INDEXING MUSIC COLLECTIONS THROUGH GRAPH SPECTRA

Alberto Pinto*, Reinier H. van Leuken, M. Fatih Demirci, Frans Wiering, Remco C. Veltkamp
Department of Information and Computing Sciences - Universiteit Utrecht (The Netherlands) and
* Dipartimento di Informatica e Comunicazione - Università degli Studi di Milano (Italy)
pinto@dico.unimi.it, {reinier, mdemirci, Frans.Wiering, Remco.Veltkamp}@cs.uu.nl

ABSTRACT

Content based music retrieval opens up large collections, both for the general public and music scholars. It basically enables the user to find (groups of) similar melodies, thus facilitating musicological research of many kinds. We present a graph spectral approach, new to the music retrieval field, in which melodies are represented as graphs, based on the intervals between the notes they are composed of. These graphs are then indexed into a database using their laplacian spectra as a feature vector. This laplacian spectrum is known to be very informative about the graph, and is therefore a good representative of the original melody. Consequently, range searching around the query spectrum returns similar melodies.

We present an experimental evaluation of this approach, together with a comparison with two known retrieval techniques. On our test corpus, a subset of a well documented and annotated collection of Dutch folk songs, this evaluation demonstrates the effectiveness of the overall approach.

1 INTRODUCTION

Singing songs has always been an important way of passing on stories and expressing emotions, religious beliefs or social values. Most of these folk songs were transferred orally, often significantly changing over time and location. This resulted in the existence of many versions of the same songs, often displaying considerable variations. In *Onder de groene linde*, a collection of Dutch folk songs has been assembled by Ate Doornbosch, a Dutch radio broadcaster and researcher [1]. By recording many singers in the countryside during a period of over three decades, he captured this cultural heritage counting more than 7300 songs on tape. A large part of these melodies and songs has now been transcribed to music notation. The collection is becoming available to the general public and research community [7]. Content-based music retrieval opens up this great resource in such a way that both audiences can access it better. Through music information retrieval, songs belonging to the same class of songs can be grouped, or songs with only slight variations can be found. It can help identify the composer of a song, or assist in any other scholarly musicological task.

We have three main contributions in this paper. Firstly, we introduce a new approach to music retrieval in which the music is represented as graphs, and the matching is based on specific features of these graphs. Our graph representation encodes the interval structure of a melody; it is a global time-independent signature of the melody, displaying the network of connections that exists between the pitch classes.

Secondly, we introduce our indexing approach, which is new to music retrieval. To compute similarity between melodies, an algebraic structure is associated to each graph: an $n \times n$ matrix, with n equal to the number of vertices in the graph.

Thirdly, we evaluate our method on a test corpus of Dutch folk songs. In this evaluation, we compare our method to two other methods: one approach specifically targeted towards folk song collections, and one approach using a time-independent structural approach as well. In this comparison, our method outperforms the other methods in terms of three well-known performance measures, namely, nearest neighbor, first and second tier.

2 RELATED WORK

Melodic similarity has been investigated by many authors from very different points of view, using different kind of song collections as dataset.

One of the most complete and recent studies has been performed by Müllensiefen and Frieler, who explored the concept melodic similarity within a collection of folk songs [10]. Using a collection of 577 Luxembourg folk songs, they empirically established an optimal similarity measure (the Opti3) that combines several known methods into one unifying expression. Out of 50 implemented musical similarity measures, taking into account all sorts of musical features, a weighted combination of methods was chosen to create one measure that best reflected the results of an extensive human listening experiment. Since their approach is specifically targeted towards a collection of folk songs, we compare it to our method as well.

Another example of a representation/matching/indexing paradigm is the weighted point set on which the Earth Mover's Distance or Proportional Transportation Distance can be applied. This kind of approach has been applied to test music similarity as well [12]. The notes of a melody are encoded as weighted points in a two-dimensional space where pitch and onset time are the axes; the duration of a note determines its weight. Similarity between two melodies can now be computed by measuring the effort it takes to transform one weighted pointset into the other.

