Quantify Music Artist Similarity Based on Style and Mood

Bo Shao, Tao Li
School of Computer Science
Florida International University
Miami, FL 33199
{bshao001,taoli}@cs.fiu.edu

Mitsunori Ogihara
Department of Computer Science
University of Miami
Coral Gables, FL 33134
ogihara@cs.miami.edu

ABSTRACT

Music artist similarity has been an active research topic in music information retrieval for a long time since it is especially useful for music recommendation and organization. However, it is a difficult problem. The similarity varies significantly due to different artistic aspects considered and most importantly, it is hard to quantify. In this paper, we propose a new framework for quantifying artist similarity. In the framework, we focus on style and mood aspects of artists whose descriptions are extracted from the authoritative information available at the All Music Guide website. We then generate style–mood joint taxonomies using hierarchical co-clustering algorithm, and quantify the semantic similarities between the style/mood terms based on the taxonomy structure and the positions of these terms in the taxonomies. Finally we calculate the artist similarities according to all the style/mood terms used to describe them. Experiments are conducted to show the effectiveness of our framework.

Categories and Subject Descriptors

H.5.5 [**Information Systems**]: Sound and Music Computing; H.3.3 [**Information Systems**]: Information Search and Retrieval

General Terms

Algorithms, Experimentation

Keywords

Music Artist Similarity, hierarchical co-clustering, Similarity Quantification

1. INTRODUCTION

Music artist similarity has been an active research topic in music information retrieval for a long time since it is especially useful for music recommendation and organization [10, 19]. Many characteristics can be brought into consideration for defining similarity, e.g., sound, lyrics, genre, style, and mood. Methods for calculating artistic similarity include recent proposals that are based on the similarity information provided by the All Music Guide website (`http://www.allmusic.com`) as well as those based on the user access history (e.g., see [10]). Although there has been considerable effort into developing effective and efficient method for calculating artist similarity, several challenges still exist. First, artist similarity varies considerably when considering different aspects of artists such as genre, mood, style, culture, and acoustics. Second, the user access history data are often very sparse and hard to acquire. Third, even if we can obtain the categorical descriptions of two artists using All Music Guide, comparing the descriptions is not trivial since there are semantic similarities among different descriptions. For example, given two mood terms *witty* and *thoughtful*, we cannot simply quantify their similarity as 0 just because they are different words or as 1 because they are synonyms.

In this paper, we propose a new framework for quantifying artist similarity. In this framework, we focus on two very important aspects of music: style and mood [15]. The style and mood descriptions of famous artists are publicly available on All Music Guide website. We collect the information of the artists and their style and mood descriptions. The All Music style terms are nouns and adjectives while its mood terms are adjectives only. These terms carry significant linguistic meanings given some context, but the use of the terms at the All Music web site is little contextual. In this paper, we study how these terms are collectively used in describing artists. To capture the semantic similarity among different style and mood descriptions, we generate a style taxonomy and a mood taxonomy using hierarchical co-clustering algorithm. Then we quantify the semantic similarities between the style/mood terms based on the taxonomy structure and the positions of these terms in the taxonomies. Finally we calculate the artist similarities according to all the style/mood terms used to describe them.

WIDM'08, October 30, 2008, Napa Valley, California, USA.

2. SIMILARITY TAXONOMY GENERATION

The style and mood labels of 2431 artists are collected for those artists having both labels appearing in the All Music Guide website. Altogether 601 style terms (nouns like *Electric Chicago Blues*, *Greek Folk*, and *Chinese Pop*, as well as adjectives like *Joyous*, *Energetic*, and *New Romantic*), and 254 mood terms (such adjectives as *happy*, *sad*, and *delicate*) are used to describe these artists. Table 1 lists an example of the mood and style descriptions of two randomly picked artists: ABBA and The Beatles. The mood and style are subjective. However, we view All Music Guide as representing collective opinions of many music experts/critics thereby representing the subjective opinions of a large proportion of music listeners.

In order to organize the style terms and mood terms into the corresponding taxonomies, we need to apply clustering algorithms to them. Clustering is the problem of partitioning a finite set of points in a multi-dimensional space into classes (called clusters) so that (i) the points belonging to the same class are *similar* and (ii) the points belonging to different classes are *dissimilar* [1, 2].

