Music Summarization Via Key Distributions: Analyses of Similarity Assessment Across Variations

Arpi Mardirossian and Elaine Chew
University of Southern California Viterbi School of Engineering
Epstein Department of Industrial & Systems Engineering
Integrated Media Systems Center
Los Angeles, CA 90089, USA
{mardiros, echew}@usc.edu

Abstract

This paper presents a computationally efficient method for quantifying the degree of tonal similarity between two pieces of music. The properties we examine are key frequencies and average time in key, and we propose two metrics, based on the L^1 and L^2 norms, for quantifying similarity using these descriptors. The methods are applied to 711 classical themes and variations over 71 variation sets by 10 composers of different genres. Quantile-quantile plots and the Kolmogorov-Smirnov measure show that the proposed metrics exhibit strongly distinct behaviour when assessing pieces from the same variation set, and those that are not. Comparisons across variation sets by the same composer, and comparisons of pieces by different composers although result in similar distributions, are derived from fundamentally different underlying distributions, according to the K-S measure. We present probabilistic analyses of the two methods based on the distributions derived empirically. When the discrimination threshold is set at 55, the probabilities of Type I and Type II errors are 18.41% and 20.56% respectively for Method 1, and 15.72% and 22.94% respectively for Method 2. Method 1 has a success rate of 99.48% when labeling pieces as dissimilar (not from the same variation set), while the corresponding rate for Method 2 is 99.45%.

Keywords: Music similarity, similarity assessment, music representation, music summarization, key distribution, pitch, music information retrieval.

1. Introduction

This paper presents a computationally efficient method for determining tonal similarity between two pieces of music. The information required by the system is pitch information for any time segment, which can be derived from MIDI or audio. We focus on the properties of key frequencies and the average time in each key, and propose similarity metrics based on these descriptors.

Music similarity is a complex problem because the definition of similarity can be widely divergent and highly subjective. Music similarity has been viewed from many angles with different assumptions. Some aspects of similarity include: instrumentation, timbre, melody, harmony, rhythm, tempo, mood, lyrics, socio-cultural backgrounds, structure, and complexity [1].

Subsequently, a challenge in music similarity research is the determining of appropriate ground truth data. In this paper, we have chosen variation sets as our ground truth information on which to verify our proposed metrics for similarity assessment. The theme and variations genre consists of music in which an initial melody, the theme, is first presented in an introductory section; it is then altered as variations to the original theme in subsequent sections. We will refer to each set of theme and variations as the "Variation Set."

2. Related Work

Any study of music similarity must first define its subject of focus, whether it be low- or high-level, melodic or rhythmic, or in linear or vertical time. We present here some recent work that spans several representative domains. Our work differs from these approaches in that it focuses on pitch structure at a relatively high level, allowing for more general classification based on vectors describing key frequency and average time-in-key information.

One domain of music similarity research is melody. The melody is often the "star" of a piece. It is what we often remember about a song. Hu, Dannenberg and Lewis [2] used dynamic programming algorithms to compute an 'edit distance' as a measure of melodic dissimilarity. Typke et al [3] developed a method where notes were mapped to weighted points in two-dimensional space, and melodic similarity was measured using the Earth Mover's Distance and the Proportional Transportation Distance.

Another domain in music similarity research is rhythm, the pattern of proportional durations of notes. Paulus and Klapuri [4] developed a system that measures the similarity of two rhythmic patterns by using a probabilistic musical meter estimation process. Hofmann-Engl [5] represent durations as chains based on atomic beats. They derived rhythmic similarity from how much two rhythms deviate in shape. Chew, Volk and Lee [6] used the method of Inner Metric Analysis to compute a metric, and Dixon, Pampalk and Widmer [7] used periodicity patterns, to assess rhythmic similarity for dance music classification.

Work that is most closely related to the present one with regards to the domain focus and methods used is that by Tzanetakis, Ermolinskyi and Cook [8]. Their method creates pitch histograms, and then extracts several features from them for genre classification. They mention that pitch histograms may be utilized in the determining of music similarity. Our work goes one step further to consider key histograms, and their behaviour under different test situations. Since each key can be summarized as a pitch distribution, our approach essentially considers the distribution of pitch distributions.

