Melody Retrieval using the Implication/Realization Model

Maarten Grachten, Josep Lluís Arcos and **Ramon López de Mántaras**
IIIA, Artificial Intelligence Research Institute
CSIC, Spanish Council for Scientific Research
Campus UAB, 08193 Bellaterra, Catalonia, Spain
{maarten,arcos,mantaras}@iiia.csic.es

ABSTRACT

We present a method for melody retrieval using the Implication/ Realization (I/R) model, a model of melodic structure. As a preprocessing step to retrieval, all melodies to be compared are analysed using our I/R parser. The similarity ranking of melodies from data base melodies with respect to the query is determined by computing the edit distance between the I/R analysis of the query and those of the data base melodies. The parameters in the distance measure were optimized by a genetic algorithm, evaluating on melodic similarity ground truth. In the MIREX '05 contest for symbolic melodic similarity, our method was compared to various other methods using the RISM A/II database, and performed best. The results of the contest are presented and discussed at the end of the paper.

Keywords: Melody Retrieval, Symbolic Melodic Similarity, Implication/Realization Model, Edit Distance, MIREX

1 INTRODUCTION

We present a method for melody retrieval using the Implication/ Realization (I/R)[1] model [Narmour (1990)], a model of melodic structure. This model characterizes consecutive melodic intervals by the expectation they generate with respect to the continuation of the melody, and whether or not this expectation is fulfilled. The model states a number of data driven principles that govern the expectations. We have used the most important of these principles to implement an Implication/Realization parser for monophonic melodies. The output of this parser is a sequence of labeled melodic patterns, so called I/R structures. An I/R structure usually represents two intervals (three notes) Eighteen basic I/R structures are defined using labels that signify the implicative/realizing nature of the melodic fragment described by they I/R structure. Apart from its I/R label, a note pattern that has been recognized as particular I/R structure can be further characterized by features like the melodic direction of the pattern, the amount of overlap between consecutive I/R structures, and the number of notes spanned.

[1]Not to be confused with IR, the widely used acronym for Information Retrieval

Since the I/R analysis of the melody is of sequential nature, it allows for comparison using known string comparison methods. We use the edit distance to compare sequences of I/R structures. The measure allows for deletion, insertion, and replacement of I/R structures. The parameters in the cost functions are optimized to predict known melodic similarity ground truth, using a genetic algorithm.

In the rest of this paper we will give some background and details of the method outlined above. Additionally, we will pay attention to the MIREX contest for symbolic melodic similarity (part of the ISMIR 2005 conference). In this contest, the performance of our method and various other methods for melody retrieval are compared. At the end of the paper, the results of the contest are presented.

2 THE IMPLICATION/REALIZATION MODEL

Eugene Narmour proposed a theory of perception and cognition of melodies, the Implication/Realization model, or I/R model [Narmour (1990, 1992)]. According to this theory, the perception of a melody continuously causes listeners to generate expectations of how the melody will continue. The sources of those expectations are twofold: both innate and learnt. The innate sources are 'hardwired' into our brain and peripheral nervous system, according to Narmour, whereas learnt factors are due to exposure to music as a cultural phenomenon, and familiarity with musical styles and pieces in particular. The innate expectation mechanism is closely related to the *gestalt theory* for visual perception [Koffka (1935); Köhler (1947)]. Gestalt theory states that perceptual elements are (in the process of perception) grouped together to form a single perceived whole (a 'gestalt'). This grouping follows certain principles (*gestalt principles*). The most important principles are *proximity* (two elements are perceived as a whole when they are perceptually close), *similarity* (two elements are perceived as a whole when they have similar perceptual features, e.g. color or form, in visual perception), and *good continuation* (two elements are perceived as a whole if one is a 'good' or 'natural' continuation of the other). Narmour claims that similar principles hold for the perception of melodic sequences. In his theory, these principles take the form of *implications*: Any two consecutively perceived notes constitute a melodic interval, and

if this interval is not conceived as complete, or closed, it is an *implicative interval*, an interval that implies a subsequent interval with certain characteristics. In other words, some notes are more likely to follow the two heard notes than others. Two main principles concern *registral direction* and *intervallic difference*. The principle of registral direction (PRD) states that small intervals imply an interval in the same registral direction (a small upward interval implies another upward interval, and analogous for downward intervals), and large intervals imply a change in registral direction (a large upward interval implies a downward interval and analogous for downward intervals). The principle of intervallic difference (PID) states that a small (five semitones or less) interval implies a similarly-sized interval (plus or minus two semitones), and a large intervals (seven semitones or more) implies a smaller interval.

