Similarity Measures for Traditional Turkish Art Music

Ali C. Gedik *a.cenkgedik@musicstudies.org*
Faculty of Fine Arts/Department of Musicology
Dokuz Eylül University
İzmir, 35320, Turkey

Abstract

Pitch histograms are frequently used for a wide range of applications in music information retrieval (MIR) which mainly focus on western music. However there are significant differences between pitch spaces of traditional Turkish art music (TTAM) and western music which prevent to apply current methods. In this sense comparison of pitch histograms for TTAM corresponds to the research domain in pattern recognition: finding an appropriate similarity measure in relation with the metric axioms and characteristics of the data. Therefore we have evaluated various similarity measures frequently used in histogram comparison such as L1-norm, L2-norm, histogram intersection, correlation coefficient measures and earth mover's distance (EMD) for TTAM. Consequently we have discussed one of the problems of the domain, about measures regarding overlap or/and non-overlap between ordinal type histograms and presented an improved version of EMD for TTAM.

Keywords: Similarity Measure, Histogram Comparison, Earth Mover's Distance, Music Information Retrieval, Traditional Turkish Art Music.

1. INTRODUCTION

Measurement of similarity between histograms is one of the important fields in pattern recognition especially for the image retrieval applications. The applications in music information retrieval (MIR) are also simply based on the comparison of 2 pitch histograms, similar to the applications in other domains. Despite the frequent use of pitch histograms in MIR applications such as tonality finding, chord recognition and segmentation of musical pieces, the research on histogram based similarity measure is rather restricted in MIR literature.

MIR is mainly based on western music and the representation of pitch histogram for western music is rather simple in comparison with the image histograms. The distribution of pitches in a given musical piece is represented as pitch-class histogram which is a 12 dimensional vector corresponding to the equal tempered 12 pitch-classes (C, C#, D, D#, E, F etc.) in western music. Mostly, correlation coefficient or its variants are applied as similarity measure for the comparison of pitch histograms in western music [1] and since such measures give successful results, research on histogram comparison in MIR is limited in number.

However there are significant differences between pitch spaces of traditional Turkish art music (TTAM) and western music as discussed in detail [2]. The number of pitch-classes and the pitch interval values are still subject to discussions in TTAM. Therefore it is not possible to define pitch-classes and thus construct pitch-class histograms in TTAM as in western music. Pitch histograms in TTAM can only be constructed based on continuous representation of pitches as shown in Fig.1. Although it is possible to detect the peaks of the histograms as shown in Fig. 1, these peaks do not correspond to pitch-classes in TTAM as in western music. Therefore we prefer to call pitch histograms of TTAM as "pitch-frequency histograms" rather than pitch-class histograms. As a result, it is not straightforward to apply current histogram-based MIR methods to TTAM.

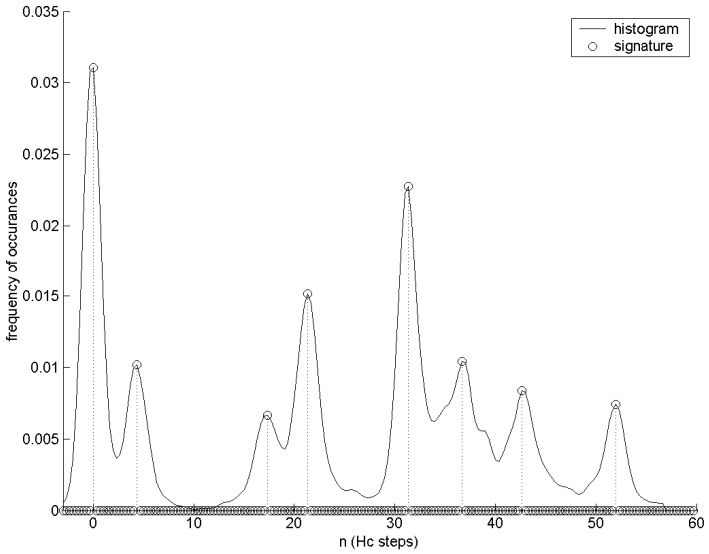


FIGURE 1: Pitch-frequency histogram of Tanburi Cemil Bey's *hicaz taksim*.

Consequently the problem of pitch histogram comparison for TTAM corresponds to the research domain in pattern recognition: finding an appropriate similarity measure in relation with the metric axioms and characteristics of the data. Therefore we have evaluated following similarity measures for TTAM used in histogram comparison studies and discuss the problems of the relevant literature: bin-by-bin measures L1-norm (Manhattan), L2-norm (Euclidean), intersection and correlation coefficient and, cross-bin and parameter based measure earth mover's distance (EMD).

Especially we introduced the discussion and evaluation of these similarity measures used in histogram comparison related with the pitch space characteristics of TTAM to the literature for the first time. Therefore the main contribution of the study is the application of histogram comparison methods to a new domain, TTAM recordings. The inadequacy of bin-by-bin measures for the comparison of ordinal type histograms due to their disadvantage has been demonstrated by regarding only the overlap between histograms [3]. They argue that the earth mover's distance (a cross-bin and parameter based measure) which regard the non-overlap between histograms as well as the overlap is more adequate than conventional measures for ordinal type histograms. Contrary to that study [3], we found bin-by-bin measures much more successful than EMD when applied to TTAM recordings represented as pitch-frequency histograms which are ordinal type histograms. Furthermore despite the high success rates we also discussed the adequacy of these measures for pitch-frequency histograms of TTAM.

