Cross-cultural aspects of perceptual features in K-Pop: A pilot study comparing Chinese and Swedish listeners

Anders Friberg
KTH Royal Institute of technology,
Stockholm, Sweden
afriberg@kth.se

Ragnar Schön
KTH Royal Institute of technology,
Stockholm, Sweden
ragnarschon@gmail.com

Anders Elowsson
KTH Royal Institute of technology,
Stockholm, Sweden
elov@kth.se

Kahyun Choi
University of Illinois
Champaign, IL, 61820
ckahyu2@illinois.edu

J. Stephen Downie
University of Illinois
Champaign, IL, 61820
jdownie@illinois.edu

ABSTRACT

In previous studies it has been shown that perceptual features can be used as an intermediate representation in music processing to model higher-level semantic descriptions. In this pilot study, we focused on the cross-cultural aspect of such perceptual features, by asking both Chinese and Swedish listeners to rate a set of K-Pop samples using a web-based questionnaire. The music samples were selected from a larger set, previously rated in terms of different emotion labels. The selection procedure of the subset was carefully designed to maximize both the variation of emotion and genre. The listeners rated eight perceptual features: dissonance, speed, rhythmic complexity, rhythmic clarity, articulation, harmonic complexity, modality, and pitch. The results indicated a small but significant difference in the two groups, regarding the average speed and rhythmic complexity. In particular the perceived speed of hip hop was different for the two groups. We discuss the overall consistency of the ratings using this methodology in relation to the interface, selection and number of subjects.

1. INTRODUCTION

1.1 Cross-cultural MIR

Much of MIR research has primarily focused on the Western music tradition. Since instrumentation and style varies across the world, the result is that MIR is culturally biased towards Western music. Serra [17] argues that the plurality of robust non-Western music traditions form a counterpoint to the Western tradition and suggests that a cross-fertilization between musicology and music cognition methodologies could be of benefit to MIR research. Cross-cultural MIR may be able to catch the richness of the world's diverse musical cultures. One attempt to appoach this lack of non-western material was by introducing a new dataset consisting of Korean pop (K-Pop) songs including extensive human annotations of both genre [13] and mood [12]. A subset of this dataset was used in the present study.

Several studies have identified cultural differences in the perception of music. Hu et al. [12] showed that Koreans were more prone to classify music with simple emotions, such as *sad*, as compared to Americans. Yang and Hu [19] observed that Koreans were more likely to rate music with negative moods than Americans. According to Hu and Lee [11], cultural context is an important factor, affecting how people determine the mood of music. The different languages of lyrics are perhaps one of the most obvious cultural differences in music. For example, the lyrics is a strong factor for mood perception. In this study, we focus on the musical aspects by using a dataset comprising of songs in a language not native to either of the two participant groups. The cross-cultural aspect of the communication of emotion in music has been further investigated by, for example, Thompson and Balkwill [18]. A general conclusion is that the emotional expression of music that is unknown, and coming from another cultural tradition, can be perceived to a certain extent, but less accurate than by native listeners. There is also evidence for cross-cultural differences of basic pitch processing, for example regarding the tritone paradox [3]. Since we are focusing on basic aspects such as *speed* and *pitch* in this study, we expect the differences to be relatively small.

1.2 Perceptual features

Content-based analysis of audio is a fundamental part of MIR research. It has to a large extent been based on the use of features extracted directly from the audio waveform and spectrogram data (see e.g. [1]). These features may be low-level, such as MFCC's and spectral descriptors, or psychoacoustic measures such as loudness. Mid-level features traditionally use slightly larger analysis windows, giving features more closely related to music theory, such as key, harmony, beat and meter. Finally, top-level semantic descriptors such as genre and mood are typically acquired, not by computational audio analysis, but by collecting human annotations. A typical MIR task is the prediction of semantic descriptors from these low- and mid-level features. To facilitate this kind of task, a set of *perceptual features* has been proposed as an intermediate layer [7-9]. These features have been chosen so that non-expert everyday listeners may rate them without difficulty. The goal has been to find features that are

relevant in everyday listening, and that can be used to describe the music so that predictions of for example emotional expression can be done directly from the perceptual features. Previous feature-driven models are often based on a pure engineering approach mainly focusing on the resulting accuracy or starting from a music theoretic viewpoint. Perceptual features may provide the intermediate step required to bridge low-level features with semantic features in order to ease the prediction of the latter using the former.