In [9], we introduced the use of key distributions in measuring similarity, and a sum-of-squared-difference metric for quantifying similarity, and tested it on a limited set of Mozart variations, showing the results in a self-similarity matrix. In the current paper, we use an L^1 metric for key distribution similarity assessment, and provide in-depth probabilistic and statistical analyses of the outcomes of this method. We consider the additional statistic – the mean time in key – and use the L^2 norm for quantifying similarity for (key distribution, mean-time-in-key) pairs. The present test data set is also vastly increased from the previous one, and now contains 711 variations from 71 Variation Sets by 10 composers.

Work that is related to ours with regards to the use of Variation Sets is that of Pickens [10]. This work develops a music information retrieval system where given a piece of music as input, it returns an ordered list of similar pieces.

3. System Description

This section describes the techniques we use to assess music similarity. We begin by slicing a piece of music into uniform segments and determining the key for each slice using a key-finding algorithm. We then generate a key histogram as well as a mean-time-in-key histogram for the piece. These histograms summarize the tonal patterns in the piece. The process repeated for the comparison piece, we compare result from the two pieces. We present two methods of this comparison. This section concludes with a short example.

3.1 Segmentation

Each piece, say of length n, is segmented into a given number of segments, m, of uniform length. It follows that the length of each segment is n/m. We refer to a complete sample of music as a "piece". In our experiments, this typically refers to either the theme section of a Variation Set, or one of its variations. When comparing pieces of differing lengths, m, remains constant while the length of each segment depends on n. m is constant so that the summary description of different performances of the same piece will be approximately the same. In our experience, the choice of m has some effect on the final result, but is reasonably stable over a range of m values. If m is very small, then each segment will be too large to provide reasonable discriminatory information. If m is very large, then each segment will be too small to produce any meaningful high-level pitch structure information. The selection of m will be further discussed in Section 4.

3.2 Key Determination

The key of each segment must be determined in order to generate the key distribution of each piece. Any key finding algorithm may be invoked at this stage (see [11] for references to key finding algorithms). We use the Center of Effect Generator algorithm [12,13] based on the Spiral Array, a mathematical model for tonality that uses nested helixes to represent tonal elements, such as pitch classes and keys. The pitches in each segment of a piece are mapped to pitch class positions on the helix using a pitch spelling algorithm [14]. An aggregate position of these positions is obtained by weighting each pitch class representation by its proportional duration in the segment. The key is then determined through a nearest neighbor search for the nearest key representation on the major and minor key helixes. This key finding algorithm can be used for both MIDI and audio input [11],[13],[15]. Even though we focus here on MIDI input, it is easy to see how our approach may be extended and used for audio input.

3.3 Key Histograms

We use the sequence of keys calculated for the segments to generate the key histograms. We represent the sequence of keys as an m-dimensional vector $\mathbf{K} = \{k_1, k_2, \ldots, k_m\}$. Each k_i is a key identified by the key finding algorithm for segment i. The bins of the key histogram are the 55 possible major and minor keys from Cbb to C##, shown as a vector of pitch names, $\mathbf{P} = \{p_1, p_2, \ldots, p_{55}\}$. $\mathbf{P}$ has 55 elements because the Spiral Array does not assume enharmonic equivalence. The key histogram values are stored in the vector $\mathbf{F} = \{f_1, f_2, \ldots, f_{55}\}$ where f_i represents the number of times an element of $\mathbf{K}$ is equal to the i–th element of $\mathbf{P}$.

Let us consider a simple example. If there were only two possible keys (A and B), we would have $\mathbf{P} = \{A, B\}$. Assume that $m = 5$ and the sequence of key segments is $\mathbf{K} = \{A, A, B, B, A\}$. Then $\mathbf{F} = \{3, 2\}$.

3.4 Mean-Time-In-Key Histograms

We use the vector $\mathbf{K}$ to generate the Mean-Time-In-Key histograms. Let $\mathbf{O} = \{o_1, o_2, \ldots, o_{55}\}$ be a vector such that o_i is the number of times a continuous sequence of elements corresponding to p_i occurs in the vector $\mathbf{K}$. The mean-time-in-key histogram is stored in the vector $\mathbf{M} = \{m_1, m_2, \ldots, m_{55}\}$, where $m_i = f_i / o_i$. Continuing with our previous example, $\mathbf{O} = \{2, 1\}$ and $\mathbf{M} = \{1.5, 2\}$.

