generated mood tags were also provided by users. This indicates a significant discrepancy between the users' and experts' mood terms. Although there were a large number of unique mood terms, due to the limited space, we present only the tags that appeared 10 or more times in our dataset [Table 2].

Based on the assumption that the first mood tag users provide represents the perceived predominant mood, we looked at just the first tag for each song provided by each user. We were looking to see if there was a difference in the number and distribution of terms among first-tags. Among the 340 first-tag instances, there were 169 unique terms with a long tail distribution of frequencies. The top 10 tags and their frequencies were: *happy* (28), *sad* (19), *calm* (13), *energetic* (10), *romantic* (9), *relaxed* (8), *relaxing* (7), *angry* (6), *upbeat* (6), *bored* (5); under the tail, there were 112 tags with a frequency of one. The top 16 (~10%) most frequently used tags comprise 39.7% of all first-tag instances, which suggests that there is a small set of core mood terms that are used heavily.

In order to understand the relationships among these mood tags, we examined their co-occurrence patterns. We plotted the cosine distances of the mood tags in a 2-dimensional space using classic multidimensional scaling (see Figure 2). There appears to be some structure to the grouping of tags in the MDS plot. We identified three dimensions to explain the mood tag space:

1. Valence, ranging from Positive (POS) to Negative (NEG);
2. Energy, ranging from High Energy (HE) – Low Energy (LE); and
3. Intensity ranging from High Intensity (HI) – Low Intensity (LI).

Examples of moods representing different combinations of these three dimensions are shown in Figure 3.

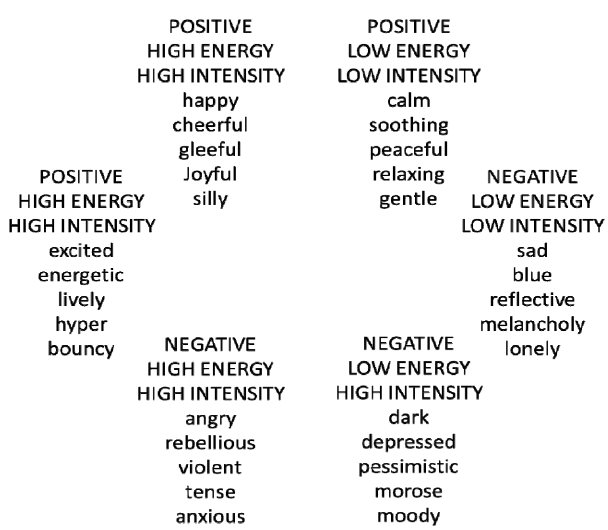

Figure 3. Examples of mood terms representing six clusters

In previous literature, several studies attempted to identify different categories or dimensions in the mood space. The first two dimensions we found are also present in the seminal work by Russell [24]. This model of valence and arousal was adopted in a number of studies on music mood [12], [18].

Wedin [28] conducted a principal components analysis on the music selections rated by his subjects on 125 unipolar adjective scales. The results indicated three dimensions: intensity-softness, pleasantness-unpleasantness, and solemnity-triviality. In Wedin's study, both intensity and energy were adjectives that were strongly associated with the intensity-softness dimension; however, we feel that these can be separated into two different dimensions similar to Thayer's: energy and tension [25].

Other studies have attempted to expand Russell's valence & arousal model, although the semantic nature of a third dimension is subject to disagreement [1], [13]. Third dimensions proposed in previous studies include: "tension" [6], "kinetics" [23], and "dominance" [21].

We have formulated a categorical model of six clusters with examples of moods in each cluster inspired by Hevner's [7] model consisting eight clusters of mood adjectives [Fig. 3]. Hevner's [7] model consists of eight clusters of mood adjectives [Fig. 4]. Our model bears some resemblance with Hevner's with regards to the positions of the clusters; the positive and negative mood clusters appearing in the opposite sides as well as the similarities of the neighboring clusters. There are some dissimilarities as well; we identified a single cluster which seems to combine the cluster 5 (*humorous, playful, whimsical, fanciful, quaint, sprightly, delicate, light, graceful*) and 6 (*merry, joyous, gay, happy, cheerful, bright*) in Hevner's model and we observed few labels representing the 1st (*spiritual, lofty, awe-inspiring, dignified, sacred, solemn, sober, serious*) and the 8th cluster (*vigorous, robust, emphatic, martial ponderous, majestic, exalting*). This may be due to the kinds of music we had in our test collection, thus further studies should be conducted to better understand the reason for the absence of these particular clusters.