3 REPRESENTATION

Our goal is to provide a sufficiently abstract representation of a melodic line that actually makes sense from a musical point of view. With this aim, we start looking just at melodies, not considering the rhythm. Melodies are generally studied from a pitch sequence/contour point of view. Our approach is different: we take as a starting point the interval structure, by which we mean the network of connections between pitches. We remark that melodies use only a subset of all possible connections, and with different frequencies. To model such relationships we use graphs, which have various and significant applications throughout mathematics, computer science, and physics. As such, the graph is a projection of the time-dependent concept of melody to a time-independent concept of intervallic structure. The next level of abstraction is to leave out pitch class information so that only the "interval connectivity" of the melody remains, and this means that certain operations such as inversion, transposition, retrogradation, other kind of permutations in the pitch class set and (some) shifting of fragments does not affect the graph. In this perspective what we are modelling is a global, time-independent signature of the melody [11], [8]. Melodies that display a similar interval behaviour have similar graphs, for example melodies in which there are one or two central notes (with many connections) and a number of peripheral notes (few connections).

Let M be a melodic sequence of length $m = |M|$ and consider the sequence of pitches $\{p_j\}_{j \in I}$, $\{I = 1, ..., m\}$. Then let $V = \mathbb{Z}_{12}$ be the (metric) space of pitches, or pitch classes, in the 12-tone system. We define the graph G with vertex set $V_G = V$ and edge set whose elements are the edges a_j such that

$$a_j : \begin{cases} p_j \to p_{j+1} & \text{for every couple } (p_j, p_{j+1}) \subseteq M \\ p_m \to p_1 & \text{for the couple } (p_m, p_1) \end{cases}$$

where $j = 1, \ldots, m-1$ (see also [2] and [5]).

The arrow $a_m : p_m \to p_1$ does not represent an actual interval in the melody but it has been added for symmetry reasons and in order to take into account the relationship between the last and the first note as well, which otherwise would not have been reflected in the model.

4 INDEXING

The graph representation described up to now is a geometric one. In order to allow computations with this representation, we need to associate an algebraic structure to it. The most common algebraic structure to represent a graph is the adjacency matrix.

The adjacency matrix $A(G)$ of a graph G is a square matrix of size equal to the order of the graph and where the entry (i, j) represents the number of oriented edges from vertex i to vertex j. This adjacency matrix therefore contains all the information to reconstruct the connectivity of the graph. A matrix closely related to the adjacency matrix is the laplacian matrix $L(G)$, computed as $L(G) = D(G) - A(G)$, where $D(G)$ is the degree matrix of G. The degree matrix is also a square matrix of size equal to the order of the graph, but all values are zero except for those on the main diagonal. Here, the entry (i, i) represents the number of outgoing edges of vertex i.

Given the laplacian matrix of a melody graph, the question remains how to compute the similarity to another melody. For this purpose, we first compute the eigenvalues of the laplacian matrix and sort them by magnitude. [1] Hereby, we obtain the laplacian spectrum of the graph, that is known to reflect a number of important properties of the graph. These properties include the diameter (related to the second smallest eigenvalue), mean distance, minimum degree and algebraic connectivity. Furthermore, the spectrum is invariant under permutations of the matrix (i.e. swapping columns or rows). Together with the absence of pitch information stored in the matrix, this makes the representation invariant under transpositions and note permutation. This is an important property, because as pointed out before, our concept of similarity is also independent from note permutation and transposition.

Our main motivation for encoding the topology of a graph using the laplacian matrix comes from the fact that laplacian matrices are more natural, more important, and more informative than other matrices about the input graphs [9]. Previously, Godsil and McKay [4] and more recently Haemers and Spence [6] have also shown that the laplacian matrix has more representational power than the adjacency matrix, in terms of resulting in fewer cospectral graphs. Recall that two graphs are called cospectral (or, isospectral) if they have the same eigenvalues.

Given a query graph and a large database, the objective of an indexing algorithm is to efficiently retrieve a small set of candidate matches, that share topological similarity with the query. As pointed out, we encode the topology of a graph through its laplacian spectrum, which is used as a signature for the database object. This spectrum can be seen as a point in a high dimensional space. To compute similarity between two graphs, we compute the Euclidean distance between their signatures, which is inversely proportional to the structural similarity of the graphs. Therefore, for a given query, retrieving its similar graphs can be reduced to a nearest neighbor search among a set of points. A set of candidate matches can now be found without having to inspect the entire database. For more details on this indexing strategy, the reader is referred to [3].

[1] Since the graphs are directed, the laplacian matrix is not necessarily symmetric. Consequently, some of the eigenvalues may be complex numbers and there exist multiple strategies for sorting these. As in [13], we sort these eigenvalues by modulus.

