EXPLORING MOOD METADATA: RELATIONSHIPS WITH GENRE, ARTIST AND USAGE METADATA

Xiao Hu **J. Stephen Downie**

International Music Information Retrieval Systems Evaluation Laboratory
The Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
{xiaohu, jdownie}@uiuc.edu

ABSTRACT

There is a growing interest in developing and then evaluating Music Information Retrieval (MIR) systems that can provide automated access to the mood dimension of music. Mood as a music access feature, however, is not well understood in that the terms used to describe it are not standardized and their application can be highly idiosyncratic. To better understand how we might develop methods for comprehensively developing and formally evaluating useful automated mood access techniques, we explore the relationships that mood has with genre, artist and usage metadata. Statistical analyses of term interactions across three metadata collections (`AllMusicGuide.com`, `epinions.com` and `Last.fm`) reveal important consistencies within the genre-mood and artist-mood relationships. These consistencies lead us to recommend a cluster-based approach that overcomes specific term-related problems by creating a relatively small set of data-derived "mood spaces" that could form the ground-truth for a proposed MIREX "Automated Mood Classification" task.

1 INTRODUCTION

1.1 Music Moods and MIR Development

In music psychology and education, the emotional component of music has been recognized as the most strongly associated with music expressivity [6]. Music information behaviour studies (e.g., [10]) have also identified music mood as an important criterion used by people in music seeking and organization. Several experiments have been conducted to classify music by mood (e.g., [7][8][9]). However, a consistent and comprehensive understanding of the implications, opportunities and impacts of music mood as both metadata and content-based access points still eludes the MIR community. Since mood is a very subjective notion, there has yet to emerge a generally accepted mood taxonomy that is used within the MIR research and development community. For example, each of aforementioned studies used different mood categories, making meaningful comparisons between them difficult.

Notwithstanding that there is a growing interest in tackling mood issues in the MIR community--as evidenced by the ongoing discussions to establish a "Audio Mood Classification" (AMC) task at the Music Information Retrieval Evaluation eXchange (MIREX)[1] [3], this lack of common understanding is inhibiting progress in developing and evaluating mood-related access mechanisms. In fact, it was the MIREX discussions that inspired this study. Thus, this paper is intended to contribute our general understanding of music mood issues by formally exploring the relationships between: 1) mood and genre; 2) mood and artist; and, 3) mood and recommended usage (see below). It is also intended to contribute more specifically to the MIREX community by providing recommendations on how to proceed in constructing a possible method for conducting an "AMC" task.

Our primary dataset is derived from metadata found within the `AllMusicGuide.com` (AMG) site, a popular music database that provides professional reviews and metadata for albums, songs and artists. Secondary data sets were derived from `epinions.com` and `Last.fm`, themselves both popular music information services. The fact that real world users engage with these services allows us to ground our analyses and conclusions within realistic social contexts of music seeking and consumption.

In a previous study [5], we examined a relatively novel music metadata type: "recommended usage". We explored the relationships between usages and genres as well as usages and artists using a set of 11 user recommended usages provided by `epinons.com`, a website specializing in product reviews written by customers. Because both music moods and usages involve subjective reflections on music, they can vary greatly both among, and within, individuals. It is therefore interesting to see whether there is any stable relationship between these two metadata types. We explore this question by examining the set of albums common to the AMG mood dataset and our `epinions.com` usage dataset [5].

The rest of the paper is organized as follows: Section 2 describes how we derived the mood categories used in the analyses. Sampling and testing method is described in Section 3. Sections 4 to 6 report analyses of the relationships between mood and genre, artist and usage respectively. In Section 7, the results from Sections 4-6

[1] http://music-ir.org/mirexwiki

undergo a corroboration analysis using an independent dataset from `Last.fm`. Section 8 concludes the paper and provides recommendations for a possible MIREX "Audio Mood Classification" task.