However, most clustering algorithms aim at clustering *homogeneous data*, i.e, the data points of a single type [3]. In our application, the data set to be analyzed involves more than one type, e.g. *styles* and *artists*. Furthermore, there are close relationships between these types of data. It is difficult for the traditional clustering algorithms to utilize those relationship information efficiently.

Co-clustering algorithms are designed to cluster different types of data simultaneously by making use of the dual relationship information such as mood–artist matrix. For instance, *Dhillon* [6] and Zha et al [8] proposed bipartite spectral graph partitioning approaches to co-cluster words and documents, *Cho. et al* [4] proposed to co-cluster the experimental conditions and genes for microarray data by minimizing the *Sum-Squared Residue*, *Long et al.* [5] proposed a general principled model, called *Relation Summary Network*, to co-cluster the heterogeneous data on a *k-partite graph*.

To further utilize the cluster information obtained from the co-clustering algorithms and generate the taxonomies we needed in our system, we have to use the hierarchial co-clustering algorithms [16].

In this application, the artist style description is represented as a 2431×601 artist–style matrix, S, and the artist mood description as a 2431×254 artist–mood matrix, M. In the following, we will describe our algorithm for the artist–style matrix S. The algorithm is the same for the artist–mood matrix M. The core idea behind the procedure is to combine Singular Value Decomposition (SVD) and K-means using a top-down iterative process [16]. The procedure is described as follows:

1. Given an $m \times n$ artist–style matrix, S, perform SVD on S to obtain: $S = U \times \Lambda \times V^T$.

2. Let $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_m$ be the largest m singular values. Then the number of clusters is k where:

$$k = argmax_{(m \geqslant i > 1)}(\lambda_{i-1} - \lambda_i)/\lambda_{i-1}$$

3. Find k singular vectors of S: $u_1, u_2, \ldots, u_k$ and $v_1, v_2, \ldots, v_k$, and then form a matrix Z by:

$$Z = \begin{bmatrix} D_1^{-1/2}[u_1, ..., u_k] \\ D_2^{-1/2}[u_1, ..., u_k] \end{bmatrix}$$

4. Apply K-means clustering algorithm to cluster Z into k clusters.

5. For each cluster, check the number of artists in it. If the number is higher than a given threshold (in our experiment, we set the threshold = 3), construct a new artist–style matrix formed by the artists and styles in that cluster, and continue to the first step.

According to this algorithm, 601 style terms are clustered into 20 clusters, and 254 mood terms are clustered into 68 clusters. They are further recursively clustered into many subclasses until the algorithm converges. We organized the generated taxonomies and present them in two trees, which can be viewed at `http://www.newwisdom.net/MIR/styletree.jsp` and `http://www.newwisdom.net/MIR/moodtree.jsp`.

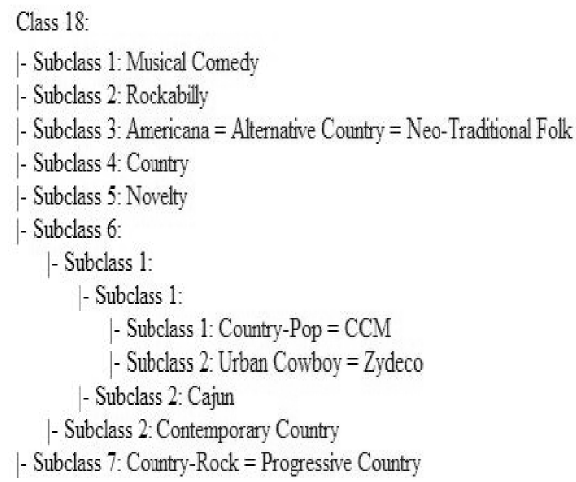

Figure 1: A sample cluster content from style similarity tree. (= means the most similar)

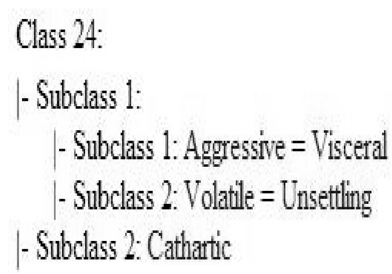

Figure 2: A sample cluster content from mood similarity tree. (= means the most similar)

Figure 1 is a sample cluster obtained from the style similarity tree. In this cluster, we observe that *Country-Rock* and *Progressive Country* are the most similar (similarity value between them equals to 1 in our system) in style description, and the similarity between *Country-Pop* and *Urban Cowboy* is greater than the similarity between *Country-Pop* and *Cajun* as well as the similarity between *Urban Cowboy* and *Cajun*. Figure 2 is a sample cluster obtained from the mood similarity tree following the same construction rule.

