# Seven Years of Music Retrieval @ UU

## Remco C. Veltkamp, Frans Wiering Multimedia and Geometry

## Introduction

Suppose you are looking for music on the Web. It would be nice to have a search engine that helps you find what you are looking for. An important task of such a search engine is to find melodies that are sounds like the tune you have in your head. Most current music search engines work on the basis of song title and artist name, not on melodies. In our research we are looking at the fundamental issues of how to compute the similarity between pieces of music, and how to retrieve music from a collection that is similar to a query.

The music information retrieval research in our department started in the Spring of 2002. In the coffee lounge, Frans Wiering and Remco Veltkamp were discussing melodic and harmonic "shape" in music, drawing diagrams on the whiteboard. We considered the possibility to model music notes as weighted points, and compute similarity by matching two sets of weighted points. Frans first drew a diagram of Monteverdi madrigal (Ohimè il bel viso) and later transcribed by hand themes from Bach's Das Wohltemperierte Klavier II (BWV 870-893), together with another cluster of similar themes from Die Kunst der Fuge (BWV 1080) into weighted point sets. Panos Giannopoulos, doing his PhD research in shape matching, did the first feasibility tests using the Earth Mover's Distance on these weighted points as similarity measure [5]. The results were convincing enough to start a whole new PhD project on this topic. That proved to be the start of a whole line of research. After the seminal test by Panos, a range of other researchers have worked in a number of projects. Figure 28 lists the people involved in the research. In the next sections we describe some of the work done in the various music retrieval projects.

## Orpheus

The first project on music retrieval we ran was Orpheus. The name of the project is an acronym,

161



Figure 28: People that worked on music information retrieval: Panos Giannopoulos, Rainer Typke, Martijn Bosma, Bas de Haas, Peter van Kranenbrug, Anja Volk, Jörg Garbers, Reinier van Leuken, Frans Wiering, Remco Veltkamp.

standing for "On-line Retrieval from Polyhymnia: the Human-oriented Experimental Utrecht Searcher". The project is of course named after the Argonaut Orpheus, the greatest musician in Greek mythology, whose songs could even charm wild beasts. When Orpheus' wife, Eurydice, was killed by the bite of a serpent, he descended into the underworld to bring her back. His songs were so beautiful that Hades, the lord of the dead and ruler of the netherworld, finally agreed to allow Eurydice to return to the world of the living. This looks like a fine example of a successful retrieval, a 'true positive'. However, Orpheus had to meet one condition: he must not look back as he was conducting her to the surface. Just before the pair reached the upper world, Orpheus looked back, and Eurydice slipped back into the netherworld once again.

In the Orpheus project, we continued the work on comparing melodies by matching weighted points sets [7]. Each note is represented in a twodimensional point with a weight that denotes its importance. The more important the note in the melody, the larger the weight. With the so-called Earth Mover's Distance we compute to what extent two weighted point sets are similar. One set of notes is considered a set of piles of earth, the other as a set of holes. The Earth Mover's Distance computes the minimal cost flow, the minimal amount of transportation effort needed to fill the holes with the piles of earth, which is a measure of how similar the two sets of weights are. When the amounts of earth in the piles and capacities of the holes are close to each other, it takes little effort to fill the holes with earth, and the weighted points (and thus the melodies) are similar. Figure 29 gives an illustration of the flow from the piles (upper row) to the holes (lower row). If you listen to the two melodies, they sound very similar, and also the transportation effort, the Earth Mover's Distance, is small.

Rainer Typke developed the Orpheus search engine based on this method. It was tested on a collection of about 480.000 'incipits', short fragments from a catalogue of music works. A substantial number of these works, about 80,000, are listed as anonymous. We have used our method to assign candidate composers to these anonymous incipits, by comparing them to all other incipits in the collection. If the comparison would have been done by matching each of the 80,000 anonymous incipits with the other 400,000 ones, that would take 32,000,000,000 comparison computations. If a single comparison would take 1 ms, the whole procedure would take more than a year! Therefore, we have applied the 'vantage' indexing method developed in our group [8], so that only a small number of expensive Earth Mover's Distance computations is necessary, while the rest of the comparisons are replaced by efficient Euclidean distance computations. Using this indexing method, comparing all the anonymous incipits only take a few minutes. In this way, we have found candidate composers for 18,000 anonymous incipits: see figure 30 for an example.

That the method works well was also demonstrated in the international Music Information Retrieval Exchange (MIREX) 2006, were we ranked first in the track on comparing notated melodies [6].