2 MOOD CATEGORIES

2.1 Mood Labels on AMG

AMG claims to be "the most comprehensive music reference source on the planet"[1] and supports access to music information by mood label. There are 179 mood labels in AMG where moods are defined as "adjectives that describe the sound and feel of a song, album, or overall body of work"[2] and include such terms as "happy", "sad", "aggressive", "stylish", "cheerful", etc. These mood labels are created and assigned to music works by professional editors. Each mood label has its own list of representative "Top Albums" and its own list of "Top Songs". The distribution of albums and songs across these mood lists is very uneven. Some moods are associated with more than 100 albums and songs while others have as few as 3 albums or songs. This creates a data sparseness problem when analysing all 179 mood labels. To alleviate this problem, we designed three alternative AMG datasets:

1. *Whole Set*: Comprises the entire 179 AMG mood label set. Its "Top Album" lists include 7134 album-mood pairs. Its "Top Song" lists include 8288 song-mood pairs.
2. *Popular Set*: Comprises those moods associated with more than 50 albums and 50 songs. This resulted in 40 mood labels and 2748 album-mood and 3260 song-mood pairs.
3. *Cluster Set*: Many albums and songs appear in multiple mood label lists. This overlap can be exploited to group similar mood labels into several mood clusters. Clustering condenses the data distribution and gives us a more concise, higher-level view of the mood "space". The set of albums and songs assigned to the mood labels in the mood clusters forms our third dataset (described below).

2.2 Mood Clustering on Top Albums and Top Songs

In order to obtain robust and more meaningful clustering results, it is advantageous to use more than one view of the available data. The AMG dataset provides two views: "Top Albums" and "Top Songs". Thus, we performed the following clustering methods independently on both the "Top Albums" and the "Top Songs" mood list data of the *Popular Set*.

First, a co-occurrence matrix was formed such that each cell of the matrix was the number of albums (or songs) shared by two of the 40 "popular" mood labels specified by the coordinates of the cell. Pearson's correlation was calculated for each pair of rows (or columns) as the similarity measure between each pair of mood labels. Second, an agglomerative hierarchical clustering procedure using Ward's criterion [1] was applied to the similarity data. Third, the resultant two cluster sets (derived from album-mood and song-mood pairs respectively) were examined and found to have 29 mood labels out of the original 40 that were consistently grouped into 5 clusters at a similar distance level. Table 1 presents the resultant 5 mood clusters along with their constituent mood terms ranked by the number of associated albums.

Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Rowdy	Amiable/ Good natured	Literate	Witty	Volatile
Rousing		Wistful	Humorous	Fiery
Confident	Sweet	Bittersweet	Whimsical	Visceral
Boisterous	Fun	Autumnal	Wry	Aggressive
Passionate	Rollicking	Brooding	Campy	Tense/anxious
	Cheerful	Poignant	Quirky	Intense
			Silly	

Table 1. *Popular Set* mood label clustering results

Note the high level of synonymy within each cluster and the low level of synonymy across the clusters. This state of affairs suggests that the clusters are both reasonable and potentially useful. The high level of synonymy found within each cluster helps to define and clarify the nature of the mood being captured better than a single term label could (i.e., lessens ambiguity). For this reason, we are NOT going to assign a term label to any of these clusters in order to stress that the "mood spaces" associated with each cluster is really the aggregation of the mood terms represented within each column.

3 SAMPLING AND TESTING METHOD

In each of the following sections, we analyse the relationship of mood to genre, artist and usage using our three datasets. We focus on the "Top Album" lists from each of these sets rather than their "Top Song" lists because the album is the unit of analysis on `epinions.com` to which we will turn in Section 6 when looking at usage-mood interactions.

At the heads of Sections 4-6, you will find information about the specific (and slightly varying) sampling methods used for each of the relationships explored. In general, the procedure is one of gathering up the albums associated with a set of mood labels and their genre, artist or usage information and then counting the number of [genre|artist|usage]-mood label pairs that occur for each album. The overall sample space is the total number of [genre|artist|usage]-mood label pairs across all relevant albums.

To test for significant [genre|artist|usage]-mood label pairs, we chose the Fisher's Exact Test (FET) [2]. FET is used to examine the significance of the association/dependency between two variables (in our case [genre|artist|usage]-mood), regardless of whether the sample sizes are small, or the data are very unequally distributed. All of our significance tests were performed using FET.

[1] `AllMusicGuide.com`: "About Us".
[2] `AllMusicGuide.com`: "Site Glossary".