3.5 Comparing Two Pieces

This section details the methods we propose for obtaining a similarity measure. We present two methods and will later compare the results obtained for both. Our first method uses vector $\mathbf{F}$, and computes a distance between them as the measure of similarity. The second method uses both vectors **F** and **M**, and gets the distance between pairs of values of **F** and **M** as the measure of similarity.

The selected features measure the degree of tonal stability in a piece. A piece with an **F** vector containing peaks is more stable than a piece that has a uniformly distributed **F** vector. For an **F** with peaks, consider its corresponding **M** vector. If the values of **M** corresponding to the peaks of **F** are large, then the piece is more stable than if these values were small. We consider both the one- and two-vector methods to see if including the additional information in **M** gives better results.

3.5.1 Comparing Two Key Distributions

We use the distance between two probability mass functions (p.m.f.'s) as our first measure of similarity, which we will refer to as "Method 1". Consider two pieces, Piece A and Piece B, with key histograms, $\mathbf{F} = \{f_1, f_2, \ldots, f_{55}\}$ and $\mathbf{F}' = \{f'_1, f'_2, \ldots, f'_{55}\}$ respectively. We treat **F** and **F'** as p.m.f.'s, and measure the distance between them using the L^1 norm, shown in (1).

$$\sum_{i=1}^{55} \left| f_i - f_i' \right| \tag{1}$$

3.5.2 Comparing Pairs of Key and Mean-Time-In-Key Distributions

We use the Euclidean distance between two (**F**, **M**) pairs as our second measure of similarity, which we will refer to as "Method 2". Let $\mathbf{M} = \{m_1, m_2, \ldots, m_{55}\}$ and $\mathbf{M}' = \{m'_1, m'_2, \ldots, m'_{55}\}$ be the respective mean-time-in-key histograms for Piece A and Piece B. The measure of similarity is based on the L^2 norm, shown in (2).

$$\sum_{i=1}^{55} \sqrt{(f_i - f_i')^2 + (m_i - m_i')^2} \tag{2}$$

3.6 Example

At this point, we present a real example to illustrate the methods. We use three pieces for this example. Piece A is the theme section from Beethoven's *La Molinara*, Piece B is the third variation of the same piece, and Piece C is the second variation of Schumann's *Symphonische Etüden*. Since Piece B is a variation of Piece A, they should be more similar than Pieces A and C. Note that $m = 50$.

Consider the plots of **F** shown in Figure 1. Our assumption that Pieces A and B are similar and that Pieces A and C are different is supported by direct inspection of these plots. Using Method 1, we calculate a distance of 10 for Pieces A and B, and 98 for Pieces A and C.

The plots of **M** are shown in Figure 2. Using Method 2, we calculate a distance of 11.82 for Pieces A and B and 103.55 for Pieces A and C. These findings further support our initial assumptions. Method 1 performs better when comparing pieces A and B, while Method 2 performs better when comparing pieces B and C.

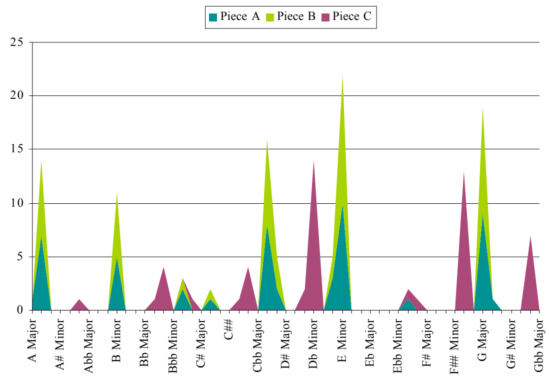

Figure 1. Plot of vector F for Pieces A, B, and C.

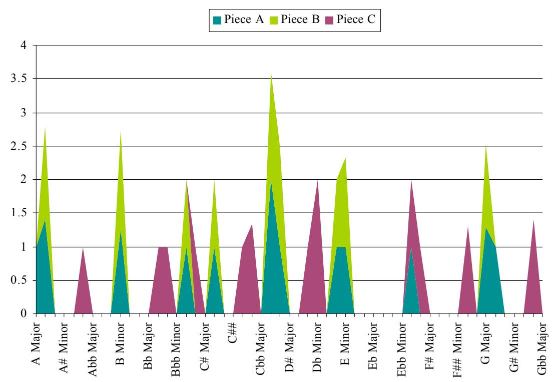

Figure 2. Plot of vector M for Pieces A, B, and C.