Figure 3 shows the distribution of the sizes of all the 20 style clusters and Figure 4 shows the distribution of the sizes

Artist	Mood Description	Style Description
ABBA	Light, Delicate, Rousing, Sentimental, Joyous, Fun, Sweet, Sparkling, Sugary, Cheerful, Happy, Playful, Naive, Plaintive, Gentle, Gleeful, Giddy, Stylish, Romantic, Energetic, Exuberant, Ambitious, Complex, Exciting, Fun, Bright, Lively, Witty, Carefree, Happy, Sentimental, Wistful	Euro-Pop, Pop/Rock, Swedish Pop/Rock, Pop, British Invasion, Psychedelic
The Beatles	Wistful, Searching, Sweet, Warm, Yearning, Whimsical, Amiable/Good-Natured, Poignant, Lush, Laid-Back/Mellow, Literate	Merseybeat, Pop/Rock, British Psychedelia, Folk-Rock, Rock & Roll

Table 1: An example of Artist Mood and Style Description

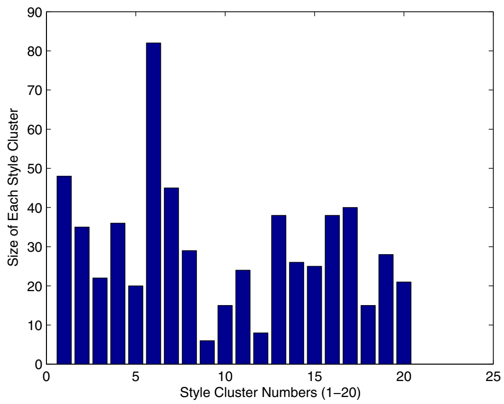

Figure 3: The distribution of sizes of style clusters

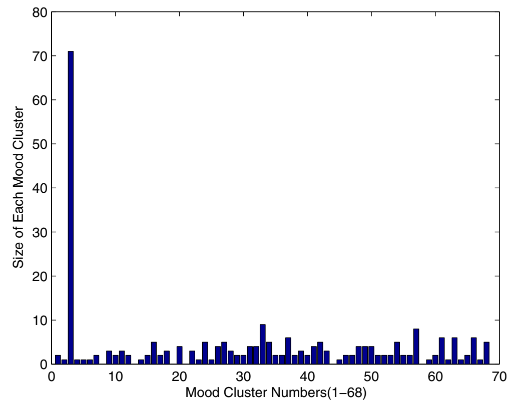

Figure 4: The distribution of sizes of mood clusters

of all the 68 mood clusters. From these two figures, we observe that both style terms and mood terms are distributed into each classes in a quite balanced manner.

Based on this co-clustering algorithm, we can also obtain the style-based artist similarity structure and mood-based artist similarity structure directly, which can be viewed at `http://www.newwisdom.net/MIR/artisttrees.jsp` and `http://www.newwisdom.net/MIR/artisttreem.jsp`. They have the similar well-balanced cluster member distributions. However, the similarity between two artists is not quantified. Furthermore, if we have new artists with style and/or mood descriptions, it is very hard for us to integrated them into the tree structures. So we need to go steps further as we will describe in details below.

3. SIMILARITY QUANTIFICATION

To calculate artist similarity, we need to quantify the semantic similarity between all pairs of style/mood terms first. In order to do this, we investigate the methods proposed by Resnik [13], Jiang and Conrath [11], Lin [12], and Schlicker et al. [9]. The approaches for calculating the similarity proposed by them are briefly described as follows:

Resnik:

$$sim_R(s_1, s_2) = \max_{s \in S(s_1,s_2)} \{-\log(p(s))\} \tag{1}$$

Jiang-Conrath:

$$dist_{JC}(s_1, s_2) = \max_{s \in S(s_1,s_2)} \{2\log(p(s)) - \log(p(s_1)) - \log(p(s_2))\} \tag{2}$$