4 MUSIC MOODS AND GENRES

Each album in each individual "Top Album" list is associated with only one genre label. However, an album can be assigned to multiple "Top Album" mood lists. Thus, our genre-mood sample space is all existing combinations of genre and mood labels with each sample being the pairing of one genre and one mood label.

4.1 All Moods and Genres

There are 3903 unique albums in 22 genres in the *Whole Set*. This set contains 7134 genre-mood pairs, but their distribution across the 22 genres is very skewed with 4564 of them involving the "Rock" genre. In order to compensate for this "Rock" bias, we conducted our association tests on the whole dataset as well as on a dataset excluding Rock albums. Table 2 shows the basic statistics of the two datasets. The mood labels "Hungry", "Snide" and "Sugary" were exclusively involved with "Rock" which resulted in a "non-Rock" mood set of 176 labels.

	Samples	Moods	Genres	Unique Albums
+Rock	7134	179	22	3903
- Rock	2570	176	21	1715

Table 2. *Whole Set* counts (+/- Rock genre)

The FET results on the *Whole Set* with "Rock" albums gives 262 genre-mood pairs whose associations are significant at $p < 0.05$. Analysis of the "non-Rock" subset yielded 205 significant genre-mood pairs. 170 of these pairs are significant in both subsets and involve 17 genres. Table 3 presents these 17 genres and the top-ranked (by frequency) associated moods.

Genre	Mood	#	Genre	Mood	#
R & B	Sensual	51	Folk	Earnest	8
Rap	Street Smart	29	Latin	Spicy	5
Jazz	Fiery	28	World	Hypnotic	4
Electronica	Hypnotic	20	Reggae	Outraged	3
Blues	Gritty	16	Soundtrack	Atmospheric	3
Vocal	Sentimental	15	Easy Listening	Soothing	2
Country	Sentimental	15	New Age	Soothing	2
Gospel	Spiritual	11	Avant-Garde	Cold	3
Comedy	Silly	8			

Table 3. *Whole Set* top-ranked genre-mood pairs

While it is interesting to note the reasonableness of these significant pairings, it is more important to note that each genre is associated with 10 significant moods on average and that the mood labels cut across the genre categories. This is strong evidence that genre and mood are independent of each other and that both provide different modes of access to music items.

4.2 Popular Moods and Genres

The 40 mood labels in the *Popular Set* involve 2748 genre-mood pairs. Again, many of the pairs are in the "Rock" genre, and thus we performed FET on both sets with and without "Rock". Table 4 presents the statistics of the two sets. There are 70 genre-mood pairs with significant relations at $p < 0.05$ in the "with Rock" set and 54 pairs in the "non-Rock" set. 41 pairs involving 16 genres are significant in both sets. Table 5 presents the top (by frequency) 16 genre-mood pairs.

	Samples	Moods	Genres	Unique albums
+ Rock	2748	40	21	1900
- Rock	927	40	20	714

Table 4. *Popular Set* counts (+/- Rock genre)

Genre	Mood	#	Genre	Mood	#
R & B	Sensual	51	Electronica	Fun	6
Jazz	Fiery	28	Gospel	Joyous	5
Vocal	Sentimental	15	Latin	Rousing	5
Country	Sentimental	15	Soundtrack	Theatrical	3
Rap	Witty	14	Reggae	Druggy	3
Comedy	Silly	8	World	Confident	2
Blues	Rollicking	8	Easy Listening	Fun	2
Folk	Wistful	8	Avant-Garde	Volatile	2

Table 5. *Popular Set* top-ranked genre-mood pairs

Because of the exclusion of less popular moods, some genres are shown to be significantly related to different moods than those presented in Table 3 (e.g., "Blues", "Electronic", "Rap", "Gospel", etc.). Note that these term changes are not contradictory but rather are suggestive of an added dimension to describing a more general "mood space". For example, in the case of "Folk" the two significant mood terms are "Earnest" and "Wistful". Similarly, the combination of "Joyous" and "Spiritual" mood terms better describes "Gospel" than either term alone. See also "Latin" ("Spicy", "Rousing") and "Reggae" ("Outraged", "Druggy").