## **C**-Minor

The Orpheus project was technology driven, developing and experimenting with a the transportation effort based similarity measure. Since music has such intricate perceptual and cognitive aspects, it makes sense to take these aspects into account. We therefore started a new project: C-Minor, Cognition-based Music Information Notation Oriented Retrieval.



Figure 29: Two melodies, their representation as weighted point sets, and the computed minimum cost flow.



Figure 30: Two versions of Mozart's "Ah! vous dirai-je Maman" (also known as Twinkle, twinkle little star, or Altijd is Kortjakje ziek) that were found.

In order to evaluate various similarity measures and query modes, Martijn Bosma developed the experimentation platform YahMuug!e (Yet Another Heterophone: Musical Utrecht University Global Lookup Engine), yahmuugle.cs.uu.nl [1]. At this moment, the user can choose four different interfaces to specify a query: a software keyboard, a software keyboard with durations, a file upload, and a Query-by-Humming interface. With the Query-by-Humming interface, a user hums or whistles a melody into a microphone. Whistling works better, because that yields a sharper sound. With the Fast Fourier Transform the audio signal is transformed into frequencies, and then into a pitch representation. Since it is not easy to whistle a melody accurately, an entered query melody can be subsequently tuned with a pianoroll editor, see figure 31. The user can insert, delete, and move notes, and also change note durations.

We have worked on a number of other similarity measures besides the Earth Movers' Distance. Bas de Haas developed a new distance function that measures the difference between chord progressions [2]. This distance function is based on Lerdahl'Šs Tonal Pitch Space. It compares the harmonic changes of two sequences of chord labels over time. This distance has proven to be effective for retrieving similar jazz standards found in the Real Book. That it matches human intuitions about harmonic similarity was demonstrated on a set of blues variations.

Reinier van Leuken developed 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 [4]. These graphs are then indexed into a database using their laplacian spectra (the eigenvalues of laplacian matrix representation of the graph) 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 presented an experimental evaluation of this approach, together with a comparison with two known retrieval techniques. On our test corpus, a subset of a welldocumented and annotated collection of Dutch folk songs (Onder de Groene Linde, see next section), this evaluation demonstrated the effectiveness of the overall approach.

## Witchcraft

After a tecnology-based and a cognition-based project, we started a domain-based project. In



Figure 31: YahMuug!e whistle and pianoroll editor interface.

the project WITCHCRAFT (the acronym stands for What Is Topical in Cultural Heritage: Contentbased Retrieval Among Folksong Tunes) we collaborate with the Meertens Institute. The project is one of 14 NWO-CATCH-projects in which computer science departments and cultural heritage institutions collaborate in the design of novel access methods to cultural heritage. The Meertens Institute researches and documents Dutch popular culture. One of its topics is Dutch popular song. Over the years, an important database has been created for this purpose, the Nederlandse Liederenbank (Dutch Song Database, www.liederenbank. nl), containing metadata about ca. 125,000 songs dating from the Middle Ages to the present. However, the musical content of the songs could not be searched, which is a serious shortcoming for both researchers and the general public. The aim of the WITCHCRAFT project is to remedy this situation. The project's overall aim is to design content-based retrieval methods for large collections of melodies stored as audio and notation. There are three aspects to this aim:

- Practical: the creation of a melody search engine.
- Scientific: the design and evaluation of methods for measuring musical similarity.

• Musicological: enabling research into the oral transmission of folk songs.

#### Onder de Groene Linde

The principal materials studied in the WITCHCRAFT project come from a song collection known as Onder de Groene Linde (Under the Green Lime Tree). Most of the songs are ballads, strophic songs with narrative content. For the song called "In Frankrijk daar staat er een herberg" (In France there is an inn) the story goes as follows:

Long ago, there was an inn outside the gate of a market town. Merchants meet there in the evening for drinking, food and enjoying female company. Everybody participates in the dissolute behaviour except one pious maidservant. One morning she finds a dead baby in her bed. Obviously, she has a secret lover, got pregnant and killed the baby. Therefore she is sentenced to the be hanged. But then a miracle happens: angels keep her alive while she is hanging from the gallows. This proves that she is innocent and she is set free. The real murderer is soon identified: the innkeeper's daughter. After killing her illegitimate baby she tried to put the blame on the maidservant. The daughter and her mother, who invented the evil plan, were hanged.

Such songs used to be sung during manual labour until the 1930s. Singing helped to synchronise movement (for example in pile driving), prevented gossiping and relieved the monotony of the work. Songs were learned by listening and participating: they were seldom written down. In this process of oral transmission the songs underwent continuous change, both in text and in melody.