4. Results

We present and discuss our results in this section. We also discuss our motivation for choosing our particular data set. We state our expected results and then present our actual findings.

We briefly mentioned the selection of the segmentation variable m. The average length of a piece in our data set is 61.3 seconds. The value of m must be such that each k_i in **K** is small enough to provide insight yet large enough so as not to contain insignificant fluctuations. Initial sensitivity analysis experiments were conducted to see the behavior of our system as the value of m changed. We set m to a number in the range [50, 300]. At $m = 50$, each segment has, on average, a length of 1.26 seconds. Our system is robust within the range of values tested. We found that at $m = 50$ there was less noise than at $m = 300$. This led us to choose $m = 50$ for further experiments.

4.1 Data: Variation Sets

We have chosen to use Variation Sets as our data set since the similarity of pieces in the same set is objectively pre-defined by the composer, and thus less subject to dispute. We have amassed a collection of Variation Sets from [16] spanning ten composers and periods ranging from Baroque and Classical, to Romantic. Table 1 summarizes the statistics on our data set, consisting of 711 theme and variations over 71 Variation Sets.

Table 1. **Summary of the pieces in the data set.**

Composer	No. of Variation Sets	No. of Pieces	Avg. Piece Length (secs)
Bach	3	48	107.48
Beethoven	20	205	50.80
Brahms	8	128	57.41
Chopin	4	21	56.90
Handel	5	40	32.30
Haydn	12	93	53.32
Liszt	3	22	36.86
Mozart	10	99	60.95
Schubert	4	34	69.56
Schumann	2	21	87.38
TOTAL	71	711	61.30

4.2 Expected Results

We assume that pieces within the same Variation Set are similar one to another. Such an assumption was also made in [10]. Note that the converse may not necessarily be true. Even though we expect pieces from different Variation Sets to be less similar than pieces from the same, we cannot assume that they will not be similar. Even though we distinguish between composers in our analysis, it is unclear if the methods would be able to distinguish between composer's styles. Because the methods so strongly favor pieces with the same tonal patterns, a characteristic of variations, we expect pieces across different Variation Sets to be less similar one to another, even if they are by the same composer. We postulate that our methods are better at capturing similarities at the mid-level of theme and variations than at the composer level.

4.3 Comparing All Pieces by All Composers

We used both Methods 1 and 2 to compare all the pieces in our data set. We compared all 711 pieces to one another and obtained a total of 505,521 comparisons for each method of similarity assessment. Note that we then discarded repeated comparisons. Like the work in [10], we have chosen to include the comparisons of pieces to themselves in our analysis as they provide a good check of our system. When comparing N pieces, each segmented into m sections, the computational complexity for calculating all pairwise comparisons is $O(mN + N^2)$.

For each method, we extracted three groups from the total measurements. Group 1 contains all comparisons of pieces from the same Variation Set. Group 2 contains all comparisons of pieces from the same composer, but different Variation Sets. Group 3 contains all comparisons of pieces from different composers. Since the number of comparisons in each group differs greatly, we normalized the results so that the distributions sum to one.

4.3.1 Analysis of Results Using Method 1

In this Section, we analyze the distributions of the three Groups of data for Method 1. Based on our expectations stated in Section 4.2, we would expect that the distribution of Group 1 would differ from the distributions of Groups 2 and 3. Also, we would expect the distributions of Groups 2 and 3 to be similar. We construct empirical quantile-quantile plots in order to make these comparisons (see Figure 3), which consist of plotting the quantiles of one empirical distribution against the corresponding one in the other. If the distributions come from the same underlying distribution, then the plot will be close to the line x=y.

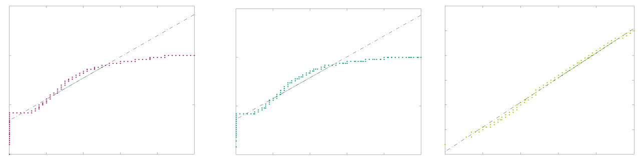

Figure 3. Quantile-quantile plots comparing Groups 1 & 2, 1 & 3, and 2 & 3, respectively, using Method 1.