Based on these two principles, melodic patterns can be identified that either satisfy or violate the implication as predicted by the principles. Such patterns are called *structures* and labeled to denote characteristics in terms of registral direction and intervallic difference. Eight such structures are shown in figure 1(top). For example, the P structure ('Process') is a small interval followed by another small interval (of similar size), thus satisfying both the PRD and the PID. Similarly the IP ('Intervallic Process') structure satisfies the PID, but violates the PRD. Some structures are said to be *retrospective* counterparts of other structures. They are identified as their counterpart, but only after the complete structure is exposed. In general the retrospective variant of a structure has the same registral form and intervallic proportions, but the intervals are smaller or larger. For example, an initial large interval does not give rise to a P structure (rather to an R, IR, or VR, see figure 1, top), but if another large interval in the same registral direction follows, the pattern is a pair of similarly sized intervals in the same registral direction, and thus it is identified as a retrospective P structure, denoted as (P).

Additional principles are assumed to hold, one of which concerns *closure*, which states that the implication of an interval is inhibited when a melody changes in direction, or when a small interval is followed by a large interval. Other factors also determine closure, like metrical position (strong metrical positions contribute to closure), rhythm (notes with a long duration contribute to closure), and harmony (resolution of dissonance into consonance contributes to closure).

We have designed an algorithm to automate the annotation of melodies with their corresponding I/R analyses. The algorithm implements most of the 'innate' processes mentioned before. It proceeds by computing the level of closure at each point in the melody using metrical and rhythmical criteria, and based on this, decides the placement and overlap of the I/R structures. For a given set of closure criteria, the procedure is entirely deterministic and no ambiguities arise. The learnt processes, being less well defined by the I/R model, are currently not included. Nevertheless, we believe that the resulting analysis have a reasonable degree of fidelity with respect to the I/R model. An example analysis is shown in figure 1(bottom).

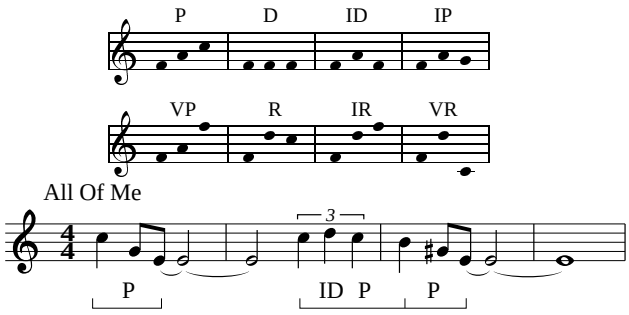

Figure 1: Top: Eight of the basic structures of the I/R model. Bottom: First measures of *All of Me* (Marks & Simons, 1931), annotated with I/R structures

3 AN EDIT DISTANCE FOR I/R REPRESENTATIONS

The edit distance or Levenshtein distance [Levenshtein (1966)] is a well known measure to compare sequential data. It has been applied in fields such as DNA analysis, automated spell checking, and is also commonly used in music computing. Musical applications include score following [Dannenberg (1984)], and melodic similarity computation (typically for melody retrieval). The edit distance has been applied to plain sequences of notes [e.g. Mongeau and Sankoff (1990); Smith et al. (1998)], but also to sequences of pitch intervals and contours [e.g. Müllensiefen and Frieler (2004); Grachten et al. (2004)], and even to tree representations of melodies [Rizo and Iñesta Quereda (2002)].

Since the I/R representations of melodies are of sequential nature, the distance between them can perfectly be assessed using the edit distance. To do this it is necessary to define the edit operations that can be applied to the elements in the sequences, and the functions that compute the costs of such operations. Although specialized operations such as consolidation and fragmentation have been proposed for computing the edit distance between note sequences [Mongeau and Sankoff (1990)], we decided to use the three traditional edit operations, insertion, deletion, and replacement. Costs of deletion and insertion of I/R structures are proportional to the number of notes spanned by the I/R structure, and replacement is a weighted sum of differences between features, plus an additional cost if the I/R structures under replacement do not have the same label. The latter cost is reduced if the labels are semantically related, that is one of the structures is the retrospective counterpart of the other.