4.3 Mood Clusters and Genres

In the *Cluster Set*, there are 1991 genre-mood cluster combinations, covering 20 genres. Among them, "Rock" albums again occupy a large portion of samples, and thus we made an additional "non-Rock" subset (Table 6). The FET significant results (at $p < 0.05$) on the "with Rock" set contain 20 genre-mood pairs and those on the "non-Rock" set contain 15 pairs. "Rock" was significantly related to Cluster 4 and 5 at $p < 0.001$. The 14 pairs significant in both sets are shown in Table 7.

	Samples	Clusters	Genres	Unique Albums
+Rock	1991	5	20	1446
- Rock	619	5	19	507

Table 6. *Cluster Set* counts (+/- Rock genre)

Genre	Mood	#	Genre	Mood	#
R & B	Cluster1	71	Vocal	Cluster3	18
Jazz	Cluster5	57	Vocal	Cluster2	17
Rap	Cluster4	32	Comedy	Cluster4	12
Rap	Cluster5	30	Latin	Cluster1	7
Folk	Cluster3	28	World	Cluster1	6
Country	Cluster3	24	Avant-Garde	Cluster5	4
Blues	Cluster1	20	Easy Listening	Cluster2	4

Table 7. *Cluster Set* top-ranked genre-mood pairs

It is noteworthy that “R&B” and “Blues” are both associated with Cluster1 which might reflect their common heritage. Similarly, “Country” and “Folk” are both associated with Cluster3.

5 MUSIC MOODS AND ARTISTS

Each album on AMG has a “Title” and an “Artist” field. For albums combining tracks by multiple artists, the “Artist” field is filled with “Various Artists”. In the following analyses, we eliminated “Various Artists” as this label does not signify a unique analytic unit.

5.1 All Moods and Artists

There are 2091 unique artists in our *Whole Set*. Some artists contribute as many as over 30 artist-mood pairs each while 871 artists only occur once in the dataset and thus each of them only relates to one mood. We limited this analysis to artists who have at least 10 artist-mood pairs, which gave us 142 artists, 175 mood labels and 2241 artist-mood pairs. There are 623 significant artist-mood pairs at $p < 0.05$. Table 8 presents the top 14 (by frequency) pair associations. Those familiar with these artists will find these results reasonable.

Artist	Mood	Artist	Mood
David Bowie	Theatrical	The Grateful Dead	Trippy
Wire	Fractured	The Small Faces	Whimsical
Wire	Cold	Randy Newman	Cynical/Sarcastic
T. Rex	Campy	Randy Newman	Literate
The Beatles	Whimsical	Miles Davis	Uncompromising
The Kinks	Witty	Thelonious Monk	Quirky
Brian Eno	Detached	Talking Heads	Literate

Table 8. *Whole Set* top significant artist-mood pairs

5.2 Popular Moods and Artists

The *Popular Set* contains 1142 unique artists. 29 of them appear in at least 9 artist-mood pairs, and together contribute 372 artist-mood pairs that form the testing sample space. The results contain 68 significantly associated artist-mood pairs at $p < 0.05$. Table 9 presents the top 16 (by frequency) pair associations.

Artist	Mood	Artist	Mood
David Bowie	Theatrical	The Small Faces	Whimsical
David Bowie	Campy	The Small Faces	Trippy
Talking Heads	Wry	Randy Newman	Literate
Talking Heads	Literate	Randy Newman	Cynical/Sarcastic
The Beatles	Whimsical	Hüsker Dü	Fiery
The Beatles	Trippy	The Jesus & Mary Chain	Tense/Anxious
Elton John	Wistful		
T. Rex	Campy	The Velvet Underground	Literate
The_Kinks	Witty		

Table 9. *Popular Set* top significant artist-mood pairs

Like we discussed in Section 4.2, it is important to note in Tables 8 and 9 the application of multiple significant terms to individual artists. For example, Randy Newman is associated with “Cynical/Sarcastic” and “Literate” and *Wire* is associated with “Fractured” and “Cold”. Again, we see that it is the “sum” of these mood terms that evokes a more robust sense of the general mood evoked by these artists.

