ON CULTURAL, TEXTUAL AND EXPERIENTIAL ASPECTS OF MUSIC MOOD

Abhishek Singhi **Daniel G. Brown**

University of Waterloo
Cheriton School of Computer Science
{asinghi,dan.brown}@uwaterloo.ca

ABSTRACT

We study the impact of the presence of lyrics on music mood perception for both Canadian and Chinese listeners by conducting a user study of Canadians not of Chinese origin, Chinese-Canadians, and Chinese people who have lived in Canada for fewer than three years. While our original hypotheses were largely connected to cultural components of mood perception, we also analyzed how stable mood assignments were when listeners could read the lyrics of recent popular English songs they were hearing versus when they only heard the songs. We also showed the lyrics of some songs to participants without playing the recorded music. We conclude that people assign different moods to the same song in these three scenarios. People tend to assign a song to the mood cluster that includes "melancholy" more often when they read the lyrics without listening to it, and having access to the lyrics does not help reduce the difference in music mood perception between Canadian and Chinese listeners significantly. Our results cause us to question the idea that songs have "inherent mood". Rather, we suggest that the mood depends on both cultural and experiential context.

1. INTRODUCTION

Music mood detection has been identified as an important Music Information Retrieval (MIR) task. For example, there is a MIREX audio mood classification task [12]. Though most automatic mood classification systems are solely based on the audio content of the song, some systems have used lyrics or have combined audio and lyrics features (*e.g.,* [3-5] and [6-7]) Previous studies have shown that combing these features improves classification accuracy (*e.g.*, [6-7] and [9]) but as mentioned by Downie et al. in [3], there is no consensus on whether audio or lyrical features are more useful.

Implicit in "mood identification" is the belief that songs have "inherent mood," but in practice this assignment is unstable. Recent work has focused on associating songs with more than one mood label, where similar mood tags are generally grouped together into the same label (*e.g.*, [2]), but this still tends to be in a stable listening environment.

Our focus is instead on the cultural and experiential context in which people interact with a work of music. People's cultural origin may affect their response to a work of art, as may their previous exposure to a song, their perception of its genre, or the role that a song or similar songs has had in their life experiences.

We focus on people's cultural origin, and on how they interact with songs (for example, seeing the lyrics sheet or not). Listening to songs while reading lyrics is a common activity: for example, there are "lyrics videos" (which only show lyrics text) on YouTube with hundreds of millions of views (*e.g.* "Boulevard of Broken Dreams"), and CD liner notes often include the text of lyrics. Our core hypothesis is that there is enough plasticity in assigning moods to songs, based on context, to argue that many songs have no inherent "mood".

Past studies have shown that there exist differences in music mood perception among Chinese and American listeners (*e.g.*, [8]). We surmised that some of this difference in mood perception is due to weak English language skills of Chinese listeners: perhaps such listeners are unable to grasp the wording in the audio. We expected that they might more consistently match the assignments of native English-speaking Canadians when shown the lyrics to songs they are hearing than in their absence. We addressed the cultural hypothesis by exploring Canadians of Chinese origin, most of whom speak English natively but have been raised in households that are at least somewhat culturally Chinese. If such Chinese-Canadians match Canadians not of Chinese origin in their assignments of moods to songs, this might at least somewhat argue against the supposition that being Chinese in culture had an effect on mood assignment, and would support our belief that linguistic skills account for at least some of the differences. Our campus has many Chinese and Chinese-Canadians, which also facilitated our decision to focus on these communities.

In this study we use the same five mood clusters as are used in the MIREX audio mood classification task and ask the survey participants to assign a song to only one mood cluster. A multimodal mood classification could be a possible extension to our work here. Earlier works in MIR [11] had used Russell's valence-arousal model where the mood is determined by the valence and arousal scores of the song; we stick to the simpler classification here.

In practice, our hypotheses about language expertise were not upheld by our experimental data. Rather, our data support the claim that both cultural background and experiential context have significant impact on the mood assigned by listeners to songs, and this effect makes us question the meaningfulness of "mood" as a category in MIR.