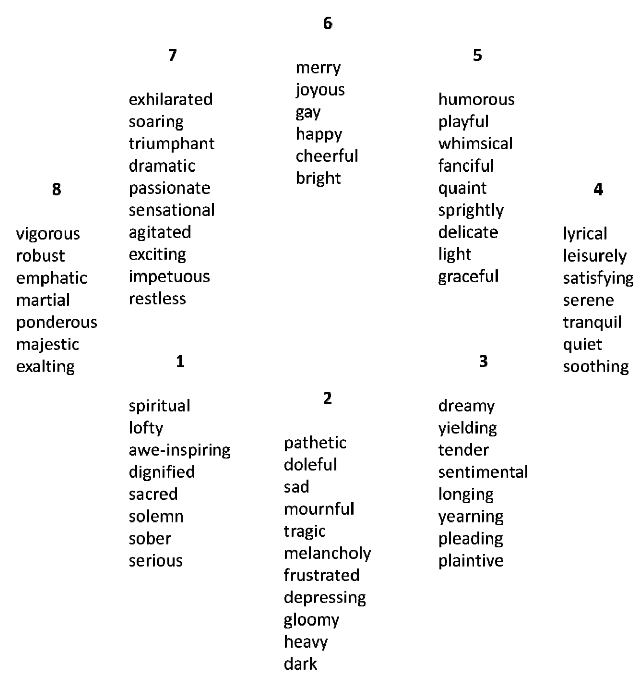

Figure 4. Hevner's adjective clusters

3.1 Factors Affecting Music Mood

To examine how much overlap exists among mood tags for different versions of the same song, we compared all user tags for the eight cover songs. As we compared the mood tags, we noticed that for each song a common set of moods emerged. Table 3 shows the number of unique tags and tags shared across at least two versions of the song. Overall, for most of the songs certain moods seem to carry over different cover versions. For instance, *happy* was found in 5 versions of *Jingle Bells*, *joyful* and *fun* in 4 versions. *Happy* and *relaxed* were found in 4 versions of *Fly me to the moon*, *dark* and *mellow* in 3 versions of *California dreamin'*, etc.

Table 3. Proportion of repeated tags in 2 or more versions

Song Title	Unique tags	Repeated tags	% of overlap
California dreamin'	71	14	19%
Jingle bells	83	16	19%
Fly me to the moon	88	16	18%
Like rain, like music	75	13	17%
Excuse	53	8	15%
Für Elise	63	8	13%
Yearning	83	8	10%
Red sunset	63	5	8%

There were definitely some noticeable differences as well in the kinds of tags each cover version received. As we closely examined the tags, we identified the following factors that seem to affect how people determine music mood.

3.1.1 Sources of Music Mood

The first point we want to raise is about different sources of music mood – *where does the mood lie?* Does it reside in the music itself, the performer, or the listener?

The fact that music mood stems from intrinsic and extrinsic sources has already been discussed in several works [8], [11], [22]. Intrinsic emotion is triggered by characteristics of music whereas extrinsic emotion derives from the semantic context outside of the music [8]. We also observed this difference in our results: for instance, compare *excited* vs. *exciting*; *annoyed* vs. *annoying; liberated* vs. *liberating; depressed* vs. *depressing; bored* vs. *boring; frightened* vs. *frightening*, etc. (e.g. a person is excited; music is exciting.)

There is another aspect to this issue, however: whose mood is actually being described? Whose perspective are the users taking when they provide these mood tags? When the user says *excited, annoyed,* or *liberated*, is it the artist/singer or the listener who is excited, annoyed, or liberated? We have noticed that listeners sometimes seem to describe what the music is making *them* feel, but in other cases, they describe the emotion they think *the singer(s)* is feeling. Take Lisa Loeb's *Stay*, for example. Some of the tags we received included: *confused, discontent, heartbroken, pouty, resentful, stressed, sulky,* and *strident*. Note that these moods strongly suggest the effects of lyrics, as it is difficult to imagine an instrumental piece carrying a pouty or strident mood. Additionally, these moods seem to be about how the singer feels about the break-up rather than how the listener is feeling. For the same song, we also found tags such as *amiable, dreamy, gentle*, and *warm*, which seem to represent the musical features. Some tags given for The Flashbulbs' *California dreamin'* can also be said to apply to the singer, who is almost whispering at some points in the song (e.g., *uncomfortable, expectant, scared, unsure*).