5.3 Mood Clusters and Artists

The *Cluster Set* contains albums by 920 unique artists. Among them, 24 artists who have no less than 8 artist-mood pairs form a testing space of 248 artist-mood pairs. Table 10 presents the 17 significant artist-mood cluster associations at $p < 0.05$.

Artist	Mood	#	Artist	Mood	#
The Kinks	Cluster4	13	Miles Davis	Cluster5	7
Hüsker Dü	Cluster5	12	Leonard Cohen	Cluster3	7
XTC	Cluster4	9	Paul Simon	Cluster3	7
Bob Dylan	Cluster3	9	John Coltrane w/ Johnny Hartma	Cluster3	6
Elvis Presley	Cluster1	8			
Elton John	Cluster3	8	David Bowie	Cluster4	6
Harry Nilsson	Cluster4	8	The Beatles	Cluster2	4
The Who	Cluster5	8	The Beach Boys	Cluster2	4
X	Cluster5	7	Nick_Lowe	Cluster2	4

Table 10. *Cluster Set* significant artist-mood pairs

The associations presented in Table 10 are again quite reasonable. For example, *The Beatles* and *The Beach Boys* are both related to Cluster2. The four artists related to Cluster5 are all famous for their “uncompromising” styles. It is noteworthy that Cluster5 members represent both the “Rock” (e.g., *Hüsker Dü*) and “Jazz” (Miles Davis) genres further indicating the independence of genre and mood to describe music. Similarly, Cluster3’s members of John, Cohen, Coltrane, and Simon also cut across genres.

6 MUSIC MOODS AND USAGES

In each of the user-generated reviews of music CDs presented on `epinions.com`, there is a field called “Great Music to Play While” where the reviewer selects a usage suggestion for the reviewed piece from a ready-made list of recommended usages prepared by the editors. Each album (CD) can have multiple reviews but each review can be associated with at most one recommended usage. Hu et al. [5] identified interesting relations between the recommended usage labels and music genres and artists as well as relations among the usages themselves. In this section, we explore possible relations between mood and usage. The following usage-mood analyses are based on intersections between our three AMG datasets and our earlier `epinions.com` dataset which contains 2800 unique albums and 5691 album-usage combinations [5].

6.1 All Moods and Usages

By matching the title and artist name of each album in our *Whole Set* and the `epinions.com` dataset, 149 albums were found common to both sets. As each album may have more than one mood label and more than one usage label, we count each combination of existing mood and usage labels of each album as one usage-mood sample. There were 1440 usage-mood samples involving 140 mood labels. 64 significant usage-mood pairs are identified by FET at $p < 0.05$. Table 11 presents the most frequent usage-mood associations for each of the 11 usage categories [1].

Usage	Mood	#	Artist	Mood	#
Go to sleep	Bittersweet	12	Hang w/friends	Fierce	5
Driving	Menacing	11	Waking up	Cathartic	4
Listening	Epic	9	Exercising	Angry	4
Reading	Provocative	7	At work	Menacing	3
Go out	Party/Celebratory	5	House clean	Carefree	2
Romancing	Delicate	5			

Table 11. *Whole Set* top significant usage-mood pairs

6.2 Popular Moods and Usages

There are 84 common albums in the *Popular Set* and the `epinions.com` dataset, which yields 527 usage-mood pairs. There are 16 pairs with 7 usages identified as significant at $p < 0.05$. Table 12 presents the most frequent usage-mood associations for each of the usage categories**.**

Usage	Mood	#	Artist	Mood	#
Go to sleep	Bittersweet	12	Go out	Fun	5
Driving	Visceral	7	Exercising	Volatile	3
Listening	Theatrical	7	House clean	Sexy	2
Romancing	Sensual	5			

Table 12. *Popular Set* top significant usage-mood pairs

6.3 Mood Clusters and Usages

There are 66 albums included in both the *Cluster Set* and the `epinions.com` dataset, yielding 358 usage-mood pairs. Table 13 presents the 6 significant pairs ($p < 0.05$).

Usage	Mood	#	Usage	Mood	#
Go to sleep	Cluster3	44	Romancing	Cluster3	17
Driving	Cluster5	20	Exercising	Cluster5	13
Hang w/friends	Cluster4	19	Go out	Cluster2	6

Table 13. *Cluster Set* significant usage-mood pairs

The usage-mood relationship appears to be much less stable than the genre-mood and artist-mood relationships. Only 6 of the 11 usages have significant cluster relationships. We believe this instability is a result of the specific terms and phrases used to denote the usage activities (also see Section 7.3).