2. RELATED WORK

Mood classification is a classic task in MIR, and is one of the MIREX challenges. Several projects have used lyrics as part of the mood prediction task. Lu et al. [1] and Trohidis et al. [2] come up with an automatic mood classification system solely based on audio. Several projects like Downie et al. [3], Xiong et al. [4] and Chen et al. [5], have used lyrics as part of the mood prediction task. Downie et al. [3] show that features derived from lyrics outperform audio features in 7 out of the 8 categories. Downie et al. [6], Laurier et al. [7] and Yang et al. [9] show that systems which combine audio and lyrics features outperform systems using only audio or only lyrics features. Downie et al. [6] show that using a combination of lyrics and audio features reduces the need of training data required to achieve the same or better accuracy levels than only-audio or only-lyrics systems.

Lee et al. [8] study the difference in music mood perception between Chinese and American listeners on a set of 30 songs and conclude that mood judgment differs between Chinese and American participants and that people belonging to the same culture tend to agree more on music mood judgment. That study primarily used the common Beatles data set, which may have been unfamiliar to all audiences, given its age. Their study collected mood judgments solely based on the audio; we also ask participants to assign mood to a song based on its lyrics or by presenting both audio and lyrics together. To our knowledge no work has been done on the mood of a song when both audio and lyrics of the song is made available to the participants, which as we have noted is a common experience. Kosta et al. [11] study if Greeks and non-Greeks agree on arousal and valence rating for Greek music. They conclude that there is a greater degree of agreement among Greeks compared to non-Greeks possibly because of acculturation to the songs.

Downie et al. [3], Laurier et al. [7] and Lee et al. [8] use 18 mood tags derived from social tags and use multimodal mood classification system. Trohidis et al. [2] use multi modal mood classification into six mood clusters. Kosta et al. [11] use Russell's valence-arousal model which has 28 emotion denoting adjectives in a two dimensional space. Downie et al. [10] use the All Music Guide datasets to come up with 29 mood tags and cluster it into five groups. These five mood clusters are used in the MIREX audio music mood classification task. We use these clusters where each song is assigned a single mood cluster.

Mood Clusters	Mood Tags
Cluster 1	passionate, rousing, confident, boisterous, rowdy
Cluster 2	rollicking, cheerful, fun, sweet, amiable/good natured
Cluster 3	literate, poignant, wistful, bittersweet, autumnal, brooding
Cluster 4	humorous, silly, campy, quirky, whimsical, witty, wry
Cluster 5	aggressive, fiery, tense/anxious, intense, volatile, visceral

Table 1. The mood clusters used in the study.

3. METHOD

3.1 Data Set

We selected fifty very popular English-language songs of the 2000's, with songs from all popular genres, and with an equal number of male and female singers. We verified that the selected songs were international hits by going to the songs' Wikipedia pages and analyzing the peak position reached in various geographies.

We focus on English-language popular music in our study, because it is the closest to "universally" popular music currently extant, due to the strength of the music industry in English-speaking countries. Our data set includes music from the US, Canada and the UK and Ireland.

3.2 Participants

The presence of a large Chinese and Canadian population at our university, along with obvious cultural differences between the two communities, convinced us to use them for the study. We also include Canadians of Chinese origin; we are unaware of any previous MIR work that has considered such a group. We note that the Chinese-Canadian group is diverse: while some speak Chinese languages, others have comparatively little exposure to Chinese language or culture.

We recruited 100 participants, mostly university students, from three groups:

- 33 Chinese, living in Canada for less than 3 years.
- 33 Canadians, not of Chinese origin, born and brought up in Canada, with English as their mother tongue.
- 34 Canadians, of Chinese origin, born and brought up in Canada.

3.3 Survey

Each participant was asked to assign a mood cluster to each song in a set of 10 songs. For the first three songs

they saw only the lyrics; for the next three songs they only heard the first 90 seconds of the audio; and for the last four songs they had access to both the lyrics and the first 90 seconds of the audio simultaneously. They assigned each song to one of the five mood clusters shown in Table 1. We collected 1000 music mood responses for 50 songs, 300 each based solely either on audio or lyrics and 400 based on both audio and lyrics together. We note that due to their high popularity, some songs shown only via lyrics may have been known to some participants. We did not ask participants if this was the case.