By way of comparison, the hammered dulcimer version of *Für Elise* received tags such as *disappointed* and *irritated*, and the banjo version of *Jingle bells* received the tags *irritated* and *annoyed*. In these cases, the listener seems to be assigning tags that convey his or her own feelings rather than the mood of the song or the artist.

Addressing this question is important as it has direct effects on how we should process, use, and interpret these kinds of data. If users are describing the mood of music from multiple sources and perspectives, perhaps this means, for example that we should be careful about using stemming, query-expansion, or other techniques which might obscure these differences.

3.1.2 Lyrics

Several tag sets suggest that lyrics do affect how people feel about music. For example, Lady Gaga's *Paparazzi* received 12 tags that have negative connotations (i.e., *angry, angst, depressed, empty, gloomy, lonely* (2), *paranoid, sad* (2), *somber, weird*). For the karaoke version of this song, however, most of the tags represented high energy and power, and there were no tags with a negative connotation. Another example is The Hollies' *He ain't heavy, he's my brother*. The version with lyrics received tags such as *altruistic, loyal, trustworthy*, and *concerned* that are closely associated with the lyrical content of the song. This contrasts with the karaoke version, where tags appeared to be more general without the input of lyrical content; for example*, happy, inspiring, hopeful* and *soothing* were used in lieu of the more specific tags previously discussed.

The instrumental version of *Fly me to the moon* received 13 negative tags (e.g., *bleak, depressing, painful, dismal, distressing, lonely, sad* (2), *unhappy* (2), *somber, sorrowful, woeful*) which was a much higher number compared to the other 4 versions; a cappella and jazz versions received no negative tags, french pop version received 1, and a ska version received 5. The lyrics in this song do carry a positive and romantic mood and the fact that they are missing in the instrumental version does seem to affect the types of tags received. For the Korean reggae song *핑계 (Excuse)* which is a monologue by a man about a break-up, most of the tags provided were positive (e.g., *relaxed* (3), *peaceful* (2), *happy* (2), *calm* (2), *cheerful* (2)) due to its melodic features.

We also noticed some mood terms imply semantic understanding of the song that would appear to stem from the lyrical content; for instance, *ironic, sarcastic, mocking, cynical, complaining,* etc. The effects of lyrics also seem to be highly dependent on how well the mood of melody and the lyrics match; for *He'll have to go* and *Any dream will do*, there did not seem to be obvious differences in the tags the original and the karaoke versions received.

3.1.3 Tempo

Of various musical features, tempo of the music seems to greatly affect users' mood with regards to valence. Early experiments in music psychology literature have also showed that swift tempo was the most important music elements for excitement [8]. When we tested cover versions of the same songs with varying tempi (e.g., contemporary classical guitar version of *Fly me to the moon*, piano ballad version of *California dreamin'*), the slower versions always received more tags representing negative feelings (e.g., *sad, depressing, lonely*). The faster versions such as *Red Sunset* by Glassbox or *Excuse* by Various Artists received more positive tags such as *energetic, upbeat, happy*. Table 4 shows, for instance, comparison of the tags provided for *California dreamin'* by The Mamas & The Papas with the version by Clare Teal, which has a much slower tempo than the original. The tags received for Clare Teal's version are mostly negative (in bold) compared to the mix of positive and negative tags for the original version.

Table 4. Comparison of two versions of *California dreamin'*

The Mamas & The Papas		Clare Teal	
alone	***mellow***	***blue***	***mellow***
cold (2)	***moody***	***concerned***	***mournful***
dark	*motivational*	***dark***	*moving*
energetic	***pessimistic***	***depressing (3)***	*romantic*
enigmatic	***sad***	***dismal***	***sad (4)***
exotic	***somber***	***down***	***somber***
hopeful (2)	*spirited*	***dreary***	***sorrow***
impulsive	*sunny*	***grief***	***sorry***
listless	*supple*	***lonely***	***uneasy***
lonely	*vibrant (2)*	***melancholy (2)***	***unhappy***
melancholic	*warm*		***uninspiring***
	wishful		

3.1.4 Instrumentation

Instrumentation also seems to affect the kinds of tags users provide. Rock music with distorted guitar almost always seems to receive tags such as *aggressive*, for instance. In this same vein, songs performed by a banjo often received positive tags such as *soothing* and *relaxing*. Tags associated with songs performed by saxophone, such as *Like rain, like music*, seem to be related to the instrumentation itself (e.g., *romantic, sexy, soft, sultry, sweet, tender, relaxing*) as they were unique to this particular version. The tags for the hammered dulcimer version of *Für Elise* also seem to be projecting a feeling onto the player of the instrument due to the sound and how the piece was played. For example, *frustrated* and *calculating* tags appear to reflect the physical necessities of playing an instrument by striking strings as well as the artistic license taken with the piece; longer pauses and strong emphasis on certain notes might have led users to provide tags such as *calculating*.