The flexible structure of EMD by the signature representation of histograms represented as peaks as shown in Fig.1 also enables us to consider the pitch-space characteristics of TTAM for the improvement of EMD. Therefore we introduced an improved version of EMD which demonstrates considerable amount of improvement in comparison with original EMD and slightly better results than bin-by-bin measures for TTAM. In addition we also empirically showed for the first time that it is not possible to represent TTAM as pitch-class histograms as in western music.

Although the problem is simply based on comparison of 2 histograms, for evaluation we have formulated experiments in a context similar to tonality finding studies in western music where the tonality of a given musical piece is found either as major or minor [1]. Since TTAM is based on a modal system, the problem is expressed as finding the modality of a musical piece. Firstly the similarity measures are evaluated by using a relatively smaller database, 41 audio recordings

from 2 modalities in TTAM, for the simplicity of presentation. Modality templates are constructed and then modality of a musical piece is found by comparing it with the two modality templates in terms of pitch-frequency histograms. As a result the modality whose template gives the highest similarity measure is found as the modality of the musical piece. Secondly in order to investigate whether the conclusions about the similarity measures are valid for a larger database, the similarity measures are evaluated by using 172 audio recordings from 9 modalities in TTAM. As a result we obtained emprical results which support the conclusions obtained from the smaller database.

Presentation of the paper is as follows: next section presents the construction of pitch-frequency histograms and *makam* templates which constitute the basis for similarity measures. In Section 3 measures for histogram comparison are evaluated and discussed. Section 4 is dedicated to the presentation and evaluation of the improved EMD for TTAM. Finally Section 5 presents the discussions and conclusion.

2. PITCH-FREQUENCY HISTOGRAMS AND *MAKAM* TEMPLATES

Construction of pitch-frequency histograms for TTAM is presented by Bozkurt [4]. Therefore this section briefly reviews the construction of pitch-frequency histograms which constitute the basis for the application of similarity measures for TTAM and presents the construction of *makam* templates.

Given an audio recording, f0 data is estimated by the YIN algorithm [5] modified to improve its performance in the analysis of TTAM and extracted from each recording to construct pitch-frequency histograms [4]. An example of pitch-frequency histogram extracted from a recording was presented in Fig.1.b in the previous section. Histograms are represented as defined by Bozkurt [4]:

$$H_{f0}[n] = \sum_{k=1}^{K} m_k, \qquad m_k = \begin{cases} 1, & f_n \le f_0[k] \le f_{n+1} \\ 0, & otherwise \end{cases} \qquad (1)$$

, where (f_n, f_{n+1}) are boundary values for f_0 which also determines the resolution of the histogram.

Since pitch perception is logarithmic, it is more convenient to use logarithmic division of the pitch frequency axis for the calculations and visual representations. Conventionally Holder comma (Hc) is used in the studies on TTAM which is a unit obtained by the logarithmic division of an octave into 53 equal parts. In this study an octave is logarithmically divided into 53 x 3 equal parts to obtain an optimum resolution for the representation of histograms. It has been shown that such representation of pitch-frequency histograms are successfully used in recent studies on TTAM [2] [4] [6] [7].

Each *makam* template is constructed by simply averaging the pitch-frequency histograms from the relevant *makam*:

$$T_m = \sum_{i=1}^{N} H_{Nf_0 i} \quad (2)$$

, where H_{Nf0i} denotes normalized pitch-frequency histogram of the *i*th recording from *makam m*, *N* refers to number of recordings from *makam m* and T_m refers to template for the *makam m*.

In addition when a pitch-frequency histogram of a recording is compared with the histograms of

the two templates, the histogram of the relevant recording does not contribute to the construction of the relevant *makam* template. Therefore both in the construction of templates and in the evaluation of similarity measures, the leave-one-out (LOO) cross validation method is applied: each time a histogram of a recording is to be compared with templates then relevant template is constructed again by excluding the contribution of the histogram of the recording subject to comparison. Histograms of *makam* templates for *makam hicaz* and *makam saba* are presented in figure 2.