4. RESULTS

We hypothesized that the difference in music mood perception between American and Chinese listeners demonstrated by Hu and Lee [8] is because of the weak spoken English language skills of Chinese students, and that this might give them some difficulty in understand the wording of songs; this is why we allowed our participants to see the lyrics for seven out of ten songs.

We had the following set of hypotheses before our study:

- People often assign different mood to the same song depending on whether they read the lyrics, or listen the audio or both simultaneously.
- Chinese-born Chinese listeners will have less consistency in the assignment of moods to songs than do Canadian-born non-Chinese when given only the recording of a song.
- Chinese-born Chinese will more consistently match Canadians when they are shown the lyrics to songs.
- Just reading the lyrics will be more helpful in matching Canadians than just hearing the music for Chinese-born Canadians.
- Canadian-born Chinese participants will be indistinguishable from Canadian-born non-Chinese participants.
- A song does not have an inherent mood: its "mood" depends on the way it is perceived by the listener, which is often listener-dependent.

4.1 Lyrics and music mood perception between cultures

We started this study with the hypothesis that difference in music mood perception between Chinese and Canadian cultures is partly caused by English language skills, and that if participants are asked to assign mood to a song based on its lyrics, we will see much more similarity in judgment between two different groups.

We used the Kullback-Leibler distance between the distribution of responses from one group and the distribution of responses from that group and another group to identify how similar the two groups' assignments of moods to songs were, and we used a permutation test to identify how significantly similar or different the two groups were. In Table 2, we show the number of songs for which different population groups are surprisingly similar. What we find is that the three groups actually agree quite a bit in uncertainty of assigning mood to songs when they are presented only with the recording: if one song has uncertain mood assignment for Canadian listeners, our Chinese listeners also typically did not consistently assign a single mood to the same song.

Our original hypothesis was that adding presented lyrics to the experience would make Chinese listeners agree more with the Canadian listeners, due to reduced uncertainty in what they were hearing. In actuality, this did not happen at all: in fact, presence of both audio and lyrics resulted in both communities having both more uncertainty and disagreeing about the possible moods to assign to a song.

This confusion in assigning a mood might be because a lot of hit songs (“Boulevard of Broken Dreams”, “Viva La Vida”, “You’re Beautiful”, *etc.*) use depressing words with very upbeat tunes. It could also be that by presenting both lyrics and audio changes the way a song is perceived by the participants and leads to a completely new experience. (We note parenthetically that this argues against using lyrics only features in computer prediction of song mood.)

The number of songs with substantial agreement between Chinese and Canadian, not of Chinese origin, participants remains almost the same with lyrics only and audio only, but falls drastically when both are presented together. (Note again: in this experiment, we are seeing how much the distribution of assignments differs for the two communities.) This contradicts our hypothesis that the difference in music mood perception between Chinese and Canadians is because of their difference in English abilities. It could of course be the case that many Chinese participants did not understand the meaning of some of the lyrics.

We had hypothesized that Canadians, of Chinese and non-Chinese origin would have very similar mood judgments because of similar English language skills but they do tend to disagree a lot on music mood. The mood judgment agreement between Chinese and Canadian, of Chinese and non-Chinese origin seem to be similar and we conclude that we can make no useful claims about the Chinese-Canadian participants in our sample.

On the whole we conclude that the presence of lyrics does not significantly increase the music mood agreement between Chinese and Canadian participants: in fact, being able to read lyrics while listening to a recording seems to significantly decrease the music mood agreement between the groups.

		lyrics	audio	audio+lyrics
Chinese	Canadians	25	22	14
Chinese	Canadian-Chinese	36	31	27
Chinese	non-Chinese Canadians	31	32	23
non-Chinese Canadians	Canadian-Chinese	36	29	31

Table 2. The number of statistically significantly similar responses between the different cultures for the three different ways they interact with the songs. "Canadians" refer to Canadians of both Chinese and non-Chinese origin.