3.1 Parametrization of Edit Operation Costs

The weight functions for computing the cost of edit operations are parametrized to allow for control and finetuning of the edit distance.

$$w(s_i, \emptyset) = \alpha_d \cdot \mathcal{S}ize(s_i)$$
$$w(\emptyset, s_j) = \alpha_i \cdot \mathcal{S}ize(s_j)$$

$$w(s_i, s_j) = \alpha_r \cdot \begin{pmatrix} \beta \cdot |\mathit{LabelDiff}(s_i, s_j)| + \\ \gamma \cdot |Size(s_i) - Size(s_j)| + \\ \delta \cdot |Dir(s_i) - Dir(s_j)| + \\ \epsilon \cdot |Overlap(s_i) - Overlap(s_j)| \end{pmatrix}$$

$$\mathit{LabelDiff}(s_i, s_j) = \begin{cases} 0 & Label(s_i) = Label(s_j) \\ \zeta & Label(s_i) = -Label(s_j) \\ 1 & otherwise \end{cases}$$

$w(s_i, \emptyset)$ is the cost of deleting I/R structure s_i from the source sequence, $w(\emptyset, s_j)$ is the cost of inserting I/R structure s_j into the target sequence, and $w(s_i, s_j)$ is the cost of replacing element s_i from the source sequence by s_j from the target sequence.

$Label$, $Size$, Dir, and $Overlap$ are functions that, given an I/R structure, respectively return its I/R label (encoded as an integer), its size (number of notes spanned), its melodic direction, and its overlap (the number notes belonging to both the current I/R structure and its successor). $\mathit{LabelDiff}$ is an additional function that determines the part of the replacement cost due to difference/equality of I/R labels. The I/R labels are mapped to integer values in such a way that the integer for the retrospective counterpart of a particular label is always the negative of the integer of that label.

The parameters come in two kinds: Firstly there are the parameters that are used to control the relative costs of the operations, α_i, α_d, and α_r. For example by setting α_i and α_d to relatively low values, the optimal alignment is more likely to include insertions and deletions of elements than replacements. The second kind of parameters, including β, γ, δ, ϵ, and ζ, are for controlling the cost of replacing one I/R structure by another as a function of the difference in attributes.

3.2 Evolutionary Optimization of Edit Distance Parameters

The choice of parameter values was determined using similarity rankings for 11 melodic incipits by human subjects [Typke et al. (2005)]. The subjects were asked to rank about 50 incipits for each query according to their similarity to that query. All incipits were extracted from the RISM A/II database. More details are given in section 5. The rankings for the queries were used to evaluate parameter settings, by calculating the ranking with the given parameter setting, and compare the computed ranking with the ranking by the human subjects. The four evaluation metrics used in this comparison were the same as the ones used in the MIREX '05 contest (see section 5).

A genetic algorithm [Goldberg (1989)] was then applied to search the space of parameter settings. The chromosomes contained the values of the parameters 8 parameters described above, and new populations were generated by an random initial population of chromosomes using an elitist approach. New chromosomes were formed by mutation and crossover of existing chromosomes. The fitness of a chromosome was defined to be the average of the outcomes of the four evaluation metrics for the corresponding parameter setting. A cross-validation setup was chosen on the 11 queries, to prevent overfitting of the parameters to the train data. The best (normalized) parameter settings found in this setup are shown in table 1.

parameter	operation/attribute	value
α_i	insertion	0.064
α_d	deletion	0.131
α_r	replacement	1.000
β	labels	0.587
γ	size	0.095
δ	direction	0.343
ϵ	overlap	0.112
ζ	retrospective counterparts	0.801

Table 1: Parameter values found by evolutionary optimization, using train data from RISM A/II database

4 GENERAL CHARACTERIZATION OF THE I/R BASED SIMILARITY MEASURE

As a representation of melodic material, the I/R analysis provides an intermediate level of abstraction from the melodic surface, between a note representation as a less abstract representation, and the pitch contour (up/down patterns) representation as being more abstract. Although the labels given to the I/R structures merely represent pitch interval relations, the overlap and boundaries of the structures convey information about meter and rhythm, be it in an implicit way.