Lin:

$$sim_L(s_1, s_2) = \max_{s \in S(s_1,s_2)} \{\frac{2 \times \log(p(s))}{\log(p(s_1)) + \log(p(s_2))}\} \tag{3}$$

Schlicker:

$$sim_L(s_1, s_2) = \max_{s \in S(s_1,s_2)} \{\frac{2 \times \log(p(s))}{\log(p(s_1)) + \log(p(s_2))}(1 - \log(p(s)))\} \tag{4}$$

Here $p(s) = freq(s)/N$ and $freq(s)$ is the number of artists that utilize the given style/mood term s to describe them, N is total number of artists, and $S(s_1, s_2)$ is the set of common subsumers of style/mood terms s_1 and s_2. The basic idea of these approaches is to capture the specificity of each style/mood term and to calculate the similarity between style/mood terms that reflects their positions in the taxonomy generated in Section 2.

Once we obtain the pairwise semantic similarity of style/mood terms, we can calculate the artist similarity based on style/mood. For example, if artist A_1 is described by a group of styles $s_1, s_2, \ldots, s_i$, and artist A_2 is described by another group of styles $s'_1, s'_2, \ldots, s'_j$, we define the style-based similarity between A_1 and A_2 as:

$$sim(A_1, A_2) = \frac{\sum_{x\in[1,i]}(max_{y\in[1,j]} sim(s_x, s'_y))}{j} \quad (5)$$

Here $sim(s_x, s'_y)$ is the similarity between style s_x and style s'_y. Mood-based artist similarity can be obtained using the same approach.

In some applications, people may see the differences among these four different approaches due to the different scales of their results and the different ways they are associating with the terms in the taxonomies. In our system, however, we compared their results and do not see any significant differences among them after normalizing them into the same scale (0∼1). To further illustrate this, let us check the data distribution of the artist similarity values generated using these four different approaches.

The distribution of artist similarity values based on style similarity calculated using the four different approaches is presented in Figure 5, and the distribution of artist similarity values based on mood similarity calculated using the four different approaches is presented in Figure 6. From these two figures, we observe that there are almost no difference among the distributions of the artist similarity values using 4 different approaches described above. Hence we use the average of all the 4 normalized quantified similarity values as the final artist similarity. We also observe that the style-based artist similarity values are a little more diverse than the mood-based artist similarity values, therefore we use a heuristic proportion value to calculate the final combined artist similarity value:

$$c = 0.4 \times m + 0.6 \times s \quad (6)$$

where, c is the combined artist similarity, and m is mood-based similarity while s is style-based similarity. In our system, 0 stands for the most different and 1 stands for the most similar.

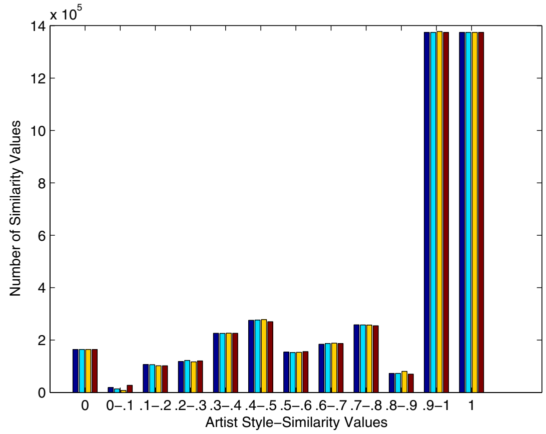

Figure 5: The distribution of artist similarity values based on style similarity calculated using the four different approaches

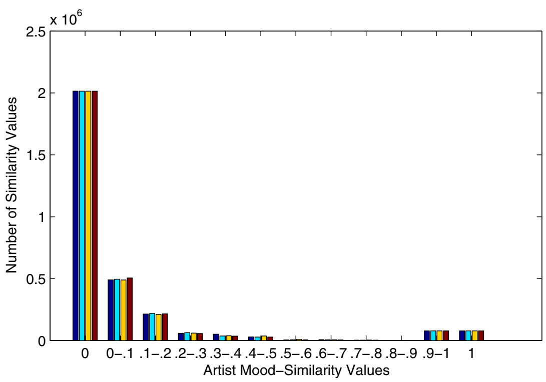

Figure 6: The distribution of artist similarity values based on mood similarity calculated using the four different approaches