One underlying idea for selecting the perceptual features has been to view music perception as a part of ecological sound perception [2, 6]. We listen to music with the same perceptual tools that are used to decipher all sounds. The ecological viewpoint suggests that it is more natural for human listeners to decode the source properties of a sound, such as "a mallet hitting a drum", than the spectral properties of the resulting sound wave.

Friberg et al. [9] proposed nine perceptual features determined by listener ratings. They are presumably easier to model, as they are both on a lower level than the semantic descriptors, while at the same time are addressing the lack of psychological validity for low-level features. The present study uses these perceptual features with some modifications as detailed in section 3.1 below.

1.3 Research question

In the previous studies of perceptual features, the agreement varied for the different features. The highest agreement was obtained for *speed* with the mean pairwise correlation equal to 0.71/0.60 and Cronbach's alpha equal to 0.98/0.97, for two different databases with 20 listeners in each [9]. These studies involved mainly European listeners and Western music. The results indicate that there is a strong agreement among these listeners.

As mentioned above, there is evidence for a cultural difference in the perception of emotion and mood. This is not surprising, given the complexity of emotion perception. However, the underlying features have a more basic character and may thus be more universally valid.

The purpose of this pilot study is to investigate if there is a cultural difference in perceiving the perceptual features.

2. K-POP DATASET SELECTION

Korean pop music, or simply K-pop, is a term often used to denote a specific style of energetic Korean pop music made known worldwide by songs such as Psy's 2012 *Gangnam Style*. In this study, a broader definition of K-pop is used, denoting several styles of Korean music including rock, hip hop and trot. Trot is a style of sentimental love songs, developed in Korea in the beginning of the 20th century. The dataset used here is a subset of the collection of 1892 K-pop songs, first introduced for the 2014 MIREX mood and genre challenges. The dataset is divided in seven genres [13] and has been mood annotated by American and Korean listeners according to two categorical models as well as the two-dimensional activity-valence model, further details in [12].

The subset used in this study consists of 98 songs from this dataset that were carefully selected to maximize the spread in genre within and across mood. Priority was given to songs with a "clear" mood rating, i.e. far from the center in the activity-valence model. Only songs having a high level of agreement in the genre ratings were used. The selection was achieved by weighting together measures for clearness of genre, distance from the center and spread between the four quadrants of the valence-activity space according to

$$S = W_g S_g + W_{va} S_{va} + W_m, \quad (1)$$

where S is the total score per song, W_g and W_{va} are the weights for genre and valence/activity respectively. S_g and S_{va} are the respective scores and W_{va} was set per genre. W_m is a weight created to steer points into the desired quadrant. S_g was set to 1 if all annotators agreed, to 0.75 if five out of six annotators agreed, to 0.5 if four out of six agreed, and to 0 in all other cases. S_{va} is the Euclidean distance from the origin in the activity-valence space

$$S_{va} = \sqrt{V^2 + A^2}\,, \quad (2)$$

where V is the valence rating and A is the activity rating. The weights were then manually adjusted in order to maximize both genre spread and emotion spread. The resulting selection can be seen in Figure 1, where the selected songs corresponds to the large blue dots. Not all genres are represented in every quadrant. For example, no hip hop songs had activity ratings below 0.