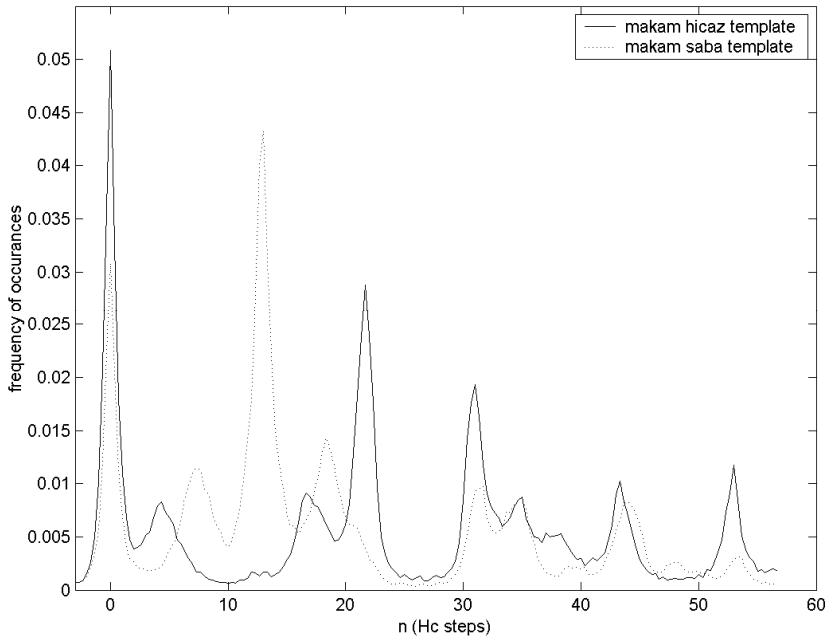


FIGURE 2: *Makam* templates for *makam hicaz* and *makam saba*.

3. SIMILARITY MEASURES FOR TTAM

As mentioned by Cha and Srihari [3], a histogram is defined as fixed-dimensional vector in the vector approach which is the most frequent approach applied in MIR studies on western music. Our study is also based on the vector approach, since we use a fixed dimension for the histograms. Types of histograms depend on the type of information represented which are called as nominal, ordinal or modulo type histograms [8]. Pitch histogram is an ordinal type histogram where the information is also present in the order of measurements.

Rubner et al. [8] also define 3 types of measures for histogram comparison: bin-by-bin dissimilarity, cross-bin dissimilarity and parameter-based dissimilarity measures. In bin-by-bin dissimilarity measure two histograms are compared with pairs of bins which have the same index. On the contrary to the bin-by-bin measure, two histograms are compared with pairs of bins which can have different indexes, as well, in cross-bin dissimilarity measure. On the other hand parameter-based dissimilarity measures use some information/parameter extracted from histograms, instead of using the histogram samples directly. Rubner et.al. [8] gives an example of this measure by mentioning a color image retrieval study of Das et.al. [9]where only peaks extracted from color histograms are used.

Since the pitch-frequency histogram model is presented in the previous section, our problem is to find the appropriate similarity measure for ordinal type histograms for TTAM. Measures used for histogram comparison can be applied to for the similarity between templates (see Fig.2) and a

recording (eg. see Fig.1). Therefore Manhattan (L1-norm), Euclidean (L2-norm), histogram intersection, correlation coefficient and earth mover's distance (EMD) are evaluated. Related with the type of measurements used in histogram comparison, we consider these measures in two sections: Manhattan, Euclidean, histogram intersection, correlation coefficient are evaluated in the section bin-by-bin measures and EMD is evaluated in the section cross-bin measure.

3.1 Bin-by-bin Measures

As described above, bin-by-bin measures compare two histograms by pairs of bins which have the same index. Bin-by-bin measures compare a sample histogram (Fig.1) with each *makam* template histograms (Fig. 2). The *makam* whose template gives the highest similarity measure is found as the *makam* of the sample. Therefore a successful measure should find the *hicaz* sample histogram (Fig. 1) more similar to the *hicaz* template histogram (Fig. 2).

The evaluated bin-by-bin measures L1-norm (Manhattan), L2-norm (Euclidean), intersection and correlation coefficient are presented below as d_1, d_2, d_3 and d_4, respectively:

$$d_1(p,q) = \sum_{i=1}^{N} |p_i - q_i| \quad (3)$$

$$d_2(p,q) = \sqrt{\sum_{i=1}^{N} (p_i - q_i)^2} \quad (4)$$

$$d_3(p,q) = \sum_{i=1}^{N} \min(p_i, q_i) \quad (5)$$

$$d_4(p,q) = \frac{\sum_{i=1}^{N} (p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum_{n=1}^{12} (p_i - \bar{p})^2 (q_i - \bar{q})^2}} \quad (6)$$

where, p_i and q_i denotes the bins of the two histograms having the same index, $\bar{p}$ and $\bar{q}$ refers to the mean of each histogram and N refers to the length of the histograms.

For the evaluation, 21 recordings from *makam saba* and 20 recordings from *makam hicaz,* totally 41 recordings are used. As a result all four measures give the same success rate in terms of F-measure as presented in Table 1 which are calculated by the set of parameters presented below:

$$recall = \frac{n_{TP}}{n_{TP} + n_{FN}},\ precision = \frac{n_{TP}}{n_{TP} + n_{FP}},\ F - measure = \frac{2 \cdot recall \cdot precision}{(recall + precision)} \quad (7)$$