4.2 Stability across the three kinds of experiences

We analyze the response from participants when they are made to listen to the lyrics, hear the audio or both simultaneously across all the three groups. We calculate Shannon entropy of this mood assignment for each of the 50 songs for the three ways we presented a song to the participants: some songs have much more uncertainty in how the participants assign mood cluster to them. We then see if this entropy is correlated across the three kinds of experience, using Spearman's rank correlation coefficient of this entropy value between the groups. A rank correlation of 1.0 would mean that the song with the most entropy in its mood assignment in one experience category is also the most entropic in the other, and so on.

	Spearman's rank correlation coefficient
only lyrics & only audio	0.0504
only lyrics & audio+lyrics	0.1093
only audio & audio+lyrics	0.0771

Table 3. Spearman's rank correlation coefficient between the groups. The groups "only lyrics" and "only audio" identify when participants had access to only lyrics and audio respectively while "audio+lyrics" refers to when they had access to both simultaneously.

The low value of the correlation analysis suggests that there is almost no relationship be-tween "certainty" in music mood across the three different kinds of experiences: for songs like "Wake Up" by Hillary Duff and "Maria Maria" by Santana, listeners who only heard the song were consistent in their opinion that the song was from the second cluster, "cheerful", while listeners who heard the song and read the lyrics were far more uncertain as to which class to assign the song to.

4.3 "Melancholy" lyrics

For each song, we identify the mood cluster to which it was most often assigned, and show these in Table 4.

Mood Clusters	only lyrics	only audio	audio+lyrics
Cluster 1	8	9	13
Cluster 2	5	15	11
Cluster 3	28	14	18
Cluster 4	4	6	3
Cluster 5	5	6	5

Table 4. The most commonly assigned mood clusters for each experimental context. Most songs are assigned to the third mood cluster when participants are shown only the lyrics.

Songs experienced only with the lyrics are most often assigned to the third mood cluster, which includes the mood tags similar to "melancholy". In the presence of audio or both audio and lyrics there is a sharp decline in the number of songs assigned to that cluster; this may be a consequence of "melancholy" lyrics being attached to surprisingly cheery tunes that cause listeners to assign them to the first two clusters. The number of songs assigned to the fourth and fifth cluster remains more similar across all experiential contexts. Even between the two contexts where the listener does hear the recording of the song, there is a good deal of inconsistency in assignment of mood to songs: for 27 songs, the most commonly identified mood is different between the "only audio" and "audio+lyrics" data.

4.4 Rock songs

We explored different genres in our test set, to see if our different cultural groups might respond in predictable ways when assigning moods to songs.

Things that might be considered loud to Chinese listeners could be perceived as normal to Canadian listeners. Thus, we examined how responses differed across these two groups for rock songs, of which we had twelve in our data set. We calculate the Shannon entropy of the response of the participants and present the result in table 5.

We see that for many rock songs, there is high divergence in the mood assigned to the song by our listeners from these diverse cultures. For seven of the twelve rock songs, the most diversity of opinion is found when listeners both read lyrics and hear the audio, while for three songs, all participants who only read the lyrics agreed exactly on the song mood (zero entropy).

We see that for 3 of 12 cases all the participants tend to agree on the mood for the song when they are given access to the lyrics. The data for lyrics only have lower entropy than audio for 5 of 12 cases and all five of these songs are "rebellious" in style. For the five cases where the audio-only set has lower entropy than lyrics-only, the song has a more optimistic feel to it. This is consistent with our finding in the last section about melancholy song lyrics.

For example, the lyrics of "Boulevard of Broken Dreams", an extremely popular Green Day song, evoke isolation and sadness, consistent with the third mood cluster. On the other hand the song's music is upbeat which

may give the increased confusion when the participant has access to both the audio and lyrics for the song.

Song	only lyrics	only audio	audio+lyrics
"Complicated"	1.0	**0.918**	1.148
"American Idiot"	1.792	**1.459**	1.792
"Apologize"	**1.0**	1.25	1.0
"Boulevard of Broken Dreams"	**0.0**	1.792	2.155
"Bad Day"	1.792	1.459	**1.061**
"In the End"	**0.65**	1.459	1.061
"Viva La Vida"	**0.0**	1.5849	1.75
"It's My life"	**0.0**	0.65	1.298
"Yellow"	1.792	**0.65**	1.351
"Feel"	0.918	**0.650**	1.148
"Beautiful Day"	1.584	**1.459**	1.836
"Numb"	1.25	1.918	**0.591**

Table 5. Entropy values for rock songs for the three different categories.