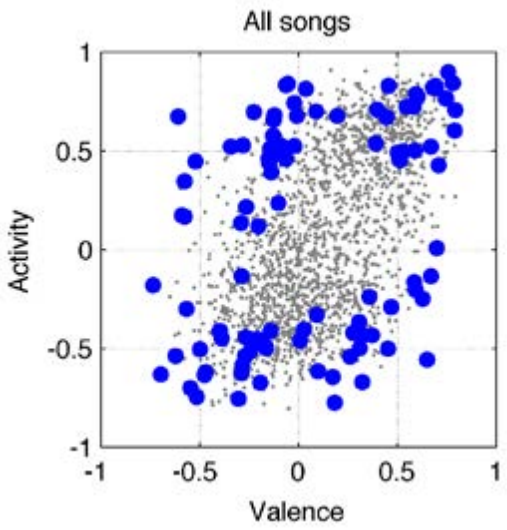

Figure 1. The small grey dots show the location of all songs of the original dataset in the activity-valence space. The selected songs are shown as large blue dots.

3. PERCEPTUAL RATINGS

3.1 Rated features

Nine perceptual features were rated by the listeners using continuous scales. The features, with their respective

scale end points were: *Speed* is how fast the music is perceived. It is closely related, but not equivalent, to tempo. It ranges from slow to fast. *Rhythmic Complexity* deals with rhythmical patterns. It ranges from simple to complex, where a simple rhythm could be a 4/4 beat, whereas a complex rhythm may contain unconventional time signatures and/or syncopations. *Rhythmic Clarity* is the clarity and accentuation of the main pulse. It ranges from flowing to firm. *Articulation* is how tied or shortened notes are, and it ranges from legato to staccato. *Modality* is a continuous measure that ranges from minor to major. *Harmonic Complexity* is a measure of how complex the harmonic progression is. It ranges from simple to complex. *Overall Pitch* is simply how high the overall pitch is perceived. It ranges from low to high. In addition, *Dynamics*, ranging from soft to loud, was also rated. However, due to a bug in the software it was not recorded.

All these features has been rated and evaluated in previous experiments (e.g. [9]). From these experiments, and from a priori expectation, it was evident that Harmonic Complexity was not as easily rated by listeners as the other features. Therefore, as motivated from a previous pilot study [15] *Dissonance* was added in order to see if it works better for capturing the variation in harmony. It ranges from consonant to dissonant. To make comparisons, *Harmonic Complexity* was also kept in the rating experiment. For more information about the experiment see [16].

3.2 Listeners and questionnaire data

Listeners were recruited from two groups representing a Western and a non-Western cultural background. The Western group consisted of Swedes and was recruited through an online social network. The non-Western group were Chinese students at KTH that was recruited with the help of a Chinese student organization.. The listeners were compensated with two cinema vouchers. Eleven Swedish and thirteen Chinese listeners volunteered. Half of the listeners were female. Mean age was 28.1 years (Ch. 27.5, Sw. 28.8).

They listened to music an average of 12.7 hours per week (Ch. 8.7, Sw. 17.4) and practiced a musical instrument an average of 2.0 hours per week (Ch. 1.9, Sw. 2.1). The corresponding values for each musical level can be seen in Table 1. Most listeners regarded themselves as being a beginner or having an intermediate musical experience level (9 and 10 listeners, respectively). Only one listener regarded himself as a professional, while four stated that they had no musical experience at all. Thus, on average the listeners had some musical experience, but were not professionals.

Using listeners with this level of musical training has been found to work well in the previous experiments. Some musical training helps to understand the different features, which presumably improves the reliability of the answers. On the other hand, by using non-professionals, we will presumably obtain results closer to an average listener experience. Professional musicians are highly trained in listening, in particular to the details of a performance, and thus may deviate from the average listener.

Level	*Subjects*	*Practice (h)*	*Listening (h)*
None	4	0.25	8.75
Beginner	10	0.80	10
Intermediate	9	3.22	14.33
Professional	1	10	40

Table 1. Mean hours practicing and listening to music grouped by self-rated level of musical experience.