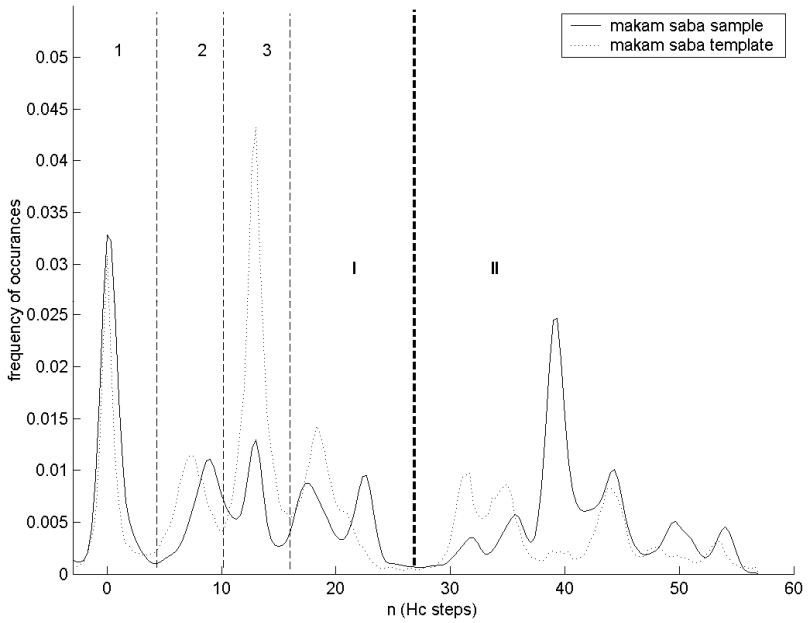
n_{TP}: # of true positives, n_{TN}: # of true negatives, n_{FP}: # of false positives, n_{FN}: # of false negatives

makam	n_{TP}	n_{TN}	n_{FP}	n_{FN}	***Recall***	***Precision***	***F-measure***
hicaz	20	19	2	0	100	90	95
saba	19	20	0	2	90	100	95
mean	19.5	19.5	1	1	95	95	95

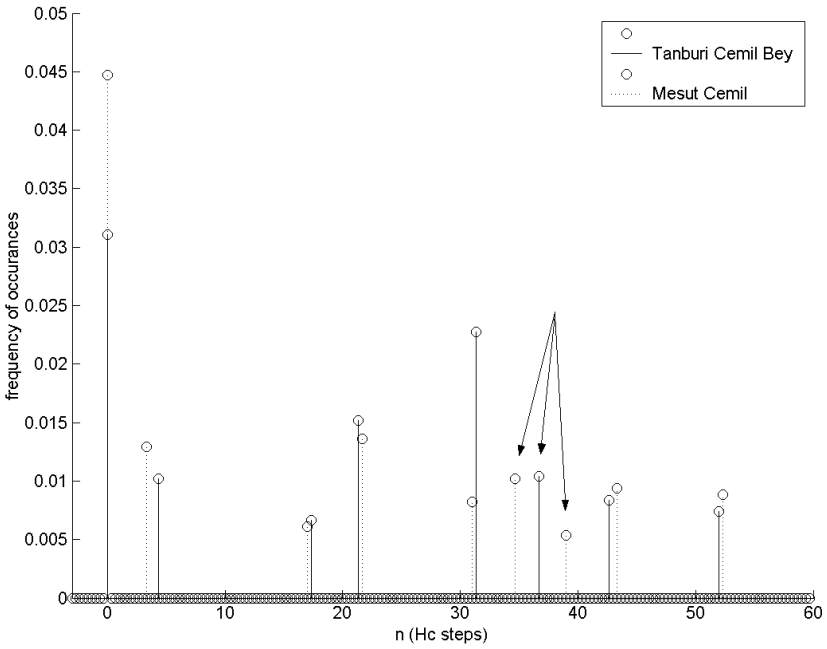
TABLE 1: Success rates of bin-by-bin measures for TTAM.

Furthermore all measures fail in the same 2 *saba* samples. Despite the high success rate, one of the samples which all measures fail, lead us to discuss these measures. In Fig. 3.a and 3.b the *saba* sample histogram is presented with *saba* and *hicaz* template histograms, respectively. To observe the reason of failure more clearly we have divided the histograms of the sample and

templates into two parts as part I and II as shown by bold dashed lines in Fig. 3. It can be observed from the part I of the figures that *saba* sample histogram is more similar to *saba* template histogram (Fig. 3.a) than *hicaz* template histogram (Fig. 3.b).



a. *saba* sample and *makam saba* template.



b. *saba* sample and *makam hicaz* template.
FIGURE 3: Histograms of a *saba* sample and *makam* templates.

Therefore bin-by-bin measures are applied to each part separately. When only part II of the histograms of sample and templates are compared, the *saba* sample is found to be more similar to *saba* template than *hicaz* template, truly. On the other hand when only part I of the histograms of sample and templates are compared, the *saba* sample histogram is found to be more similar to *hicaz* template histogram, wrongly. Thus bin-by-bin measures fail to identify the *saba* sample even for the part (part I) of the histograms which demonstrates the similarity of *saba* sample histogram with the *saba* template histogram more explicitly. Further division of part I as regions 1, 2 and 3, gives detailed evidences about this failure of the bin-by-bin measures. Measures are evaluated separately for each region in Fig.3.a and b:in all 3 regions although the *saba* sample histogram is more similar to *saba* template histogram than the *hicaz* template histogram, the bin-by-bin measures find the *saba* sample histogram more similar to the *hicaz* template histogram.

As a result, independent from the domain (music) or data type (pitch-frequency histograms) when part I of the histograms are considered, the problem corresponds to a discussion in histogram comparison literature [3]: bin-by-bin measures regarding only the overlap between histograms. However this problem can be observed only in one sample and bin-by-bin measures give considerably high success rate.