It is clear from Figure 3 that Group 1 does not come from the same underlying distribution as Groups 2 and 3. Also, Groups 2 and 3 appear to be derived from the same underlying distribution. These observations support our initial assumptions and verify that Method 1 is successful at distinguishing between pieces from the same and different Variation Sets.

Table 2. Results of K-S test for Method 1.

Comparison Groups	K-S stat	p-value	Reject H_0?
1 and 2	0.588	0.00	Yes
1 and 3	0.612	0.00	Yes
2 and 3	0.051	$1.957x10^{-10}$	Yes

We also conduct a Kolmogorov-Smirnov (K-S) test to compare the distributions of the three groups. The null hypothesis, H_0, for this test is that two sets come from the same underlying continuous distribution. We present the results of this test in Table 2. This test verifies that the distribution of Group 1 is indeed significantly different from the distributions of Groups 2 and 3. Even though the distributions for Groups 2 and 3 appear to be similar, and the K-S statistic is correspondingly smaller, the test reveals that they come from different underlying distributions.

The distributions of Groups 1, 2 and 3 are shown in Figure 4. By inspection, we can see that the plot for Group 1 is significantly different from that for Groups 2 and 3, while Groups 2 and 3 appear much more similar.

Next, we perform some probabilistic analyses of classification errors should Method 1 be used for music categorization. Recall that Method 1 returns a single value for every comparison made between two pieces. If two pieces are exactly the same, this value is equal to zero. As the degree of difference between the pieces increases, so does this measure. In a rudimentary categorization scheme, we could select a cutoff point for determining if the two can be considered variations one of another. If the value is less than this cutoff point, we conclude that the pieces are similar. If it is greater than or equal to the cutoff point, we conclude that the pieces are dissimilar.

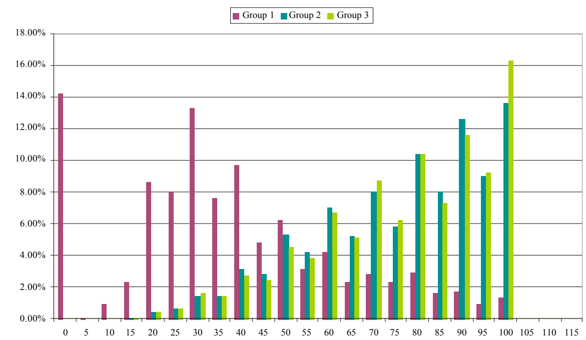

Figure 4. Distributions of similarity measures, obtained using Method 1, divided into Groups 1, 2 and 3.

Suppose the cutoff point is 55, the point at which the outlines of the three distributions cross. Let

A = "Two pieces are from the same Variation Set," and
B = "Their similarity value is less than 55."

Next we compute Type I (false positive) and Type II (false negative) probabilities for this strategy. The probability of a Type I error, P(B | A') = 18.41%. The probability of a Type II error, P(B' | A) = 20.56%.

Consider the question: if we pick a data point at random, and its value is less than 55, what is the probability that this data point belongs to Group 1? We can restate this question as P(A | B). Also consider the converse of this question: if we pick a data point at random, and its value is greater than or equal to 55, what is the probability that this data point does not belong to Group 1? This can be restated as P(A' | B'). We consider three possible scenarios: (1) when we consider all Variation Set comparisons for pieces by only one composer (i.e. A' consists of members of Group 2 only); (2) when we consider single-composer and different-composer Variation Set comparisons (A' = Group 3); and, (3) when we consider all comparisons (A' = Groups 2 and 3). The results are summarized in Table 3.

Table 3. Bayesian reasoning for Method 1.

Groups Considered as A'	P(A | B)	P(A' | B')
Group 2	34.90%	96.66%
Group 3	9.61%	99.39%
Groups 2 and 3	8.15%	99.48%

Since Groups 2 and 3 have far more data points than Group 1, a randomly selected data point is more likely to be from one of these groups than Group 1. Consider the probabilities of Groups 1, 2 and 3 in Table 3. When we restrict ourselves to only the comparisons of Variation Sets by the same composer (scenario 1), P(A | B) is 34.90% and P(A' | B') is 96.66%. When we consider same Variation Set and different composers comparisons (scenario 2), P(A | B) is 9.61% and P(A' | B') is 99.39%. When considering all comparisons (scenario 3), P(A | B) is 8.15% and P(A' | B') is 99.48%. Hence, in all cases, the probability of a false negative is far higher than that of a true positive. This is to be expected since the number of pieces in the same Variation Set is far exceeded by those that are not.