In previous work [Grachten et al. (2004)], we studied the behavior of the I/R similarity measure described here in comparison to other edit distance based measures, using note and pitch interval/contour data. We studied the similarity distributions of each measure on a set of 124 jazz melodies. The ordering of melody representations in terms of abstraction (the I/R analysis being intermediate) was reflected in tests of entropy and divergence of the distance distributions of the various measures.

There appears to be a trade-off between discriminatory power on the short range of melodic similarity on the one hand, and discriminatory power on the long range of similarity on the other. Similarity measures based on more concrete melody representations tend to favor the former and those based on more abstract melody representations the latter. In terms of applications, concrete measures would be more suitable to find the single best match for a query (e.g. to implement Google's "I'm feeling lucky" functionality), whereas abstract measures would be more useful for multidimensional scaling of a set of melodies.

5 THE *MIREX '05* CONTEST FOR SYMBOLIC MELODIC SIMILARITY

In this section we will give a short overview of the results of the symbolic melodic similarity contest, that formed part of the Music *Information Retrieval Evaluation eXchange* (MIREX '05) event, and to which we submitted the I/R based similarity measure.

The contest task consisted in ranking a subset (558 incipits) of the RISM A/II database (a database containing

bibliographic records of musical manuscripts written after 1600), according to melodic similarity against 11 query incipits. Rankings for another 11 queries were available to the participants as train data. The melody incipits were available as MusicXML. MIDI versions were also available. Grace notes were removed in the MIDI versions, since including them without altering the durations of surrounding notes would break the time structure of the melody (since the grace notes would incorrectly consume time).

The ranking computed by the participant algorithms were compared to rankings by human subjects. Those rankings were constructed from the subjects individual rankings according to a procedure documented in [Typke et al. (2005)]. Whenever the order between subject ranked incipits was not statistically significant, the corresponding incipits were grouped together. The evaluation metrics for comparison of computed rankings and target rankings were chosen such that the computed ranking was not penalized for changing the order of incipits whose order was not significant in the target ranking. The following metrics were used: *average dynamic recall* [Typke (2005)], *normalized recall at group boundaries*, *average precision* (non-interpolated), and *precision at N documents* (N is the number of relevant documents).

The algorithms of other participants include techniques like the Earth Mover's Distance (Typke, Wiering & Veltkamp), geometric matching (Lemström, Mikkilä, Mäkinen & Ukkonen), N-grams matching (Orio), (Suyoto & Uitdenbogerd), and linear combinations of many different (mostly well known) similarity measures (Frieler & Müllensiefen). Some of them (Typke et al., Lemström et al.), use a pitch/duration/onset based melodic representations, while others use pitch intervals (Suyoto & Uitdenbogerd), or a combination of concrete and abstract representations (Frieler & Müllensiefen), (Orio).

As can be seen from table 2, our I/R based similarity measure performs relatively well. It scores highest in all four evaluation metrics. It may be tempting to interpret this as a corroboration of the I/R Model. However some reservations must be made, Firstly, one should bear in mind that the I/R analysis of a melody is hypothesized to express the pattern of listening expectations (and their satisfaction/violation) that the melody generates. Evidence that perceptually similar melodies have similar I/R analyses is not necessarily evidence for this hypothesis. And secondly, the evaluation results are only partly determined by the choice of representation (in our case the I/R analysis), the actual distance metric may have a great impact as well. Nevertheless, the good performance of our method indicates that the I/R analysis provides a relevant and useful representation of melody.

An interesting question is whether combining I/R representations with other distance metrics improves the results. It is surprising to see the relatively good results of Suyoto & Uitdenbogerd, as they apparently use only pitch interval representations, and discard duration information. This implies that either duration and other non-pitch information is irrelevant for melodic similarity (which we consider very unlikely), or that the N-grams counting method they used is very effective. That leads us to conclude that it looks worth wile investigating the possibility of using a distance metric based on matching N-grams of I/R structures. The more so since Orio's N-grams based method also gives good results.

The last column in the table shows run times. In this respect, our algorithm lags behind. But our run time may well be improved, since our focus has not been on computational efficiency. In particular, the preprocessing step that performs the I/R analysis of the midi files is currently implemented as an interpreted Guile/Scheme script, which obviously runs slower than compiled code. Furthermore, we used a C++ implementation of the edit distance that is very generic (e.g. it allows for an arbitrary number of edit operations, and supports context-aware edit operations). Using an edit distance implementation that is more specialized will probably speed up similarity computations.

