PROJECTION OF ACOUSTIC FEATURES TO CONTINUOUS VALENCE-AROUSAL MOOD LABELS VIA REGRESSION

Erik M. Schmidt **Youngmoo E. Kim**
MET-lab, Drexel University
Department of Electrical and Computer Engineering
{eschmidt,ykim}@drexel.edu

ABSTRACT

The problem of organizing music by emotional content or mood is not only difficult to solve computationally, but often lacks well-defined answers. In previous work, we have presented a collaborative game, MoodSwings [1], which records dynamic (per-second) labels of players ratings of music using the valence-arousal model.

Using a small subset of the MoodSwings data, we are currently investigating the projection of various acoustic features to valence-arousal point values using regression, as opposed to discretizing emotional space into a finite number of classes [2]. We demonstrate preliminary results that indicate the effectiveness of the regression-based approach in taking advantage of the continuous range of the underlying valence-arousal space. Our data collection consists of 120, 15-second song clips, which have been selected a priori to approximate an even distribution across the four primary quadrants of the valence-arousal space.

Using least-squares regression, the system is trained to project the mean of the acoustic features to the mean valence-arousal value for each 15-second music clip. Using a combination of MFCCs and spectral shape features we show that the least-squares projection results in an average deviation of only 16.03% from the mean labels of the testing samples. We compare the Euclidean distances from the projected valence-arousal points to the mean collected labels (which are assumed to represent ground truth) to baseline distances resulting from a random permutation of the ground truth. Comparing these cases over 50 cross-validations, we compute the Student's T-test to demonstrate the statistical significance of our results.

Feature	Avg. Distance	Avg. Rand. Dist.	T-test
MFCC	0.180 ± 0.015	0.273 ± 0.020	26.610
S. Shape	0.179 ± 0.014	0.256 ± 0.018	24.292
S. Contrast	0.161 ± 0.014	0.274 ± 0.024	28.614
Chroma	0.233 ± 0.016	0.241 ± 0.017	2.5547
MFCC + S.S.	0.160 ± 0.012	0.278 ± 0.024	31.010

Table 1. Emotion regression of MoodSwings Data.

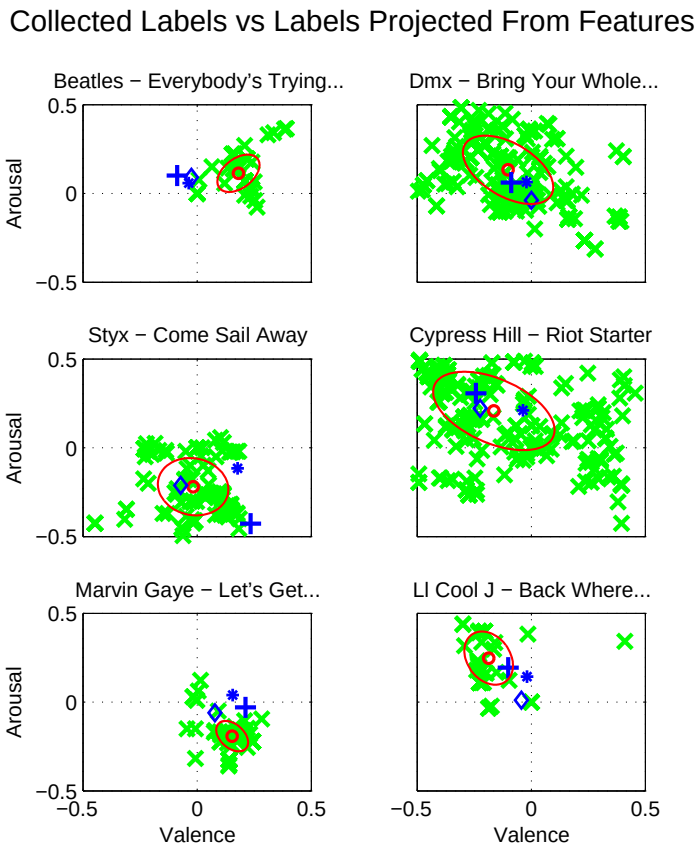

Figure 1. Labels collected for song (x), μ of collected labels (red o), σ of collected labels (red ellipse), MFCC-only projection (*), spectral shape projection ($\diamond$), and MFCC+spectral shape projection (+).

1. REFERENCES

[1] Y. Kim, E. Schmidt, and L. Emelle. Moodswings: A collaborative game for music mood label collection. In *Proc. International Conference on Music Information Retrieval*, Philadelphia, PA, September 2008.

[2] L. Lu, D. Liu, and H. J. Zhang. Automatic mood detection and tracking of music audio signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):5–18, 2006.