3.3 Rating platform and Procedure

A web-based rating platform was developed specifically for this study. It presented the user with nine continuous sliders, one for each perceptual feature. Each endpoint was marked according to descriptions in section 3.1. The same features were rated for all 98 clips, which were presented in random order. The test was done online with the listeners' own headphones. Each listener was initially asked to adjust the volume to a barely noticeable calibration tone, in order to remove listening volume as a factor. In this way, we obtained a rough calibration of the loudness. Each slider was initialized at a neutral value and had to be adjusted before they could continue to the next song. The listeners were allowed to pause and resume the test at will in order to avoid listener fatigue. The resulting numerical values from each slider were in the range -1 to 1 with 101 discrete steps.

4. RESULTS

4.1 Rating accuracy

Listeners' agreement was estimated using the mean pairwise correlation between all subject pairs, see Table 2. As seen in the table, the agreement varied substantially for the different features. The overall variation followed previous experiments [9] but the values were in general lower. Speed had the highest value (0.71) while Harmonic Complexity indicated a value of almost zero (0.09). The agreement was in general lower within the Chinese group than the Swedish group.

Feature	*All*	*Chinese*	*Swedish*
Speed	0.71	0.70	0.76
Rhy. Compl.	0.15	0.10	0.24
Rhy. Clarity	0.19	0.12	0.28
Articulation	0.45	0.33	0.63
Modality	0.18	0.17	0.26
Har. Compl.	0.09	0.07	0.15
Dissonance	0.35	0.32	0.40
Pitch	0.23	0.26	0.22

Table 2. The mean pairwise correlation for each perceptual feature for the Swedish and Chinese group. A value of 1 represents perfect inter-rater agreement and 0 represents no agreement.

The accuracy of the mean estimation was evaluated using Cronbach's alpha, see Table 3. Here we find a rather good

agreement for most features. Speed again had the highest and Harmonic Complexity had the lowest alpha values. We can see again lower values from the Chinese group compared to the Swedish.

Interestingly, the results for Dissonance was quite good both in terms of Cronbach's alpha and in the pairwise correlations, indicating that Dissonance clearly was an easier feature to rate and a more reliable feature than Harmonic Complexity as a measure for harmony perception.

Feature	*All*	*Chinese*	*Swedish*
Speed	0.982	0.967	0.968
Rhy. Compl.	0.796	0.566	0.777
Rhy. Clarity	0.846	0.644	0.800
Articulation	0.951	0.858	0.944
Modality	0.848	0.720	0.806
Har. Compl.	0.679	0.502	0.645
Dissonance	0.921	0.827	0.883
Pitch	0.877	0.817	0.751

Table 3. Cronbach's alpha for each perceptual feature for the Swedish and Chinese group. A value of 1 represents perfect inter-rater agreement and 0 represents no agreement.

	Speed	*Rhythmic Complexity*	*Rhythmic Clarity*	*Articulation*	*Modality*	*Harmonic Complexity*	*Dissonance*
Rhythmic Complexity	0.70 ***						
Rhythmic Clarity	0.77 ***	0.41 ***					
Articulation	0.86 ***	0.62 ***	0.70 ***				
Modality	0.73 ***	0.54 ***	0.65 ***	0.57 ***			
Harmonic Complexity	0.69 ***	0.67 ***	0.42 ***	0.56 ***	0.64 ***		
Dissonance	0.74 ***	0.58 ***	0.46 ***	0.78 ***	0.40 ***	0.45 ***	
Pitch	0.38 ***	0.27 **	0.31 **	0.19	0.55 ***	0.41 ***	0.23 *

Table 4. Cross-correlation among the rated features for the Chinese group. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.2 Cross-correlation of the rated features

The cross-correlations among the rated features were calculated using the average rating in each group, computed over the 98 music examples. The results are shown in Table 4 and 5. Note that no compensation for multiple comparisons was applied. Thus, the significances should be interpreted just as an indication. Most of the correlations are significant, but not extremely high. There can be two sources of this correlation. A database consisting of commercial music will have a natural covariation of the features. The second source of covariation can be related to the subjects that possibly did not judge each feature in isolation. The overall correlation was higher for the Chinese group (mean corr = 0.57) than the Swedish group (mean corr = 0.44), indicating that it was more difficult for the listeners in the Chinese group to separate the different features. For comparison, the cross-correlation for the same features in [9] was considerably lower, and only in about half of the cases reached significance.