4. EVALUATION

Our interests are in how these *professionally assigned* mood and style terms are grouped together in describing artists. We believe that neither acoustic similarity nor mood/style labels provide sufficient information to enable accurate similarity calculation. We are rather interested in how related the label-based similarity and the acoustics-based similarity are to each other. To explore more on this question, it would be ideal if we had acoustics data for all the 2431 artists in the study, but the time and cost required for collecting the data would be prohibitive. So for the experimental study in the section, we chose to look at a limited number of artists. We present a case study on six famous artists (bands): the Beatles, the Carpenters, Celine Dion, Elvis Presley, Madonna, and Michael Jackson to demonstrate the effectiveness of our framework. The quantified artist similarities among them are listed in the second, third, and fourth columns (mood-based, style-based, and the combined similarity) of Table 2.

To compare with the artist similarity, we propose to use the distances of the acoustic features extracted from the songs of these artists (bands). In the following subsection, we will briefly explain how the acoustic features are

Name Pair	Mood-based	Style-based	Combined Similarity	Average Distance
Elvis Presley : Michael Jackson	0.33	1	0.732	4.807
The Carpenters : Celine Dion	0.15	1	0.66	4.836
Michael Jackson : Madonna	0	1	0.6	6.840
The Carpenters : Michael Jackson	0	1	0.6	7.921
Celine Dion : Michael Jackson	0	1	0.6	7.991
Elvis Presley : Madonna	0	1	0.6	8.555
The Beatles : Michael Jackson	0	0.875	0.525	9.455
Celine Dion : Madonna	0	0.75	0.45	8.229
The Carpenters : Madonna	0.143	0.5	0.357	8.344
Celine Dion : Elvis Presley	0	0.75	0.45	8.655
The Carpenters : Elvis Presley	0.048	0.5	0.319	8.756
The Beatles : Madonna	0	0.5	0.3	9.688
The Beatles : Elvis Presley	0	0.5	0.3	9.324
The Beatles : Celine Dion	0	0.278	0.167	10.887
The Carpenters : The Beatles	0	0.25	0.15	11.134

Table 2: Quantified Similarity Values and Average Distance

extracted and why there are a suitable references to check with.

4.1 Music Feature Extraction

There has been a considerable amount of work in extracting descriptive features from music signals for music genre classification and artist identification. In our study, we use timbral features along with wavelet coefficient histograms. The feature set consists of the following three parts and total 80 features, which can well reflect the moods and styles of the corresponding artists [17, 14, 18, 20].

4.1.1 Mel-Frequency Cepstral Coefficients (MFCC)

Mel-Frequency Cepstral Coefficients (MFCC) is a feature set that is highly popular in speech processing. It is designed to capture short-term spectral-based features. The features are computed as follows: First, for each frame, the logarithm of the amplitude spectrum based on short-term Fourier transform is calculated, where the frequencies are divided into thirteen bins using the Mel-frequency scaling. Next, this vector is decorrelated using discrete cosine transform. This is the MFCC vector. In this work, we use the first five bins, and compute the mean and variance of each over the frames.

4.1.2 Short-Term Fourier Transform Features (STFT)

This is a set of features related to timbral textures and is not captured using MFCC. It consists of the following five types: Spectral Centroid, Spectral Rolloff, Spectral Flux, Zero Crossings, and Low Energy. More detailed descriptions of STFT can be found in [14].

4.1.3 Daubechies Wavelet Coefficient Histograms (DWCH)

Daubechies wavelet filters are a set of filters that are popular in image retrieval. The Daubechies Wavelet Coefficient Histograms, proposed in [20], are features extracted in the following manner: First, the Daubechies-8 (db_8) filter with seven levels of decomposition (or seven subbands) is applied to 30 seconds of monaural audio signals. Then, the histogram of the wavelet coefficients is computed at each subband. Then the first three moments of a histogram, i.e., the average, the variance, and the skewness, are calculated from each subband. In addition, the subband energy, defined as the mean of the absolute value of the coefficients, is computed from each subband. More details of DWCH can be found in [20, 19].