3.2 A Cross-bin and Parameter Based Measure: EMD

Cha and Srihari [3]show that for ordinal type of histograms, EMD proposed by Rubner et al. [8] give better performance than conventional bin-by-bin measures. EMD is found advantageous against other measures in terms of its 2 features, cross-bin and parameter based qualities: first EMD considers the similarity of non-overlap between histograms, as well as the overlap resulting from the cross-bin measure quality and second the parameter based quality gives the opportunity to represent histograms in a more efficient way by the use of bins only containing significant information. Due to these advantages, we expect better results from EMD than bin-by-bin measures for TTAM. However due to its high computational cost, there is also a number of algorithms proposed to outperform EMD [8] [10] [11] [12] [13]. Finally, since our database is relatively small, we prefer to apply the original EMD without considering its computational cost.

EMD is simply defined as a solution to the following transportation problem: several suppliers distribute their goods to several customers. There is an amount of commodity of several suppliers and an amount of capacity for several customers. So the problem is to find the optimal commodity flow from suppliers to customers. For the case of histogram comparison, the problem can be defined as finding the amount of optimum work necessary to resemble one histogram to another.

EMD proposed by Rubner et.al. [8] use signatures extracted from histograms, instead of histograms themselves. Signature is defined as a set of clusters which can be obtained by vector quantization of a given histogram. The main idea behind using signatures instead of histograms is to be able to work only with the bins of the histogram which contain significant information and get rid of the rest of the insignificant bins which reduces the computational cost. Rubner et.al. [8] also show that signature representation gives better results than the histogram itself in an image retrieval problem. Although Rubner et.al. [8] do not restrict the definition of cluster, signature is defined as follows: $P = \{(p_1, w_{p1})_{,\ldots,} (p_m, w_{pm})\}$, where each pair as the element of the signature set P refers to a cluster. So each cluster is represented by its mean p_m and by its proportion w_{pm}. A second signature is also defined to formulate the similarity measure between two signatures: $Q = \{(q_1, w_{q1}),\ldots, (q_n, w_{qn})\}$. Finally EMD is defined by the equation below (Rubner et al., 2000):

$$WORK(P,Q,F) = \sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij} \,, \qquad (8)$$

, with the following 4 constraints:

$$1)\quad f_{ij} \geq 0, \qquad 1 \leq i \leq m,\ 1 \leq j \leq n$$

$$2)\quad \sum_{j=1}^{n} f_{ij} \leq w_{p_i}, \qquad 1 \leq i \leq m$$

$$3)\quad \sum_{i=1}^{m} f_{ij} \leq w_{q_j}, \qquad 1 \leq j \leq n$$

$$4)\quad \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min\left(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j}\right)$$

where d_{ij} refers to ground distance between clusters p_i and q_j from signatures P and Q, and f_{ij} refers to amount of weight which is transferred from the bin with index i to the bin with index j which belongs to signatures P and Q respectively. Finally w_{pi} and w_{qj} are the weights of the bins with index i and j. As a result, the kind of histogram comparsion measure applied in EMD can be defined as a mixture of cross-bin and parameter-based dissimilarity measure. Cross-bin quality comes from comparing different bins of the histograms and parameter-based quality comes from the signature representation of histograms as defined in EMD.

The most significant information about pitch-frequency histograms exists in the peaks of the histograms as shown in Fig. 1 and these peaks can also be thought as the pitches analogous to pitches in western music. Therefore each pitch-frequency histogram is represented as a signature, collection of peaks (clusters). The bins and weights of the peaks can directly represent the distribution of pitches performed in a recording. To obtain a signature, each histogram is smoothed by finite impulse response (FIR) low pass filters to eliminate the noise like peaks and then a peak detection algorithm is applied to each histogram. *Makam* templates are also represented as signatures by applying similar procedures.

Finally, signature of each recording is compared with the signatures of two *makam* templates using the EMD measure. Again, the *makam* whose template gives the highest similarity measure is found as the *makam* of the recording. For the evaluation, the same data set presented in previous section is used. Table 2 presents the evaluation of EMD measure applied to pitch-frequency histograms of TTAM with an L_1-norm (Manhattan) distance used as a ground distance:

$$d_{ij} = \left| p_i - q_j \right| \quad (9)$$

Success rates of EMD are calculated according to the parameters presented in Eq. 7.

makam	n_{TP}	n_{TN}	n_{FP}	n_{FN}	***Recall***	***Precision***	***F-measure***
hicaz	16	12	9	4	80	64	71
saba	12	16	4	9	57	75	65
mean	24	24	6.5	6.5	68.5	69.5	69

TABLE 2: EMD applied to TTAM recordings.

As a result, on the contrary to our expectation from EMD, the success rate is found as 69 %, considerably lower than the bin-by-bin measures. Although the quality of cross-bin dissimilarity measure and representation of histograms in a more compact way are the advantages of EMD, partly these features also become disadvantageous for TTAM. While EMD succeeds in comparing similar pitches from two histograms which reside in different but close bins, it is likely that EMD can also compare two irrelevant pitches which reside in far bins if such comparison is a part of the optimum solution. As an example, assume that two pitches with different weights and in far bins from two histograms are matched; subsequently EMD would transfer some weight from one pitch to another.

Such kind of matching and transfer would be an inadequate operation when the data type is considered. An example from western music can be more explicit: it is irrelevant to compare the proportion of pitch C from one histogram with the proportion of pitch F from the other histogram.