	Speed	*Rhythmic Complexity*	*Rhythmic Clarity*	*Articulation*	*Modality*	*Harmonic Complexity*	*Dissonance*
Rhythmic Complexity	0.41 ***						
Rhythmic Clarity	0.68 ***	-0.06					
Articulation	0.89 ***	0.49 ***	0.69 ***				
Modality	0.72 ***	0.11	0.52 ***	0.62 ***			
Harmonic Complexity	0.23 *	0.76 ***	-0.18	0.24 *	-0.13		
Dissonance	0.78 ***	0.45 ***	0.59 ***	0.84 ***	0.45 ***	0.34 ***	
Pitch	0.54 ***	0.16	0.32 **	0.35 ***	0.36 ***	0.14	0.33 ***

Table 5. Cross-correlation among the rated features for the Swedish group. Significance levels, see Table 4.

4.3 Mean values across songs

The mean values of the listeners' ratings were computed across songs for the two groups. The features were examined separately as shown in Figure 2. Significant differences between groups were observed for Speed and Rhythmic Complexity only. More specifically, the Swedish group rated both of these features lower than the Chinese group. A one-way ANOVA was performed for each feature, which yielded significant effects for Speed ($p<0.002$) and for Rhythmic Complexity ($p<0.005$). The remaining features were not significant. As the scale of the ratings was between -1 and 1 all differences were relatively small, regardless of their level of significance.

In order to investigate the differences more in detail, a mixed ANOVA was conducted with Speed as the dependent variable, and with Genre (within), Nationality (between) and Gender (between) as factors. It indicated significant effects for Genre ($p=0.000$), Nationality ($p=0.004$) and the interaction between Genre and Nationality ($p=0.009$). All p-values were compensated using Greenhouse-Geisser correction. All other effects were non-significant. The interaction between Genre and Nationality indicate that hip hop in particular was perceived differently by Chinese and Swedish listeners, as seen in Figure 3.

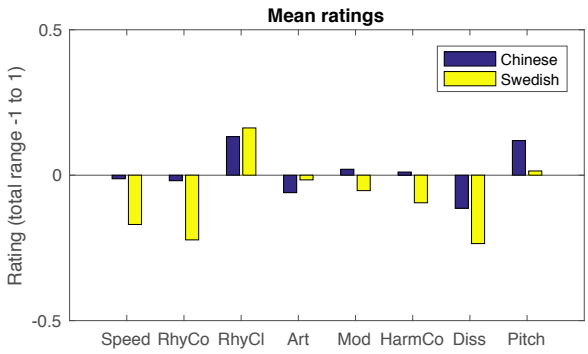

Figure 2. Average ratings for each perceptual feature.

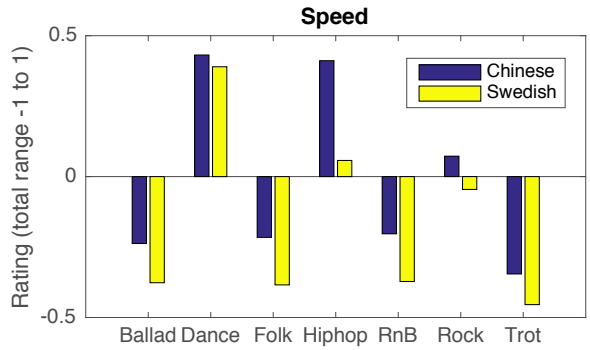

Figure 3. Average ratings for Speed divided in genres.

4.4 Prediction of emotion ratings from perceptual features

In the previous experiments we successfully predicted activity and valence from the rated perceptual features [9]. It was therefore natural to test this again in a similar manner, using the previously collected activity-valence values from [12]. Considering the comparatively low agreement between the listeners in the current experiment and the comparatively few raters per item in [12], we assumed that the prediction in terms of explained variance would be lower.