4.5 Hip-Hop/ Rap

Lee et al. [8] show that mood agreement among Chinese and American listeners is least for dance songs. Our test set included five rap songs, and since this genre is often used at dance parties, we analyzed user response for this genre. Again, we show the entropy of mood assignment for the three different experiential contexts in Table 6.

What is again striking is that seeing the lyrics (which in the case of rap music is the primary creative element of the song) creates more uncertainty among listeners as to the mood of the song, while just hearing the audio recording tends to yield more consistency. Perhaps this is because the catchy tunes of most rap music pushes listeners to make a spot judgment as to mood, while being reminded of lyrics pushes them to evaluate more complexity.

In general we see that there is high entropy in mood assignment for these songs, and so we confirm the previous claim that mood is less consistent for "danceable" songs.

5. DOES MUSIC MOOD EXIST?

For music mood classification to be a well-defined task, the implicit belief is that songs have "inherent mood(s)," that are detectable by audio features. Our hypothesis is that many songs have no inherent mood, but that the perceived mood of a song depends on cultural and experiential factors. The data from our study supports our hypothesis.

We have earlier shown that the mood judgment of a song depends on whether it is heard to or its lyrics is read or both together, and that all three contexts produce mood assignments that are strikingly independent.

We have shown that participants are more likely to assign a song to the "melancholic" mood cluster when only reading its lyrics, and we have shown genre-specific cultural and experiential contexts that affect how mood appears to be perceived. Together, these findings suggest that that the concept of music mood is fraught with uncertainty.

The result of the MIREX audio mood classification task has had a maximum classification accuracy of less than 70% [12], with no significant recent improvements. Perhaps, this suggests that the field is stuck at a plateau, and we need to redefine "music mood" and change our approach to the music mood classification problem. Music mood is highly affected by external factors like the way a listener interacts with the song, the genre of the song, the mood and personality of the listener, and future systems should take these factors into account.

Song	only lyrics	only audio	audio+lyrics
"London Bridge"	1.459	**0.918**	1.405
"Don't Phunk With My Heart"	1.459	**1.251**	1.905
"I Wanna Love You"	**0.918**	1.459	1.905
"Smack That"	1.918	**1.792**	1.905
"When I'm Gone"	1.251	**0.918**	1.448

Table 6. Entropy values for hip-hop/ rap songs for the three different categories.

6. CONCLUSION

Our experiment shows that the presence of lyrics has a significant effect on how people perceive songs. To our surprise, reading lyrics alongside listening to a song does not significantly reduce the differences in music mood perception between Canadian and Chinese listeners. Also, while we included two different sets of Canadian listeners (Canadian-Chinese, and Canadians not of Chinese origin), we can make no useful conclusions about the Chinese-Canadian group.

We do consistently see that presence of both audio and lyrics reduces the consistency of music mood judgment between Chinese and Canadian listeners. This phenomenon may be because of irony caused by negative words presented in proximity to upbeat beats, or it could be that presenting both audio and lyrics together might be a completely different experience for the listener. This is an obvious setting for further work.

We have shown that the mood of a song depends on its experiential context. Interestingly, songs where listeners agree strongly about the mood of the song when only listening to the recording are often quite uncertain in their mood assignments when the lyrics are shown alongside the recording. Indeed, there is little correlation between the entropy in mood assignment between the different ways we presented songs to participants.

We also show that many "melancholy" lyrics are found in songs assigned to a more cheerful mood by listeners, again suggesting that for such songs, the extent to which listeners focus on the lyrics may influence how sad they view a song to be. We analyzed the mood assignments of participants on rock and hip-hop/rap songs. We see that people tend to agree much more to the mood of a hip-hop/rap song when they are made to listen to the song. We found that for rebellious/negative rock songs lyrics leads to more agreement in music mood but audio is better for positive songs. In both the genres we found that hearing audio while reading lyrics lead to less agreement on music mood of songs.