CRITERIA	NN	1^{st} tier	2^{nd} tier
LAPLACIAN	66%	44%	63%
ADJACENCY	58%	28%	48%
OPTI3	40%	39%	56%
EMD	64%	33%	50%
PTD	64%	30%	46%

Table 1. Nearest neighbour (NN), first tier and second tier results on the *Onder de groene linde* collection, computed using Laplacian spectra (L) Adjacency spectra (A) of the graphs. The results are compared to the methods Opti3, EMD and PTD.

5 EXPERIMENTS

In "Onder de groene linde", a large number of Dutch folk songs is preserved. This collection consists of more than 7300 songs recorded on tape. These songs are documented and annotated in great detail, and illustrated by sheet music examples. We experimented on a subset of this resource, that consists of 141 songs, of which we used the first phrase. These songs have been classified in 18 classes or *melody groups*, that relate to the concept of *melody norms*.

At the Meertens Institute (a research institute for Dutch language and culture in Amsterdam) the concept of *melody norm* [2] is used to group historically or "genetically" related, orally transmitted melodies. Because the contents of folk song collections such as OGL are highly fragmented, it is impossible to trace back the history of melodies and to find all variants that are derived from a common 'ancestor' melody. What can be done, is to find related groups of melodies within the collection, based on both melodic similarity and available meta data, and link them to melody norms. A search engine would speed up this process of relating melodies considerably. As a ground truth in our experiments, we used a classification of the melodies into *melody groups*, that serve as candidates for the melody norms to be assigned in a later stage.

For all the melodies in our test corpus, a graph has been constructed as described in Section 3. We evaluated retrieval performance with these graphs using both the adjacency and the laplacian spectra. The results are summarised in Table 1. For both experiments, we computed some retrieval statistics, namely nearest neighbor, first and second tiers, each averaged over all possible queries. These are frequently used in information retrieval.

The first figure is the percentage of correct *nearest neighbors (NN)*, i.e. the number of cases in which the top ranked database item, discarding the query itself, belongs to the same class as the query. We also computed the *first tier*, i.e. how many melodies of the query's class are returned within the first $K-1$ matches, where K is the size of the query class. A similar performance figure is the *second tier*, i.e. how many melodies of the query's class are returned within the first $2(K-1)$ matches. The laplacian spectral method performs best with a NN score of 66%, a 1^{st} tier score of 44% and a 2^{nd} tier score of 63%.

Although these performance figures show in general the efficacy of the method, there are some interesting cases in particular we would like to point out here. In Figure 1 there is a special case of an "almost false" positive: for query OGL19205 (belonging to "Heer Halewijn - 3^{rd} version"), the nearest neighbor is OGL19107, that belongs to the group "Heer Halewijn - 4^{th} version". However, the nearest neighbor is somehow related to the query; coincidentally they share the same graph representation, as is shown in Figure 3. This example shows how two melodies can be identical from the interval connectivity point of view but can also be perceptually quite different. This may represent the main limitation of this method in perceptual similarity tasks. The second example (Figure 2) shows the nearest neighbors for the query song OGL19406. Both examples may suggest also that in the case of folksongs people tend to remember more the interval connectivity than the actual intervals of the melody.

Furthermore, we experimented with weighting the edges based on the interval they represent. For this purpose, two different sets of weights were used: one reflecting the difference in notes on the chromatic scale (ignoring differences in octaves) and one reflecting the harmonics of the interval, giving larger weights to consonant intervals and smaller weights to dissonant intervals. During this round of experiments, these methods did not improve the results obtained with normal laplacian spectra.

Using the same test corpus and performance measures we compared our method to the optimal distance measure that was established by Müellensiefen and Frieler [10]. These results are also presented in Table 1, under the name Opti3. This distance measure is a weighted combination of three distance measures, each working on different feature sets. These measures are *harmcore* (using harmonic correlation), *rhythfuzz* (using fuzzified rhythm values) and *ngrukkon* (taking into account characteristic motives). This combined distance measure was established empirically out of 50 building blocks, by searching for a weighted combination whose performance best reflected the results of an extensive human listening experiment. Consequently, this method has been fitted to the data set at hand, explaining why the results are not optimal in our experiment. We also compared our method to the Earth Mover's Distance (EMD). This distance measure takes two weighted point sets as input, and measures the minimum amount of work needed to transform one into the other by moving weight. The EMD is used in a number of different contexts; in the musical case, as pointed out in [12], the (2 dimensional) weighted point set is represented by the score itself, where the weight assigned to each note is its duration. However, since our method only takes into account the global melodic structure, we projected the weighted points on the pitch axis prior to computing the transportation distances. The "Proportional Transportation Distance" (PTD) is a modification of the EMD in order to get a similarity measure based

[2] Equivalent with "tune family" and "Melodietyp".

on weight transportation such that the surplus of weight between two point sets is taken into account.