We also observed some interesting patterns with regards to the distances between different instruments based on co-occurrence patterns of the tags; for instance, the piano and hammered dulcimer versions of *Für Elise* shared 5 common tags (*moving, uplifting, sleepy, tired, bored*) whereas the banjo version only had 2 common tags with the piano version (*relaxing, calm*) and 1 common tag (*thoughtful*) with the hammered dulcimer version. For *Jingle Bell*, 5 tags were shared between the piano and banjo versions, 3 for cello and banjo versions, and only 1 for piano and cello versions. One may think that piano and cello are more similar to each other since they are both classical instruments; however, cello is typically not used for this song and thus seems to have elicited a different set of tags.

3.1.5 Genre

User tags that reflected genre showed some distinct patterns. For example, songs in the reggae genre were tagged with positive tags such as *happy, cheerful, relaxed, peaceful,* dance genre with tags such as *happy, energetic, upbeat, fun,* rock songs *aggressive, rebellious, energized,* and so on.

The effect of genre can be seen when we compare the tags received for the original ballad version and reggae version of *Like rain, like music*. The ballad version received a number of low-energy, negative tags such as *sad* (2), *melancholy/-ic* (2), *lonely, sorrow, gloomy, mournful, sleepy, relaxed,* etc. whereas the reggae version received many tags with positive sense (e.g., *happy* (3), *cheerful* (2), *hopeful* (2), *bright, sunny, energized, joy, motivational*).

3.1.6 Delivery

For *California dreamin'* the piano ballad version, which had a more languid vocal delivery, received more negative tags [Table 4] than the other versions. Beck's *Devil's haircut* received tags such as *drunk, hungover, exhausted, lethargic,* and *mischievous* that seem to represent his unique style of monotone delivery. Although our listeners most likely did not understand the Korean lyrics of *Yearning*, they still seem to be able to detect the singer's emotion based on her delivery of the song (*heartbroken, heartfelt, yearning, determined*). The male and female versions, however, did not seem to get very different sets of tags. For instance, the male and female versions of *Like rain, like music* shared 8 tags, and overall the tags seem to describe similar negative, low energy, low intensity moods.

3.1.7 Musical Expectancy

We also observed some tags reflecting reactions due to the users' musical expectancy. In previous literature [22] musical expectancy is used to refer to a process where an emotion is induced because "specific features of the music violates, delays, or confirms the listener's expectations about the continuation of the music." [11] For songs that were recorded using unconventional musical instruments, and cover versions of songs in unexpected genres, we observed tags that suggest the users' expectations were violated (e.g., *dorky* and *silly* for the country version of *Jingle Bells*, and *clever, disjointed*, and *random* for the cello version of the same song). Tags provided for the trot version of Excuse also reflect users' reaction to this relatively unknown genre of music outside of Korea (e.g., *odd, strange, uncool, fresh*). Although musical expectancy has been discussed mostly based on sequential progression of music, it may be applied more broadly to explain how listeners respond to music when their musical expectations for familiar songs are disrupted with regards to instrumentation or genre.

3.1.8 Cultural Context

Jingle bells received tags from users that appeared to have strong cultural connotations. Strong overlap in positive tags (e.g., *happy, joyful, fun*) was observed across all versions. Tags such as *jolly* and *festive* were also prevalent and were used almost exclusively for this particular set of songs. This may be considered as a type of evaluative conditioning, a process whereby an emotion is induced by some music simply because this stimulus has been paired repeatedly with other stimuli [11]. The tags received for the original version of *California dreamin'* (e.g., *spirited, sunny, supple, vibrant, warm, wishful, hopeful, motivational*) which were unique to this version are also intriguing as they seem to reflect the cultural context (i.e., hippie culture). This poses an interesting question regarding how a song's cultural context may reflect user mood, and introduces important questions regarding how users from different cultures and countries may assign mood tags.