4.3.2 Analysis of Results Using Method 2

In this Section, we carry out the same type of analysis for Method 2. As before, we construct empirical quantile-quantile plots for comparisons of Groups 1 and 2, 1 and 3, and 2 and 3. The resulting plots are shown in Figure 5. By inspection, it is clear from Figure 5 that the data points of Group 1 are not derived from the same underlying distributions as Groups 2 and 3. And, the data points in Groups 2 and 3 seem to come from similar underlying distributions. These observations support our initial assumptions, and verify that Method 2 is also successful at distinguishing between pieces from the same Variation Set.

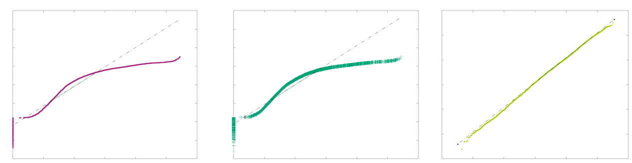

Figure 5. Quantile-quantile plots comparing Groups 1 & 2, 1 & 3, and 2 & 3, respectively, using Method 2.

We again conduct a K-S test to compare the distributions of the groups, this time using Method 2. Recall that the null hypothesis is that two sets come from the same underlying continuous distribution. We present the results of this test in Table 4.

Table 4. Results of K-S test for Method 2.

Comparison Groups	K-S stat	p-value	Reject H_0?
1 vs. 2	0.588	0.00	Yes
1 vs. 3	0.598	0.00	Yes
2 vs. 3	0.041	5.623×10^{-7}	Yes

This test verifies that the distribution of Group 1 is significantly different from those of Groups 2 and 3. The results of the comparisons of Groups 2 and 3 show that they come from different underlying distributions. However, note that the K-S statistic for this comparison is significantly smaller. We may conclude, as we did with Method 1, that even though Groups 2 and 3 do not come from the same underlying distributions, they are much more similar one with another than with Group 1.

Figure 6 is a visualization of the distributions of Groups 1, 2 and 3. Refer to this figure for further confirmation that the distribution of Group 1 is visibly different from the distributions of Group 2 and 3, while the distributions of Group 2 and 3 are more similar.

Again, we calculate the error rates and probabilities for Method 2. As before, the outlines of all three plots converge at the distance value 55, which we set as the cutoff point. In this case, the probability of a Type I error is P(B | A') = 15.72%, and the probability of a Type II error is P(B' | A) = 22.94%. Comparing these numbers with those for Method 1, we find that Method 1 has a lower Type II error (false negative) probability, while

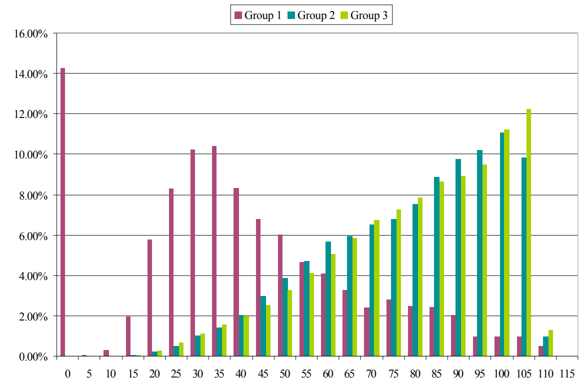

Figure 6. Distributions of similarity measures, obtained using Method 2, divided into Groups 1, 2 and 3.

Method 2 has a lower Type I error (false positive) probability.

We also compute P(A | B) and P(A' | B') for Method 2. We consider the same three scenarios as in Method 1. The results are presented in Table 5. When considering only the Variation Sets by the same composer (scenario 1), P(A | B) is 37.98% and P(A' | B') is 96.45%. When we consider same Variation Set and different composers comparisons (scenario 2), P(A | B) is 10.55% and P(A' | B') is 99.36%. When considering all possible pairs of pieces (scenario 3), P(A | B) is 9.00% and P(A' | B') is 99.45%.