Finally, since the pitch-frequency histograms are represented as signatures, we obtain representation of pitches in TTAM similar to western music. Therefore it is possible to apply correlation coefficient to signature representations of pitch-frequency histograms, to see whether this measure would give successful results for TTAM as applied to western music. Consequently application of the correlation coefficient to the same data set gives a success rate of 21 % worse than chance (50% for our data set) although it is one of the most successful measures for western music. This result empirically proves our argument about the different pitch space qualities of TTAM and western music.

4. EMD FOR TRADITIONAL TURKISH ART MUSIC

EMD bears a flexible structure arising from the signature representation which enables to make some improvements on it based on pitch space characteristics of TTAM. Therefore we try to preserve the cross-bin measure quality of EMD, while restricting this quality by an adequate distance obtained from the pitch space characteristics of TTAM.

Rubner et.al. [8] show that EMD is a true metric when the signatures have equal weights and the ground distance is a metric. It is well-known that a true metric should satisfy 4 axioms of metric space [14]. However Tversky [15] discusses the perceptual validity of these conditions for a similarity measure by comparing the human similarity judgments. Based on the psychological experiments on similarity judgments, Tversky questions each of the condition. In addition Strelkov [16] points the importance of data type for the applicability of similarity measures and proposes a new similarity measure based on expert decision in the area of data type. Therefore we try to take into account the characteristics of data, pitch space characteristics of TTAM and discuss metric axioms in the improvement of EMD.

4.1 Pitch Space Characteristics of TTAM

The pitch space characteristic of TTAM can be observed from the *hicaz taksim* performances of two outstanding musicians, Tanburi Cemil Bey and his son Mesut Cemil as shown in Fig.4. Pitch-frequency histograms of the two recordings are represented as signatures where the peaks represent pitches. It can be seen from the figure that pitches except the ones pointed by arrows constitute pitch pairs which are around 1.5 commas apart. This fact clearly demonstrates the flexibility of pitches, on the contrary to the exact pitch intervals in western music.

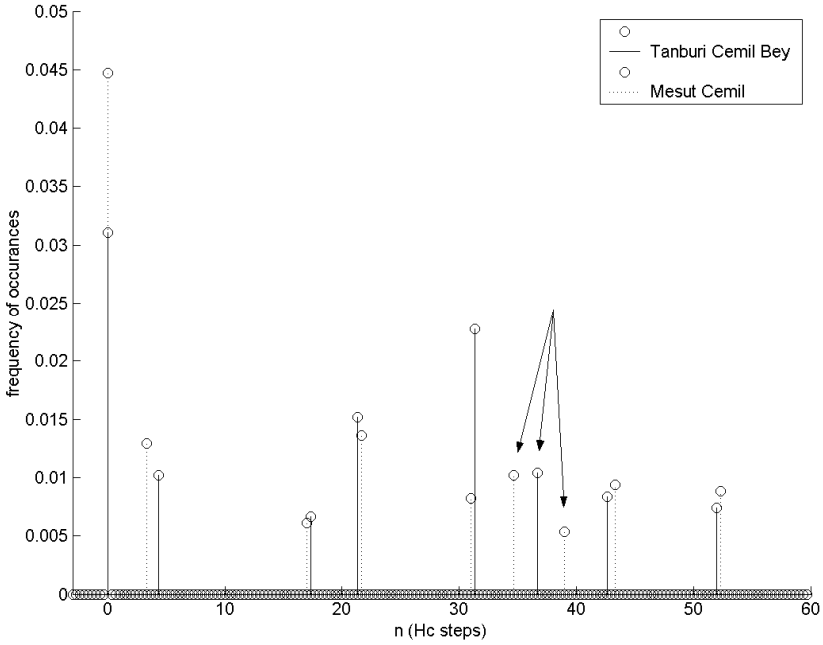


FIGURE 4: Two *hicaz taksim* performances by Tanburi Cemil Bey and Mesut Cemil.

Another distinction between two performances occurs in the number of pitches performed. While Tanburi Cemil Bey performs 8 pitches, Mesut Cemil performs 9 pitches for the same *makam* within the same octave. Therefore it is not straightforward to estimate number of pitches directly from recordings. For a detailed discussion about pitch characteristics of TTAM, tuning theories and automatic analysis methods developed for tuning analysis, the reader is referred to Bozkurt [4] and Bozkurt, et. al. [17].

4.2 Improvement of EMD for TTAM

Regarding the pitch space characteristics of TTAM and the critiques made on EMD, we can draw the framework of improved EMD. First we propose an approach based on pitches of each sample and template. Thus comparison of a recording with a *makam* template will be a matter of comparing the pitches (peaks) of a sample with the pitches (peaks) of a template. However on the contrary to the original EMD, only the peaks which are 1.5 commas apart are subjected to comparison for improvement. In other words while the cross-bin approach of EMD is preserved, cross-bin comparison is restricted by comparing bins which are only 1.5 commas apart. This restriction transforms original EMD into an adequate measure where the pitches from two histograms are matched if they reside only in relevant bins.

Consequently we consider 2 terms for the new measure: dissimilarity of the peaks matched and dissimilarity of the peaks unmatched from two histograms. First term corresponds to the peaks of the sample and the template which are matched. Second term consist of two elements: pitches of the sample which does not match with pitches of the template, and pitches of a template which do not match with any pitches of a sample.