3.1.9 Social Expectation

Some tags appeared to indicate that people may feel obligated to provide certain kinds of tags in some cases. For instance, we observed no negative tags for the version of *Jingle Bells* which was sung by a children's choir. Similarly, the original version of *Any dream will do* (which features a children's chorus) had associated tags such as *cute* and *precious* that users may feel obligated to provide. These types of tags were non-existent in the karaoke version, which did not include the chorus in the backing track.

3.1.10 Personal Preference

The listeners' personal preferences of music might also affect how they describe the mood of the music. For example, Marvin Gaye's *What's going on* received tags such as *boring, vacuous,* and *vapid*. These strongly negative tags do not appear to reflect the song's tone or intent, and therefore could be hypothesized to reflect the user's potential dislike of the type of music as a whole.

3.1.11 Familiarity

We observed that there are usually at least one or two moods that are shared among all the versions. We initially suspected that this is because people are already familiar with the original song. However, we see examples suggesting otherwise. For instance, *Für Elise* is a well-known song but there were not a lot of overlap among the tags received for the three different versions. Conversely, our users most likely did not understand Korean lyrics, but the tags we received for different versions of the Korean songs do show some overlaps.

4. CONCLUSION AND FUTURE WORK

While users generate many unique mood tags, they also frequently use a small number of fairly generic moods (such as *happy, energetic,* and *sad*). Outside of these simple mood terms, the vocabulary of music mood is highly varied among listeners, reflecting the fact that mood is about an individual's relationship with a given piece of music. In some instances, we observed that moods can carry across different versions of a song, but differences in sources of mood, lyrics, tempo, instrumentation, genre, delivery, musical expectancy, cultural context, social expectancy, personal preference and familiarity all affect how people feel about music. Ultimately, mood is a highly subjective feature for describing and organizing music, perhaps more akin to user ratings than genre labels.

Our results suggest that we may be able to reach high agreement on mood tagging if we limited the tags to a small number of generic moods. However, the potential usefulness of music mood for organizing and accessing music will be significantly reduced. On the other hand, if we permit a larger number of mood tags to be used to organize music collections, it will be much more difficult to reach agreement on tags across users, thus making it too subjective of an element to be used as an effective organizational method. This may mean that music mood is best exploited when it is used in conjunction with other means for classification such as genre, musical styles, etc. rather than being used by itself.

Perhaps most importantly, the results from this study helped us to understand which factors to investigate further in determining from where mood is derived in the music-listening experience. The results clearly suggest that the tags we are seeing are affected by the combination of multiple factors. In order to understand how these factors influence perceptions of mood, we plan to conduct follow-up studies, including more listening/tagging studies but providing users with a limited number of mood tags to choose from rather than expressing music moods in their own terms in order to reduce the subjectivity in the results. Additionally, we want do conduct user interviews in order to address extrinsic factors like culture, language, familiarity, prior musical knowledge, etc. One limitation of this paper is that we are attempting to identify factors affecting music mood from limited evidence we can find from the tags provided by users. An in-depth interview with users will help build a more solid case on how each of these factors affect the users determine the mood, and how users prioritize the different factors. We are also interested in looking at how users respond to music containing conflicting moods across different factors (e.g., depressing lyrics with a happy melody).

5. REFERENCES

[1] Barrington, L., Turnbull, D., O'Malley, and Lanckriet, G. 2009. User-centered Design of a Social Game to Tag Music. *ACM KDD (Knowledge Discovery and Data Mining) Workshop on Human Computation.*

[2] Bigand, E. 2005. Multidimensional Scaling of Emotional Responses to Music: The Effect of Musical Expertise and of the Duration of the Excerpts. *Cognition and Emotion*, 19, 8, 1113-1139.

[3] Bischoff, K., Firan, C. S., Nejdl, W., and Paiu, R. 2009. How Do You Feel About "Dancing Queen"?: Deriving Mood & Theme Annotations From User Tags. *Proceedings of the JCDL (Joint Conference on Digital Libraries),* 285–294.

[4] Bischoff, K., Firan, C. S., Paiu, R., Dejdl, W., Laurier, C., and Sordo, M. Music Mood and Theme Classification: A Hybrid Approach. *Proceedings of the 10th ISMIR (International Society for Music Information Retrieval) conference,* 657-662.

[5] Dan, Y. and Lee, W. 2009. Music Emotion Identification from Lyrics. *2009 11th IEEE International Symposium on Multimedia,* 624-629.

[6] Eerola, T, Lartillot, O., and Toivianen, P. 2009. Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models. *Proceedings of the 10th ISMIR (International Society for Music Information Retrieval) conference,*621-626.