7 EXTERNAL CORROBORATION

It is always desirable to analyse multiple independent data sources whenever conducting analyses of relationships. In this section we take our relationship findings from Sections 4-6 and attempt to re-find them using sets of data from `Last.fm`. Note that we are only looking for corroboration, not definite "proof" whether the AMG findings are "true" or "false". That is, we are exploring the `Last.fm` data sets to see whether, or not, our approach is sound and whether it merits further development.

[1] Usage labels modified for space reasons. See [5] for original labels.

`Last.fm` is a website collecting music related information from the general public, including playlists, and variety of tags associated with albums, tracks and artists, etc. The `Last.fm` tag set includes genre-related, mood-related and sometimes usage-related tags that can be used to analyse genre-mood, artist-mood and usage-mood relationships.

7.1 Corroboration of Mood and Genre Associations

`Last.fm` provides webservices [2] through which the general public can obtain lists of "Top Tracks", "Top Albums" and "Top Artists" for each user tag. As we are interested in corroborating the significance of the genre-mood pairs uncovered in the AMG datasets, we obtained the 3 `Last.fm` "top lists" for tags named by the genre-mood pairs shown in Tables 3 and 5. From these lists, we constructed three sample sets by collecting albums, tracks and artists with at least one genre tag and one mood tag. The three sample sets present three different "views" with regard to the associations between genre and mood. A FET was performed on each of the three sample sets. 21 of the 28 significant pairs presented in Tables 3 and 5 are also significantly associated in at least one of the `Last.fm` sample sets ($p < 0.05$). The 7 non-corroborated pairs are: "Electronica"–"Fun", "Latin"–"Rousing", "Reggae"–"Druggy", "Reggae"–"Outraged", "Jazz"–"Fiery", "Rap"–"Street Smart", and "World"–"Hypnotic".

The same method was applied to the corroboration of genre-mood cluster pairs. 12 of the 14 pairs in Table 7 tested to be significantly associated at $p < 0.05$. The 2 non-corroborated pairs are: "Jazz"–Cluster5 and "Latin"–Cluster1.

7.2 Corroboration of Mood and Artist Associations

`Last.fm` provides a "Top Artists" list for each user tag and a "Top Tags" list for each artist in its system. We retrieved the "Top Artists" list for each of the mood labels in Table 8 and 9, as well as the "Top Tags" list for each of the artists. 17 of the 22 artist-mood pairs in Tables 8 and 9 were corroborated either by successfully identifying the artists in the "Top Artists" lists of the corresponding tags (10 pairs) or by identifying the tags in the "Top Tags" lists of the corresponding artists (7 pairs). The 5 non-corroborated artist-mood pairs include: *The Beatles*–"Whimsical", *The Grateful Dead*–"Trippy", Miles Davis–"Uncompromising", *Thelonious Monk*–"Quirky", and David Bowie–"Campy".

To corroborate artist-mood cluster pairs, we combined the "Top artists" lists of all the mood labels in each cluster. By the same method, 15 of the 17 pairs in Table 10 (except for Miles Davis–Cluster5 and John Coltrane with Johnny Hartma–Cluster3) were corroborated.

7.3 Corroboration of Mood and Usage Associations

Using the same method as in Section 7.1, we built three sample sets based on top albums, tracks and artists with

[2] http://www.audioscrobbler.net/data/webservices

at least one usage tag and one mood tag that appeared in Tables 11 and 12. Please note that some of the usage tags are not available in `Last.fm` such as "Hanging out with friends", and "Romancing". Others have very few occurrences, such as "Cleaning the house". We tried to locate tags similar to these phrases (e.g., "hanging out", "cleaning"). Thus, results from this dataset disclose quite different associations than those from the AMG sets. The only 3 pairs corroborated are ($p < 0.01$): "Going to sleep"–"Bittersweet", "Driving"–"Menacing", and "Listening"–"Epic".