4.2 Result Analysis

For each artist (band), we randomly pick 5 songs and conduct the following procedure. Firstly, we exact the acoustic features of each song using the approach explained above. Then we calculate the pairwise Euclidean distances between the feature points that represent the songs of different artists (bands). Finally we calculate the average of all the pairwise distances as the average distance of the two artists. The results are listed in the last column of Table 2.

From the results, we observe that our quantified artist similarities match very closely the artist similarity based on the acoustic features of their music recordings. By checking the last two columns of Table 2, we can easily observe that the data variation trends from the top to the bottom, i.e, while the average distance increases one by one, the combined similarity decreases almost constantly. In other words, the acoustic feature points of songs from the artists with higher similarity values (e.g., The Carpenters versus Celine Dion) are closer than those of songs from the artists with lower similarity values (e.g., The Beatles versus Celine Dion, and The Beatles versus The Carpenters), while the acoustic feature points of songs from the artists with lower similarity values (e.g., Elvis Presley and The Beatles) are farther than those of songs from the artists with higher similarity values (e.g., Elvis Presley and Michael Jackson).

5. CONCLUSION

Music artist similarity has been an active research topic in music information retrieval for a long time since it is especially useful for music recommendation and organization. But artist similarity varies from different aspects considered, and hard to quantify although considerable efforts have been put into this venue. In this paper, we focus on two very important aspects of musical artists: style and mood. we extract authoritative information from All Music Guide, generate style and mood similarity taxonomies, and quantify the artist similarities based on the semantic similarities of the style and mood terms. We also conduct a case study, which shows the effectiveness of this proposed framework.

Acknowledgments

The work of T. Li is partially supported by NSF under IIS-0546280, HRD-0317692, and IIP-0450552.

6. REFERENCES

[1] J. Hartigan. *Clustering Algorithms*. Wiley, 1975.

[2] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, 1988.

[3] P. Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA, 2002.

[4] H. Cho, I. Dhillon, Y. Guan, and S. Sra. Minimum sum squared residue co-clustering of gene expression data. In *Proceedings of The 4th SIAM Data Mining Conference*, pages 22–24, April 2004.

[5] B. Long, X. Wu, Z. M. Zhang, and P. S. Yu. Unsupervised learning on k-partite graphs. In *Proceedings of ACM SIGKDD (KDD'06)*, pages 317–326, 2006.

[6] I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. *Proceeding of ACM SIGKDD (KDD'01)*, 2001.

[7] I. S. Dhillon, S. Mallela, and D. S. Modha. Information-theoretical co-clustering. In *Proceedings of ACM SIGKDD (KDD'03)*, pages 89–98, 2003.

[8] H. Zha, X. He, C. Ding, M. Gu, and H. Simon. Bipartite graph partitioning and data clustering. *Proceedings of ACM CIKM (CIKM'01)*, 2001.

[9] J. R. A. Schlicker, F. S. Domingues and T. Lengauer. A new measure for functional similarity of gene products based on gene ontology. In *BMC Bioinformatics*, volume 7, pages 302–317, June 2006.

[10] A. B. D Ellis, B Whitman and S. Lawrence. The quest for ground truth in musical artist similarity. In *ISMIR*, 2002.

[11] J. J. Jiang and D. W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. In *In ROCLING X*, 1997.

[12] D. Lin. An information-theoretic definition of similarity. In *ICML*, 1998.

[13] P. Resnik. Using information content to evaluate semantic similarity in a taxanomy. In *IJCAI*, 1995.

[14] G. Tzanetakis and P. Cook. Music genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[15] A. Uitdenbogerd and R. van Schyndel. A review of factors affecting music recommender success. In *ISMIR*, 2002.

[16] G. Xu and W.-Y. Ma. Building implicit links from content for forum search. In *SIGIR*, 2006.

[17] J. Foote and S. Uchihashi. The beat spectrum: a new approach to rhythm analysis. In *IEEE ICME*, 2001.

[18] B. Logan and A. Salomon. A content-based music similarity function. Technical Report CRL 2001/02, Cambrige Research Laboratory, June 2001.

[19] T. Li and M. Ogihara. Towards Intelligent Music Retrieval. In *IEEE Transactions on Multimedia*, 8(3): 564-574 (2006).

[20] T. Li, M. Ogihara, and Q. Li. A comparative study on content-based music genre classification. In *SIGIR'03*, pages 282–289, 2003.