As independent input parameters, we used the eight rated perceptual features. In addition, since dynamics is well known to be an important factor for emotional expression (e.g. [4]), we also optionally added the data from the computational model of dynamics [5] developed using the data in the previous experiments [9]. This model was able to explain above 80% of rated dynamics in two different data sets consisting of film music clips and popular music. While the ground truth data consisted only of Western music and listeners, the high generalizability of the model indicate that a major part of the variation might be predicted also in the current data.

For the prediction, we used partial least-square regression (PLS) with 10-fold cross-validation. PLS is particularly useful in this case, since there are rather few cases (98) compared to the number of features (9), and since there is a considerable correlation among the features. The number of factors used in the PLS regression was manually varied in order to obtain the highest possible cross-validation results for each prediction. All results were adjusted for the number of factors.

The resulting explained variance R^2 is shown in Table 6 for all subjects, the two groups, and with or without computed dynamics. For the results including dynamics, we see that the prediction of activity is rather good and the result is only slightly decreasing when it is cross-validated. The prediction of valence is moderate and reaches only about 40 % explained variation for all subjects when cross-validated. In comparison, in [9] we obtained cross-validated R^2 = 0.93 and 0.90 for activity and R^2 = 0.88 and 0.75, for valence for two different data sets.

Comparing the cross-validated R^2 for the two groups and all subjects, we see that the best result for activity was obtained when all subjects were included (84%). For valence, the best result was obtained for the Swedish group (41%). For activity both groups had about the same R^2, while for valence there was a drop in R^2 for the Chinese group compared to the Swedish.

Surprisingly, the resulting R^2 only changed slightly when dynamics was omitted. The R^2 for activity even increased slightly while the R^2 for valence decreased. Given the documented importance of dynamics, it is possible that the concept of dynamics was captured to some extent also in the ratings of the other perceptual features. Such covariation among the features are expected to some degree and also indicated in Table 4 and 5.

With Dynamics		*All*	*Chinese*	*Swedish*
Activity	R^2	0.854	0.851	0.831
	R^2 crossval	0.834	0.807	0.807
	Factors	2	3	2
Valence	R^2	0.509	0.406	0.486
	R^2 crossval	0.401	0.245	0.406
	Factors	4	5	3
Without Dynamics				
Activity	R^2	0.857	0.862	0.824
	R^2 crossval	0.841	0.826	0.815
	Factors	2	4	1
Valence	R^2	0.450	0.377	0.464
	R^2 crossval	0.341	0.234	0.387
	Factors	4	6	3

Table 6. The explained adjusted variance (R^2) including cross validation (R^2 crossval), predicting activity and valence from perceptual features using PLS regression.

5. CONCLUSION AND DISCUSSION

Comparing the Chinese and Swedish group, we found small but significant differences of the average ratings for Speed and Rhythmic complexity (Table 6). Difference in such absolute levels of the ratings can be attributed to different listener habits. If a listener is used to listen to rather slow music, the judgments will presumably be biased so that the rating of Speed will be higher than a listener that usually listen to faster music.

Interestingly, the difference in the average speed ratings could be mainly attributed to the difference in the perception of the hip hop genre. The Chinese listeners perceived hip hop to be fast, while the Swedish listeners

perceived it to be in the middle range. One possibility could be that the Chinese group focused on the vocals (often quite fast) while the Swedish group focused on the supporting beat (often in medium tempo).

Dissonance was found to be a good alternative to Harmonic Complexity. Dissonance obtained considerably higher values for Cronbach's alpha and pair-wise correlation than Harmonic Complexity (Table 2, 3).

There were differences between the two groups in the pair-wise correlation among subjects (Table 2), Cronbach's alpha (Table 3), the cross-correlation among the features (Table 4,5), and the prediction of emotion (Table 6). These differences all indicate that it was a bit harder for the Chinese group to rate the features. The origin of these differences is unclear but could be due to differences in musical experience, problems understanding the instructions, or the use of an online test.