Table 5. Bayesian reasoning for Method 2.

Groups Considered as A'	P(A | B)	P(A' | B')
Group 2	37.98%	96.45%
Group 3	10.55%	99.36%
Groups 2 and 3	9.00%	99.45%

5. Conclusion

We have shown that key histograms can be used to develop musically accurate summarization of pieces. Our selection of a data set has helped us establish ground truth in an area of research that often lacks it. We have provided two efficient methods for determining the level of similarity between two pieces. Both methods are rather successful in identifying similarity at the level of Variation Sets. Both methods are comparable in their level of success. When the cutoff statistic is set at 55, Method 1 has a success rate of 99.48% in labeling pieces as dissimilar, while the corresponding rate for Method 2 is 99.45%. Future work will consider comparisons of data sets normalized to the same key. This will allow us to better compare transposed variations and fluctuations in key. We will also consider variations on the same theme by several composers.

6. Acknowledgments

We thank Sheldon Ross for suggesting the Mean-Time-In-Key statistic and Kurt Palmer for suggesting some Bayesian analysis. This work was supported in part by the NSF under ITR Grant No. 0219912, and the shared facilities at, the Integrated Media Systems Center, an NSF ERC, under Cooperative Agreement No. EEC-9529152. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect those of NSF.

References

[1] E. Pampalk, Computational Models of Music Similarity and Their Application in Music Information Retrieval. Ph.D. thesis, Vienna University of Technology, 2006.

[2] N. Hu, R. Dannenberg, and A. Lewis, "A Probabilistic Model of Melodic Similarity," in *Proceedings of the International Computer Music Conference,* 2002.

[3] R. Typke, P. Giannopoulos, R. Veltkamp, F. Wiering, and R. van Oostrum, "Using Transportation Distances for Measuring Melodic Similarity," in *International Symposium on Music Information Retrieval*, 2003.

[4] J. Paulus, and A. Klapuri, "Measuring the Similarity of Rhythmic Patterns," in *Proceedings of the International Symposium on Music Information Retrieval*, 2002.

[5] L. Hofmann-Engl, "Rhythmic Similarity: A Theoretical and Empirical Approach," in *Proceedings of the 7th International Conference on Music Perception and Cognition*, 2002.

[6] E. Chew, A. Volk, and C.-Y. Lee, "Dance Music Classification Using Inner Metric Analysis – a computational approach and case study using 101 Latin American Dances and National Anthems," In Bruce Golden, S. Raghavan, Edward Wasil (eds.): The Next Wave in Computing, Optimization and Decision Technologies: Springer OR/CS Interfaces Series No.29, pp.355-370.

[7] S. Dixon, E. Pampalk, and G. Widmer, "Classification of Dance Music by Periodicity Patterns," in *Proceedings of the International Conference on Music Information Retrieval*, 2003.

[8] G. Tzanetakis, A. Ermolinskyi, and P. Cook, "Pitch Histograms in Audio and Symbolic Music Information Retrieval," *Journal of New Music Research*, 2003.

[9] A. Mardirossian and E. Chew "Key Distributions as Musical Fingerprints for Similarity Assessment," *1st IEEE International Workshop on Multimedia Information Processing and Retrieval*, 2005.

[10] J. Pickens, Harmonic Modeling for Polyphonic Music Retrieval. Ph.D. thesis, University of Massachusetts Amherst, 2004.

[11] 1st Annual Music Information Retrieval Evaluation eXchange," [Web site] 2006, url: *http://www.music-ir.org/mirex2005/index.php*

[12] E. Chew, Towards a Mathematical Model of Tonality. Ph.D. thesis, Massachusetts Institute of Technology, 2000.

[13] E. Chew, "Modeling Tonality: Applications to Music Cognition", in *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, 2001.

[14] E. Chew and Y.-C. Chen, "Real-Time Pitch Spelling Using the Spiral Array", *Computer Music Journal*, 2005.

[15] C.-H. Chuan, and E. Chew, "Fuzzy Reasoning in Pitch Class Determination for Polyphonic Audio Key Finding", *6th International Conference for Music Information Retrieval*, 2005.

[16] "Classical Music Archives," [Web site] 2006, Available: http://www.classicalarchives.com