These arguments are expressed by the measure presented below:

$$EMD_{imp} = \sum_{i=1}^{m}\sum_{j=1}^{n} F_{ij} \qquad \begin{cases} d_{ij} \le 1.5, \;\; F_{ij} = f_{ij} \\ otherwise, \;\; F_{ij} = f_i + f_j \end{cases} \tag{10}$$

, where d_{ij} denotes the distance between two bins from two histograms. It should be noted that for the case $d_{ij} \leq 1.5$, our distance measure is defined with the constraints of the original EMD. The first term f_{ij} denotes the transfer (difference) of weights of the peaks matched and the second term $f_i + f_j$, denotes the summation of weights of the peaks unmatched. L_1-norm ground distance, d_{ij} is used again as presented in Eq. 3 but only for the decision of applying either the first term or the second term. Since this measure is an improved version of EMD for TTAM, we refer to it as EMD_{imp}.

Finally it should be noted that EMD_{imp} is not a true metric. Although it is trivial to prove that EMD_{imp} satisfies the 3 axioms of the true metric, clearly it does not satisfy the second axiom: $d(p, q) = 0$, if and only if $p = q$. This axiom does not satisfy due to fact that the ground distance includes a condition of 1.5 commas. Therefore two signatures which are different ($p \neq q$) but each pitch pairs are 1.5 commas apart would give zero similarity value as if they are equal. Consequently EMD_{imp} is a pseudometric [18].

Success rates of EMD_{imp} are calculated according to the parameters presented in Eq. 7 and shown in Table 3.

makam	n_{TP}	n_{TN}	n_{FP}	n_{FN}	***Recall***	***Precision***	***F-measure***
hicaz	20	20	1	0	100	95	97.5
saba	20	20	0	1	95	100	97.5
mean	20	20	0.5	0.5	97.5	97.5	97.5

TABLE 3: Success rate of EMD_{imp} for TTAM.

It is clear that EMD_{imp} demonstrates a significant amount of improvement in comparison with the original EMD. While the success rate of original EMD is found as 69 %, the succes rate of EMD_{imp} is foundas 97.5 % in terms of F-measure. Furthermore EMD_{imp} succeed to identify the *makam* of one sample which other bin-bybin measures failed as discussed in Subsection 3.1.

5. COMPARATIVE EVALUATION

In order to investigate whether the conclusions about the similarity measures are valid for a larger database, the similarity measures are evaluated by using 172 audio recordings from 9 modalities in TTAM. The same evaluation context is applied to all measures as used throughout the paper and the overall success rates are presented in Table 4 in terms of F-measure.

measure	***F-measure***
L1-norm	68
L2-norm	63
intersection	68
correlation	58.5
EMD	17.5
EMD_{imp}	72

TABLE 4: Success rate of all similarity measures when applied to 172 audio recordings from 9 modalities in TTAM.

As a result we obtained emprical results for a significantly larger database which supports the conclusions obtained for the smaller database. While the success rate of EMD is found considerably less than bin-by-bin measures, the succes rate of EMD_{imp} is found significantly more than the original EMD and slightly better than bin-by-bin measures.

In order to present a comparative evaluation we have to look for studies using similarity measures for TTAM, since similar studies based on musics other than TTAM would not provide apropraite comparison. The differences between western musics and TTAM, and the similarity measures used for western musics was discussed in the previous sections. Thus it is not possible to compare our evaluations with the ones found for western musics. However, there are also not much study on TTAM for the comparison of our results.

In this sense we compare the evaluations of the study [2] held by Gedik and Bozkurt. Although this study uses various similarity measures for TTAM, these measures are evaluated according to their success on automatic tonic finding, not makam recognition. This study supplies an appropriate comparative evaluation of our current study in the sense that the bin-by-bin similarity measures for the comparison of pitch-frequency histograms are used.

150 synthetic and 118 real audio files are evaluated by cross-correlation, Euclidean (L2-norm), city block (L1-norm), intersection and Bhattacharyya measures in that study. All those measures except Bhattacharyya are the same measures used in our study. The tests of synthetic audio files consist of 7 makams which are the ones used in our study. The results of automatic tonic detection for synthetic audio files supports our results. City block (L1-norm) and intersection measures are found as the most successful measures. While city block (L1-norm) and intersection measures give no error rate, cross-correlation and Euclidean measures fail on 4 samples in automatic tonic detection.

The tests on 118 real audio files consist of the same 9 makams with our study. Similarly city block (L1-norm) and intersection measures are found as the most successful measures. They failed only one sample in automatic tonic detection, while the other failed more than one sample.

As can bee seen from the Table 4, except the EMD_{imp} measure, the most successful measures we found in our study is also the same as the study we compared: City block (L1-norm) and intersection measures. Of course, since EMD_{imp} is a measure we proposed for the first time in this study, it is not possible to compare its success with other studies.

6. DISCUSSION AND CONCLUSION

In this study we presented the problem of pitch histogram comparison for TTAM. Therefore we have evaluated following similarity measures for TTAM used in histogram comparison studies and discuss the problems of the relevant literature: bin-by-bin measures L1-norm (Manhattan), L2-norm (Euclidean), intersection and correlation coefficient and, cross-bin and parameter based measure earth mover's distance (EMD).