Figure 1. Example of false positive for the query song "Heer Halewijn" (3^{rd} version) OGL19205 with its NN, OGL19107, instance of "Heer Halewijn" (4^{th} version).

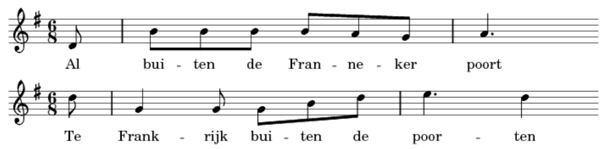

Figure 2. Example of true positive for the query song "In Frankrijk buiten de poorten" (2^{nd} version) OGL19406 with its NN, OGL41709.

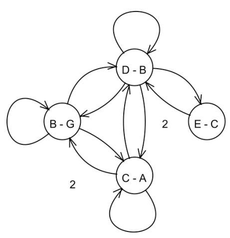

Figure 3. Graph representation of the folk songs OGL19205 and OGL19107 (see Figure 1). The two letters in each circle represent the pitch classes respectively in the first and in the second song.

6 CONCLUDING REMARKS

We presented a graph spectral approach that is new to music retrieval. Our method is focussed on the intervallic structure of the melody. This structure is encoded in a graph whose vertices correspond to the 12 pitch classes and whose edges reflect the interval sequence of the melody; an edge is added to the graph if the pitch classes of the corresponding vertices appear consecutively in the melody. The graphs are indexed into a database using their laplacian spectra, a feature vector that reflects the original topology and graph structure to a large extent.

We evaluated our approach using a subset of a large collection of Dutch folk songs. On this test corpus, our method clearly outperforms existing methods. It is our intention to investigate this method further, for instance by weighting the edges with the duration of the target note and to extend the test corpus. Furthermore we feel that the results can improve by incorporating more detailed musical features.

7 ACKNOWLEDGMENTS

This research was supported by the FP6 IST project 511572-2 PROFI and WITCHCRAFT NWO project 640-003501. The authors wish to thank Peter van Kranenburg (Universiteit Utrecht) and Ellen van der Grijn (Meertens Instituut). Many thanks also to Daniel Müllensiefen and Klaus Frieler for making available the SIMILE package, which includes Opti3.

8 REFERENCES

[1] *Onder de groene linde*. Uitgeverij Unipers, 1987-1991.

[2] Béla Bollobás. *Modern graph theory*. Springer-Verlag, New York, 1998.

[3] M.F. Demirci, R.H. van Leuken, and R.C. Veltkamp. Indexing through laplacian spectra. *Computer Vision and Image Understanding*, To appear, 2007.

[4] C.D. Godsil and B.D. McKay. Constructing cospectral graphs. In *Aequationes Mathematicae*, pages 257–268, 1982.

[5] Chris Godsil and Gordon Royle. *Algebraic Graph Theory*, volume 207 of *Graduate Texts in Mathematics*. Springer Verlag, 2001.

[6] W. H. Haemers and E. Spence. Enumeration of cospectral graphs. *Eur. J. Comb.*, 25(2):199–211, 2004.

[7] Meertens Instituut. Nederlandse liederenbank. http://www.liederenbank.nl/.

[8] P. Leonardo. A Graph Topological Representation of Melody Scores. *Leonardo Music Journal*, 12(1):33–40, 2002.

[9] B. Mohar. The laplacian spectrum of graphs. In *Sixth International Conference on the Theory and Applications of Graphs*, pages 871–898, 1988.

[10] D. Müllensiefen and K. Frieler. Optimizing measures of melodic similarity for the exploration of a large folk song database. In *Proc. of ISMIR*, 2004.

[11] A. Pinto and G. Haus. A novel xml music information retrieval method using graph invariants. *ACM Transactions on Information Systems*, 2007.

[12] Remco C. Veltkamp. Rainer Typke, Frans Wiering. Transportation distances and human perception of melodic similarity. *Musicae Scientiae, Discussion Forum 4A, 2007 (special issue on similarity perception in listening to music), p. 153-182.*

[13] A. Shokoufandeh, D. Macrini, S. Dickinson, K. Siddiqi, and S.W. Zucker. Indexing hierarchical structures using graph spectra. *Pattern Analysis and Machine Intelligence*, 27(7), 2005.