By combining the albums/tracks/artists lists with all the mood labels in each cluster, we corroborated only 2 usage-mood cluster pairs found in Table 13: "Going to sleep"–Cluster3 ($p = 0.001$), "Driving"–Cluster5 ($p < 0.015$). Again, these observations indicate that the relationship between usage and mood is not stable and is most likely dependent on the specific vocabularies present in the datasets they are derived from.

8 RECOMMENDATIONS

The usage-mood relationships are not stable enough to warrant further consideration. However, the genre-mood and artist-mood relationships explored in this study show great promise in helping construct a meaningful MIREX "AMC" task. The corroborative analyses using the `Last.fm` data sets provide additional evidence that the nature of these two relationships is generalizeable beyond our original AMG data source.

Mood term vocabulary size (and its uneven distribution across items) is a huge impediment to the construction of useable ground-truth sets (e.g., AMG's 179 mood terms). Throughout this study we saw that many of the individual mood terms were highly synonymous or described aspects of the same underlying, more general, "mood space". Thus, we found that decreasing mood vocabulary size in some ways actually clarified the underlying mood of the items being described. We therefore recommend that MIREX members consider constructing an "AMC" task based upon a set of "mood space" clusters rather than individual mood terms. The clusters themselves need not be those presented here but should be relatively small in number. As Table 14 shows, a cluster-based approach also improves the distribution of albums and artists in AMG across the clusters.

	Cluster1	**Cluster2**	**Cluster3**	**Cluster4**	**Cluster5**
Albums	355	285	486	493	372
Artist	14	16	85	87	46

Table 14. AMG sample distributions across mood clusters

Under a fully automated scenario (i.e., no human evaluation), ground-truth sets could be constructed by locating those works, across both artists and genres, which are represented in each cluster by mapping the constituent mood terms back to those artists and genres to which they have statistically significant relationships.

Under a human evaluation scenario (e.g. [4]), training sets would be similarly constructed. However, for evaluation itself, the human evaluators would be given exemplars from each of the 5 (or so) clusters to give them an understanding of their "nature". The limited number of clusters increases the probability of evaluator consistency. Scoring would be based on the agreement between system and evaluator assigned cluster memberships.

9 ACKNOWLEDGEMENTS

This project is funded by the National Science Foundation (IIS-0327371) and the Andrew W. Mellon Foundation.

10 REFERENCES

[1] Berkhin, P. *Survey of Clustering Data Mining Techniques*, Accrue Software, 2002.

[2] Buntinas, M. and Funk, G. M. *Statistics for the Sciences*, Brooks/Cole/Duxbury, 2005.

[3] Downie, J. S. "The Music Information Retrieval Evaluation eXchange (MIREX)", *D-Lib Magazine 2006, Vol. 12(12).*

[4] Gruzd, A. A., Downie J. S., Jones, M. C. and Lee, J. H. "Evalutron 6000: collecting music relevance judgments", *Proceedings of the Joint Conference on Digital Libraries (JCDL 2007).*

[5] Hu, X., Downie, J. S. and Ehmann, A. F. "Exploiting recommended usage metadata: Exploratory analyses", *Proceedings of the 7th International Conference on Music Information Retrieval, ISMIR'06,* Victoria, Canada.

[6] Juslin, P.N., Karlsson, J., Lindström E., Friberg, A. and Schoonderwaldt, E. "Play it again with feeling: computer feedback in musical communication of emotions", *Journal of Experimental Psychology: Applied 2006, Vol.12(1).*

[7] Li, T. and Ogihara, M. "Detecting emotion in music", *Proceedings of the 4th International. Conference on Music Information Retrieval ISMIR 2003*, Washington, D.C.

[8] Lu, L., Liu, D. and Zhang, H. "Automatic Mood Detection and Tracking of Music Audio Signals", *IEEE Transaction on Audio, Speech, and Language Processing 2006, Vol.14(1).*

[9] Mandel, M. Poliner, G. and Ellis, D. "Support vector machine active learning for music retrieval", *Multimedia Systems 2006, Vol.12(1).*

[10] Vignoli, F. "Digital Music Interaction concepts: a user study", *Proceedings of the 5th Int. Conference on Music Information Retrieval, ISMIR 2004*, Barcelona, Spain.