Despite the comparatively lower agreement and emotion predictability, a testing of a new interface for music browsing revealed that the average values across all listeners seems to be effective for selecting songs according to emotion [14]. Further evaluation and development will be undertaken to validate these preliminary results.

6. ACKNOWLEDGEMENTS

We would like to thank the listeners for their participation and the help from the Chinese student organization at KTH. This work was supported by an exchange grant from UIUC and KTH within the Inspire program and the Swedish Research Council, Grant No. 2012-4685.

7. REFERENCES

[1] J. J. Burred and A. Lerch: “Hierarchical automatic audio signal classification,” *The Journal of the Acoustical Society of America*. Vol. 52, No. 7/8, pp. 724–738, 2004.

[2] E. F. Clarke: *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford University Press (OUP), Oxford, 2005.

[3] D. Deutsch: "The tritone paradox: An influence of language on music perception." *Music Perception,* Vol. 8, No. 4, pp. 335-347, 1991.

[4] T. Eerola, A. Friberg and R. Bresin: ”Emotional expression in music: contribution, linearity, and additivity of primary musical cues,” *Frontiers in Psychology*, Vol. 4, No. 487, pp. 1-12, 2013.

[5] A. Elowsson and A. Friberg: “Predicting the perception of performed dynamics in music audio with ensemble learning,” Journal of the Acoustical Society of America, Vol. 141, No. 3, pp. 2224-2242, 2017.

[6] A. Friberg: “Music listening from an ecological perspective,” Poster presented at the 12th ICMPC and the 8th ESCOM, Thessaloniki, 2012.

[7] A. Friberg and A. Hedblad: ”A Comparison of Perceptual Ratings and Computed Audio Features,” *Proceedings of the 8th Sound and Music Computing Conference*, pp. 122-127, 2011.

[8] A. Friberg, E. Schoonderwaldt, and A. Hedblad: “Perceptual ratings of musical parameters,” *Loesch and Weinzierl (eds.): Gemessene Interpretation: computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen*, pp. 237-253, Schott, 2011.

[9] A. Friberg, E. Schoonderwaldt, A. Hedblad, M. Fabiani and A. Elowsson: “Using listener-based perceptual features as intermediate representations in music information retrieval,” *The Journal of the Acoustical Society of America*, Vol. 136, No. 4, pp. 1951-1963, 2014.

[10] A. Gabrielsson and E. Lindström: ”The role of structure in the musical expression of emotions,” *P. N. Juslin and J. A. Sloboda (eds.): Handbook of music and emotion: Theory, research, applications*, pp. 367-400, OUP, New York, 2010.

[11] X. Hu and J. H. Lee: “A cross-cultural study of music mood perception between American and Chinese listeners,” *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, pp. 535–540, 2012.

[12] X. Hu, J. H. Lee, K. Choi, J. S. Downie: “A cross-cultural study of mood in K-POP songs,” *Proceedings of ISMIR*, pp. 385–390, 2014.

[13] J. H. Lee, K. Choi, X. Hu and J. S. Downie: “K-Pop genres: A cross-cultural exploration,” *Proceedings of ISMIR*, pp. 529–534, 2013.

[14] R. Nysäter: *Music discovery methods using perceptual features*. Master thesis KTH, 2017.

[15] R. Schön: “Vertical dissonance as an alternative harmonic-related perceptual feature,” Technical report, KTH, Stockholm, 2015.

[16] R. Schön: "A cross-cultural listener-based study on perceptual features in K-pop," Master thesis, KTH, Stockholm, 2015.

[17] X. Serra: "A multicultural approach in music information research," *Proceedings of ISMIR*, pp. 151–156, 2011.

[18] W. F. Thompson and L-L. Balkwill: “Cross-cultural similarities and differences,” *J. Sloboda (ed.): Handbook of music and emotion: Theory, research, applications*, pp. 755-788, OUP, Oxford, 2010.

[19] Y-H. Yang and X. Hu: “Cross-cultural music mood classification: A comparison on English and Chinese songs,” *Proceedings of the International Symposium on Music Information Retrieval*, pp. 19–24, 2012.