Although pitch histograms and histogram comparison are frequently used in MIR studies on western music, it was not possible to apply the current methods due to significant differences between pitch spaces of TTAM and western music. Therefore we have presented appropriate methods for the representation of pitch histograms and histogram comparison for TTAM.

Contrary to Cha and Srihari [3], we found bin-by-bin measures much more successful than EMD when applied for TTAM recordings. However we also discussed the adequacy of bin-by-bin measures over one sample which supports the arguments of Cha and Srihari [3], partly. Since the signature representation of histograms in EMD enables us to represent pitch histograms of TTAM similar to western music, we also empirically showed that it is not possible to represent TTAM as pitch-class histograms as in western music: correlation coefficient, the most frequently used similarity measure in western music, gives success rate worse than chance when applied to signature representation of TTAM.

We have also introduced an improved version of EMD, EMD_{imp} which demonstrates considerable amount of improvement in comparison with original EMD and slightly better results than bin-by-bin measures for TTAM. Finally L1-norm, L2-norm, intersection and EMD_{imp} measures are found successful when applied for the comparison of pitch histograms of TTAM.

Besides the success rate, the most important advantage of our proposed measure EMD_{imp} in comparison to bin-by-bin measures is the ease of its representation of pitch-frequency histograms. While 60 dimensional vector is necessary for the comparison of each pitch-frequency histograms by bin-by-bin measures, 10-15 dimensional vectors are enough to represent each pitch-frequency histograms. This amount of dimension reduction is no doubt an important improvement both for the representation of music files in databases and the computational cost of comparison.

However,the most important drawback of our study is the volume of the database. While the database of similar studies on western musics can reach to thousands of audio files, the number of audio files used in studies on non-western musics such as the TTAM we studied are much smaller. This is not surprising since the number of studies on western musics are also much higher than the number of studies on non-western musics. Therefore we hope to evalaute our study on a much larger database in the future.

7. REFERENCES

1. D. Temperley. The Cognition of Basic Musical Structures. MIT Press, Cambridge, Massachusetts, 2001

2. A. C. Gedik and B. Bozkurt. Pitch frequency histogram based music information retrieval for Turkish music, Signal Processing, Vol. 90, No. 4, pp. 1049-1063, 2010.

3. S. H. S. Cha and N. Srihari. On measuring the distance between histograms, Pattern Recognition, Vo. 35, pp. 1355–1370, 2002.

4. B. Bozkurt. An Automatic Pitch Analysis Method for Turkish Maqam Music, Journal of New Music Research, Vol. 37, No. 1, pp. 1–13, 2008.

5. A. de Cheveigne and H. Kawahara. YIN, a fundamental frequency estimator for speech and music, Journal of the Acoustical Society of America, Vol. 111, No. 4, pp. 1917-1930, 2002.

6. C. Akkoç. Non-deterministic scales used in traditional Turkish music, Journal of New Music Research, Vol. 31, No. 4. pp. 285-293. 2002.

7. A. C. Gedik and B. Bozkurt., Evaluation of the Makam Scale Theory of Arel for Music Information Retrieval on Traditional Turkish Art Music, Journal of New Music Research, Vol. 38, No. 2, pp. 103-116, 2009.

8. Y.Rubner, C. Tomasi, and L. J. Guibas., The Earth Mover's Distance as a Metric for Image Retrieval, International Journal of Computer Vision, Vol. 40, No. 2, pp. 99-121, 2009.

9. M. Das, E.M. Riseman, and B.A. Draper. FOCUS: Searching for multi-colored objects in a diverse image database, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1997, pp. 756–761.

10. J. Morovic, J. Shaw, P.L. Sun. A fast, non-iterative and exact histogram matching algorithm, Pattern Recognition Lett., Vol. 23, pp. 127–135, 2002.

11. J.K., Kamarainen, V. Kyrki, J. Llonen, H. Kälviäinen. Improving similarity measures of histograms using smoothing projections, Pattern Recognition Lett., Vol. 24, pp. 2009–2019, 2003.

12. F. Serratosa and A. Sanfeliu. A fast distance between histograms, Lecture Notes on

Computer Science, Vol. 3773, pp. 1027 – 1035, 2005.

13. Ling, H. and Okada, K. An Efficient Earth Mover's Distance Algorithm for Robust Histogram Comparison, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 5, pp. 840-853, 2007.

14. M. Rosenlicht. Introduction to analysis, Dover Pub., New York, 1968.

15. A. Tversky. Features of similarity. Psychological Review, Vol. 84, No. 4, pp. 327–352, 1977.

16. V.V. Strelkov. A new similarity measure for histogram comparison and its application in time series analysis, Pattern Recognition Letters, Vol. 29, pp. 1768–1774, 2008.

17. B. Bozkurt, O.Yarman, M. K. Karaosmanoğlu and C. Akkoç. Weighing Diverse Theoretical Models On Turkish Maqam Music Against Pitch Measurements, Journal of New Music Research, Vol. 38, No. 1, pp. 45-70, 2009.

18. K. Ito, (ed.). "Metric Space", Encyclopedic Dictionary of Mathematics, Vol.2, The Mathematical Society of Japan, MIT Press, 2nd edition, 1993.