[7] Hevner, K. 1936. Experimental Studies of the Elements of Expression in Music. *American Journal of Psychology*, 48, 2, 246-268.

[8] Hu, X. 2010. Music and Mood: Where Theory and Reality Meet. *Proceedings of iConference.*

[9] Hu, X., and Downie, J. S. 2007. Exploring Mood Metadata: Relationships with Genre, Artist, and Usage Metadata. *Proceedings of the 8th ISMIR (International Society for Music Information Retrieval) conference,* 67–72.

[10] Hu, Y., Chen, X., and Yang, D. 2009. Lyric-based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method. *Proceedings of the 10th ISMIR (International Society for Music Information Retrieval) conference,* 123-128.

[11] Juslin, P. N. and Vastfjall, D. 2008. Emotional Responses to Music: The Need to Consider Underlying Mechanisms. *Behavioral & Brain Sciences,* 31, 559-621.

[12] Kim, Y., Schmidt, E., and Emelle, L. 2008. Moodswings: A Collaborative Game for Music Mood Label Collection. *Proceedings of the 9th ISMIR (International Society for Music Information Retrieval) conference,* 231-236.

[13] Kim, Y., Schmidt, E., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A. and Turnbull, D. 2010. Music Emotion Recognition: A State of the Art Review. *Proceedings of the 11th ISMIR (International Society for Music Information Retrieval) conference*, 255-266.

[14] Laurier, C., Sordo, M., Serra, J., and Herrera, P. 2009. Music Mood Representations from Social Tags. *Proceedings of the 10th ISMIR (International Society for Music Information Retrieval) conference,* 381-386.

[15] Law, E. L. M., Von Ahn, L, Dannenberg, R. B., and Crawford, M. 2007. TagATune: A Game for Music and Sound Annotation. *Proceedings of the 8th ISMIR (International Society for Music Information Retrieval) conference,* 361-364.

[16] Lee, J. H. 2010. Crowdsourcing Music Similarity Judgments Using Mechanical Turk. *Proceedings of the 11th ISMIR (International Society for Music Information Retrieval) conference,* 183-188.

[17] Levy, M. and Sandler, M. 2007. A Semantic Space for Music Derived From Social Tags. *Proceedings of the 8th ISMIR (International Society for Music Information Retrieval) conference,* 411-416.

[18] Lu, L., Liu, D., and Zhang, H. 2006. Automatic Mood Detection and Tracking of Music Audio Signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14, 1, 5-18.

[19] Mandel, M., Eck, D, and Bengio, Y. 2010. Learning Tags That Vary Within a Song. *Proceedings of the 11th ISMIR (International Society for Music Information Retrieval) conference*, 399-404.

[20] Mandel, M. I., and Ellis, D. P. W. 2007. A Web-based game for collecting music metadata. *Proceedings of the 8th ISMIR (International Society for Music Information Retrieval) conference,* 365-366.

[21] Meharabian, A. and Russell, J. A. 1974. *An Approach to Environmental Psychology.* MIT Press.

[22] Meyer, L. 1956. *Emotion and Meaning in Music.* University of Chicago Press, Chicago.

[23] Mion, L., and Poli, G. D. 2008. Score-independent Audio Features for Description of Music Expression. *IEEE Transactions on Audio, Speech and Language Processing,* 16, 2, 458-466.

[24] Russell, J. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39, 1161-1178.

[25] Thayer, R. E. 1989. *The Biopsychology of Mood and Arousal.* Oxford University Press.

[26] Turnbull, D., Liu, R., Barrington, L. and Lanckriet, G. 2007. A Game-based Approach for Collecting Semantic Annotations of Music. *Proceedings of the 8th ISMIR (International Society for Music Information Retrieval) conference,* 535-538.

[27] Urbano, J. Morato, J., Marrero, M., and Martin, D. 2010. Crowdsourcing preference judgments for evaluation of music similarity tasks. *Proceedings of the SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation,* 9-16.

[28] Wedin, L. 1972. A Multidimensional Study of Perceptual-Emotional Qualities in Music. *Scandinavian Journal of Psychology*, 13, 241-257.

[29] Xia, Y., Wang, L., Wong, K., and Xu, M. 2008. Sentiment Vector Space Model for Lyric-based Songs Sentiment Classification. *Proceedings of the Association for Computational Linguistics.* Columbus, Ohio. 133-136.