As a consequence of labour mechanisation and the introduction of the radio, ballad singing has completely disappeared from the Netherlands. That we still have access to this tradition is the consequence of a long-term effort to record these songs, started by Will Scheepers in the early 1950s and continued by Ate Doornbosch. They collected around 7000 field recordings, which were usually broadcasted in the radio programme Onder de groene linde (1957-1994). This programme was an early example of interactive radio. Listeners were encouraged to contact Doornbosch if they knew more about the songs that were played. Doornbosch would then record their version an broadcast it. In this manner a unique collection was created that documents, in addition to a now disappeared aspect of Dutch cultural heritage, the textual and melodic variation that results from oral transmission. The melodic variation makes this collection

also very interesting for music information retrieval research.

#### **Corpus creation**

Given the fact that the songs are available as audio recordings, it would seem that the most logical approach to searching the songs would be to digitise these recordings and to extract features for matching from these. This proved not to be an option for two reasons. One is that audio-based methods are not developed far enough yet to support reliable matching of the kinds of high-level features one needs for retrieving related melodies. The other reason is the quality of the recordings. Although the technical quality is often already problematic, the real problem resides in the singers: their wornout, uneducated voices are often unstable as regards pitch and rhythm. Even for specialists it is often hard to determine their musical intention. The good news is that specialists actually did so: they wrote out some 5000 songs in music notation, on paper. These 'transcriptions' can relatively easily be encoded in a searchable music format.

Corpus creation was a necessary precondition for WITCHCRAFT. For this task, Jörg Garbers and Peter van Kranenburg developed the WitchCraftEditor (WCE), which in turn relies on the LilyPond software of former ICS PhD student Han-Wen Nienhuys (www.lilypond.org). Songs are encoded in a subset of LilyPond and then stored MIDI, \*\*kern (an encoding system for music analysis, see kern.humdrum.net) and of course as image of the notation, as illustrated in Figure 32.

The actual corpus has been created mainly in a project of the Meertens Institute, Dutch Songs as Musical Content. It consists now of around 2500 folk songs from Onder de Groene Linde, 1500 folk songs from written sources, and 1500 instrumental melodies, mostly dances, also from written sources.

#### **Ground Truth for Similarity Measures**

When are two melodies similar? In the end this is a subjective judgement; nevertheless, certain patterns can be discerned in the way listeners make their judgement. Important factors are rhythm, pitch succession and melody segmentation. 'Local' changes such as addition, deletion or substitution of notes, ornamentation, tempo differences and transposition have little influence on the perception of similarity.

During the corpus creation, musical experts from the Meertens Institute have classified the a large part of the melodies into groups called 'melody norms.' The members of a melody norm are melodies that, in the experts' opinion, derive from a common ancestor. When however historical evidence is lacking (as is mostly the case in orally transmitted folksongs) experts base their judgement on musical similarity between melodies. In order to understand how experts perform this classification task, an experiment was done involving a subset of melody norms. It appears that experts employ family resemblance, which is based on the presence of a set of features that apply most of the time, but not necessarily always, to group members. Some groups cohere mainly by rhythm, others by melody contour or the occurrence of a salient musical pattern [10]. In the testing of similarity measures, melody norms can be used as a ground truth, representing the ideal answer set for musical queries. The fact that melody norms are constituted in very different ways make it very unlikely that a single similarity measure will be able to emulate human melody norm assignment. Consequently, a useful music search engine must possess a number of complementary similarity measures.



Figure 32: In Frankrijk daar staat een herberg (NLB 73515): (top) manual transcription, (middle) LilyPond encoding, (bottom) computer-generated score.

#### Similarity measures

Initial retrieval tests on the Onder de groene linde corpus were done using the Earth Mover's Distance (see above). The results were generally quite reasonable, but often properties of the measure itself caused musically unrelated items to appear quite prominently amongst the retrieved items. Therefore alternatives have been investigated. One of these is Inner Metric Analysis, a method developed by Anja Volk. This method measures the contribution of each note to the formation of rhythmicmetric patterns. One measure the similarity between two melodies by calculating the correlation between their IMA patterns [9] (see Figure 33).

Another approach, currently investigated by Peter van Kranenburg, uses string alignment methods to establish the similarity between melodies. Melodies are represented as strings, for example of characters representing pitches. In more complex representations durations are also accounted for. The general idea of string alignment is to minimise the total penalty incurred by substitutions, deletions and gap insertions. When applied to folk songs, the aim is to choose the penalties in such a way that the more likely a change is to occur in oral transmission of melodies, the smaller the penalty. Figure 34 gives an example of the alignment of two melodies. By providing different (combinations of) raters that determine the cost of substitutions, deletions and gap insertions, searching can be adapted to different kinds of musical similarity. A variation of the string matching approach is to use a group of melodies as the query [3]. In this way one can use known members of a melody norm to retrieve new candidates.