Our results suggest that music mood is so dependent on cultural and experiential context to make it difficult to claim it as a true concept. With the classification accuracy of mood classification systems reaching a plateau with no significant improvements we suggest that we need to re-define the term "music mood" and change our approach toward music mood classification problem.

A possible extension to our work could be running a similar study using a larger set of songs and more participants, possibly from more diverse cultures than the ones we studied. Future studies could focus on multi-modal music mood classification where a song could belong to more than one mood, to see if even in this more robust domain there is a stable way to assign songs to clusters of moods when they are experienced in different contexts. We also wonder if other contextual experiments can show other effects about mood: for example, if hearing music while in a car or on public transit, or in stores, makes the "mood" of a song more uncertain.

We fundamentally also wonder if "mood" as an MIR concept needs to be reconsidered. If listeners disagree more or less about the mood of a song when it is presented alongside its lyrics, that suggests a general uncertainty in the concept of "mood". We leave more evidence gathering about this concept to future work as well.

7. ACKNOWLEDGEMENTS

Our research is supported by a grant from the Natural Sciences and Engineering Research Council of Canada to DGB.

8. REFERENCES

[1] L. Lu, D. Liu, and H. Zhang, "Automatic mood detection and tracking of music audio signals", in *IEEE Transactions on Audio, Speech, and Language Processing*, volume 14, number 1, pages 5-18, 2006.

[2] K. Trohidis, G. Tsoumakas, G. Kalliris and I. Vlahvas, "Multi-label classification of music into emotions", in *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR '08)*, pages 325-330, 2008.

[3] X. Hu and J.S. Downie, "When lyrics outperform audio for music mood classification: a feature analysis", in *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR '10)*, pages 1–6, 2010.

[4] H. He, J. Jin, Y. Xiong, B. Chen, W. Sun, and L. Zhao, "Language feature mining for music emotion classification via supervised learning from lyrics", in *Proceedings of Advances in the 3rd International Symposium on Computation and Intelligence (ISICA '08)*, pages 426-435, 2008.

[5] Y. Hu, X. Chen and D. Yang, "Lyric-based song emotion detection with affective lexicon and fuzzy clustering method", in *Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR '09)*, pages 123-128, 2009.

[6] X. Hu and J. S. Downie, "Improving mood classification in music digital libraries by combining lyrics and audio", in *Proceedings of Joint Conference on Digital Libraries (JCDL)*, pages 159–168, 2010.

[7] C. Laurier, J. Grivolla, and P. Herrera, "Multimodal music mood classification using audio and lyrics", in *Proceedings of the International Conference on Machine Learning and Applications (ICMLA '08)*, pages 688-693, 2008.

[8] X. Hu and J.H. Lee, "A cross-cultural study of music mood perception between American and Chinese listeners," in *Proceedings of International Society for Music Information Retrieval (ISMIR '12)*, pages 535-540, 2012.

[9] Y.H. Yang, Y.C. Lin, H.T. Cheng, I.B. Liao, Y.C. Ho, and H.H. Chen, "Toward multi-modal music emotion classification", in *Proceedings of Pacific Rim Conference on Multimedia (PCM '08)*, pages 70-79, 2008.

[10] X. Hu. and J.S. Downie, "Exploring mood metadata: relationships with genre, artist and usage metadata", in *Proceedings of the 8th International Conference on Music Information Retrieval, (ISMIR '07)*, pages 67-72, 2007.

[11] K. Kosta, Y. Song, G. Fazekas and M. Sandler, "A study of cultural dependence of perceived mood in Greek music", in *Proceedings of the 14th International Society for Music Information Retrieval Conference, (ISMIR '13)*, pages 317-322, 2013.

[12] X. Hu, J. S. Downie, C. Laurier, M. Bay, and A. F. Ehmann: "The 2007 MIREX audio mood classification task: Lessons learned," in *Proceedings of the 9th International Society for Music Information Retrieval Conference, (ISMIR '08)* , pages 462–467, 2008.