ACKNOWLEDGEMENTS

We wish to thank Rainer Typke and his colleagues at University of Utrecht for preparing the ground truth experiments, setting up the symbolic melodic similarity contest and evoking discussion. Furthermore, we thank the IMIRSEL group at University of Illinois for organising the MIREX '05 contests, especially Stephen Downie, and Xiao Hu for their help and cooperation. Thank you. This research has been partially supported by the Spanish Ministry of Science and Technology under the project TIC 2003-07776-C2-02 "CBR-ProMusic: Content-based Music Processing using CBR" and EU-FEDER funds.

References

R. Dannenberg. An on-line algorithm for real-time accompaniment. In *Proceedings of the 1984 International Computer Music Conference*. International Computer Music Association, 1984.

D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

M. Grachten, J. Ll. Arcos, and R. López de Mántaras. Melodic similarity: Looking for a good abstraction level. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*. Pompeu Fabra University, 2004.

K. Koffka. *Principles of Gestalt Psychology*. Routledge & Kegan Paul, London, 1935.

W. Köhler. *Gestalt psychology: An introduction to new concepts of modern psychology*. Liveright, New York, 1947.

V. I. Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10:707–710, 1966.

M. Mongeau and D. Sankoff. Comparison of musical sequences. *Computers and the Humanities*, 24:161–175, 1990.

D. Müllensiefen and K. Frieler. Optimizing measures of melodic similarity for the exploration of a large folk

Rank	Participant	Average Dynamic Recall	Normalized Recall at Group Boundaries	Average Precision (non-interpolated)	Precision at N documents	Input Data Format	Runtime (seconds)
1	Grachten, Arcos & Mántaras	65.98%	55.24%	51.72%	44.33%	MIDI	80.174*
2	Orio	64.96%	53.35%	42.96%	39.86%	XML	24.610
3	Suyoto & Uitdenbogerd	64.18%	51.79%	40.42%	41.72%	MIDI	48.133
4	Typke, Wiering & Veltkamp	57.09%	48.17%	35.64%	33.46%	MIDI	51240
5	Lemström, Mikkilä, Mäkinen & Ukkonen (P3)	55.82%	46.56%	41.40%	39.18%	MIDI	10.007*
6	Lemström, Mikkilä, Mäkinen & Ukkonen (DP)	54.27%	47.26%	39.91%	36.20%	MIDI	10.106*
7	Frieler & Müllensiefen	51.81%	45.10%	33.93%	33.71%	MIDI	54.593

Table 2: Results for the MIREX 2005 contest for symbolic melodic similarity, ranked according to Average Dynamic Recall. Note *: These runs were executed in M2K environment, and thus the runtime includes evaluation time

song database. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*. Pompeu Fabra University, 2004.

E. Narmour. *The Analysis and cognition of basic melodic structures : the implication-realization model*. University of Chicago Press, 1990.

E. Narmour. *The Analysis and cognition of melodic complexity: the implication-realization model*. University of Chicago Press, 1992.

D. Rizo and J.M. Iñesta Quereda. Tree-structured representation of melodies for comparison and retrieval. In *Proceedings of International Workshop on Pattern Recognition in the Information Society, PRIS 2002,*, page 155, 2002.

Ll. A. Smith, R. J. McNab, and I. H. Witten. Sequence-based melodic comparison: A dynamic programming approach. In W. B. Hewlett and E. Selfridge-Field, editors, *Melodic Similarity. Concepts, Procedures, and Applications*, Computing in Musicology, pages 101–118. MIT Press, 1998.

R. Typke. Proposed measure for comparing ranked lists, for MIREX '05 (Symbolic Melodic Similarity). Unpublished; Available at: `http://teuge.labs.cs.uu.nl/Ruu/orpheus/groundtruth/measure.pdf`, 2005.

R. Typke, M. den Hoed, J. de Nooijer, F. Wiering, and R.C. Veltkamp. A ground truth for half a million musical incipits. In *Proceedings of the Fifth Dutch-Belgian Information Retrieval Workshop*, pages 63–70, 2005.