#### Search engine

As a first step in the creation of a music search engine for the Nederlandse Liederenbank, the Yah-Muugle search engine has been integrated by Jörg Garbers into the Liederenbank. At this stage the similarity measure is still the Earth Mover's Distance, but this will change in the course of 2009, when other measures will be added as well. Figure 35 shows the entry of the song discussed above. It has a link "find similar melodies" that retrieves a list of musically related items, the top of which is shown in Figure 35.

Queries can also be newly created using software keyboards or Query-By-Humming, using methods that have been developed in the C-Minor project.



Figure 33: Establishing the similarity between two melodies using IMA patterns.



Figure 34: String alignment of two melodies

## **Future**

By working closely together with musicologists from the Meertens Institute, we were able to incorporate a lot of musical domain knowledge in the design of novel music retrieval methods. At the end of the WITCHCRAFT project we plan to offer two interfaces: one for general users, offering the generally best-performing similarity measure and a sim-

ple search interface, and one offering a full suite of query entry methods and similarity measures designed to support the domain experts in their research. While the research into melodic similarity measures was motivated first of all by their wishes, the similarity measures have been designed with more general application in mind and will form the basis of future MIR projects.



Figure 35: Liederenbank entry for In Frankrijk daar staat er een herberg (top), and the items retrieved by the melody search (bottom).

We have plans to further develop our research into recommendation services based on the meaning that music has to people. The specific meaning that a piece of music has to someone is a personal matter by definition. Therefore scientists and scholars have often been reluctant to study musical meaning. But as the meaningfulness of music is the fundamental reason why people engage in musical activities, the issue cannot be ignored. There are obvious regularities in how listeners experience music. We regard such regularities as high-level musical features. The modelling and matching of such features is an important direction in research that aims at creating high-quality computational methods for accessing musical content.

So there remains enough to be researched for the next seven years. We strive to provide the music information community with many exciting new results. At least one thing we are sure we wil provide to this community is the organization of the International Conference on Music Information Retrieval (ISMIR) in Utrecht in 2010.

### Acknowledgments

This work has been partially supported by the Dutch ICES/KIS III bsik project MultimediaN, and by the Netherlands Organisation for Scientific Research (NWO) project 640-003-501.

## **Bibliography**

- M. Bosma, R. C. Veltkamp, and F. Wiering. Muugle: A modular music information retrieval framework. In *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR 2006)*, 2006.
- [2] B. de Haas, R. C. Veltkamp, and F. Wiering. Tonal pitch step distance: a similarity measure for chord progressions. In *Proceedings of the Eighth International Conference on Music Information Retrieval (ISMIR)*, 2008.
- [3] J. Garbers and F. Wiering. Towards structural alignment of folk songs. In *Proceedings of the Eighth International Conference on Music Information Retrieval (ISMIR)*, 2008.

- [4] A. Pinto, R. H. van Leuken, M. F. Demirci, F. Wiering, and R. C. Veltkamp. Indexing music collections through graph spectra. In *Proceedings of the Seventh International Conference on Music Information Retrieval* (ISMIR), pages 153–156, 2007.
- [5] R. Typke, P. Giannopoulos, R. C. Veltkamp, F. Wiering, and R. van Oostrum. Using transportation distances for measuring melodic similarity. In *Proceedings of the Fourth International Conference on Music Information Retrieval (ISMIR)*, pages 107–114, 2003.
- [6] R. Typke, F. Wiering, and R. C. Veltkamp. Mirex symbolic melodic similarity and query by singing/humming. In *Proceedings MIREX 2006*, 2006.
- [7] R. Typke, F. Wiering, and R. C. Veltkamp. Transportation distances and human perception of melodic similarity. *Musicae Scientiae*, Discussion Forum 4A, 2007 (special issue on similarity perception in listening to music), pages 153–182, 2007.
- [8] J. Vleugels and R. C. Veltkamp. Efficient image retrieval through vantage objects. *Pattern Recognition*, 35(1):69–80, 2002.
- [9] A. Volk. Persistence and change: local and global components of metre induction using inner metric analysis. *Journal of Mathematics and Computation in Music*, 2(2):99–115, 2008.
- [10] A. Volk, P. van Kranenburg, J. Garbers, F. Wiering, R. C. Veltkamp, and L. P. Grijp. A manual annotation method for melodic similarity and the study of melody feature sets. In *Proceedings of the Eighth International Conference on Music Information Retrieval (ISMIR)*, 2008.