Using the Implication/Realization Model for Measuring Melodic Similarity

Maarten Grachten and **Josep-Lluís Arcos**
IIIA-CSIC - Artificial Intelligence Research Institute
CSIC - Spanish Council for Scientific Research
Campus UAB, 08193 Bellaterra, Catalonia, Spain.
Vox: +34-93-5809570, Fax: +34-93-5809661
Email: {maarten,arcos}@iiia.csic.es

Abstract. The concept of melodic similarity has become increasingly relevant in the light of music retrieval and music content processing systems. We propose a new way of measuring melodic similarity, based on analyses of the melody according to the Implication/Realization (I/R) model [7] for melodic structure and cognition. The similarity is assessed as the edit-distance between these I/R analyses. We present some experiments and results, comparing this approach to edit-distance based measures on other melodic representations. We argue that the I/R representation combines some advantages of the other measures, and is a good compromise between concrete and abstract levels of melody representation.

1 Introduction

Computing similarities in sequences of notes is a very general problem with diverse musical applications ranging from music analysis to content-based retrieval. Choosing the appropriate level of representation is a crucial issue and depends on the type of application. For example, in applications such as pattern discovery in musical sequences [1], [3], or style recognition [3], it has been established that melodic comparison requires taking into account not only the individual notes but also the structural information based on music theory and music cognition [9].

Our research interest concerns the development of a CBR system for expressive music processing. In that context (e.g. for retrieval and reuse mechanisms), a well chosen distance measure for melodies is of importance. Some desirable features of such a measure are the ability to distinguish phrases from different musical styles and to recognize phrases that belong to the same song. We propose a new way of assessing melodic similarity, representing the melody as a sequence of I/R structures (conform Narmour's Implication/Realization (I/R) model for melodic structure [7]). The similarity is then assessed by calculating the edit-distance between I/R representations of melodies. We compared this assessment to assessments based on note representations [6], and melodic contour representations [2, 5].

We have found that the discriminatory power of the note level distance measure is much lower than that of the contour and I/R level measures. Also, taking into account interval durations within the contour level measures, tended to decrease the discriminatory power. We argue that the I/R level measure is an appropriate compromise that takes into account rhythmical/temporal information in an implicit way, without losing discriminatory power.

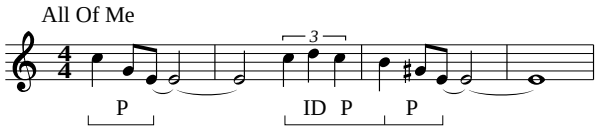

Figure 1. First measures of All of Me, annotated with I/R structures.

2 The Implication/Realization Model

Narmour [7, 8] has proposed a theory of perception and cognition of melodies, the Implication/Realization (I/R) model. According to this model, melodies can be analyzed into melodic structures, that are identified based on principles similar to those of Gestalt theory [4]. Figure 1 shows an I/R analysis of a melodic fragment. The labels of the structures (*P*, and *ID*) denote whether, and in which dimensions (e.g. the direction and size of the melodic interval) the second melodic interval of the structure realizes the implications raised by the first interval. The extent to which structures are *chained* (i.e. share notes), is related to the strength of *closure* at structure boundaries, which is determined by factors like meter, rhythm, and underlying harmony.

3 Experimentation

We compared the behavior of the I/R measure with three more familiar melodic distance measures: an edit-distance for note representations of the melody, one for interval-contour representations (a sequence of numbers representing the number of semitones between each subsequent pair of notes, and one for direction-contour representations (a sequence of +1,0,-1, denoting the sign of the melodic interval between consecutive notes). We used these four distance measures to assess pairwise distances for a number of musical phrases. The comparison was performed using 124 different musical phrases from 40 different jazz songs from the Real Book. The musical phrases have a mean duration of eight bars. Among them are jazz ballads like 'How High the Moon' with around 20 notes, many of them with long duration, and Bebop themes like 'Donna Lee' with around 55 notes of short duration. Jazz standards typically contain some phrases that are close variations of each other (e.g. only different beginning or ending) and some that are more distinct. This is why

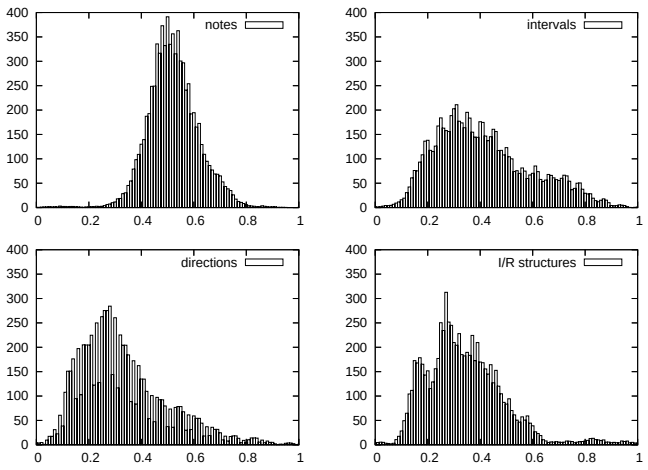

Figure 2. Distribution of distances for four melodic similarity measures. The x axis represents the normalized values for the distances between pairs of phrases. The y axis represents the number of pairs that have the distance shown on the x axis.

the structure of the song is often denoted by a sequence of labels such as A1, A2 and B, where labels with the same letters denote phrases that are similar.

Figure 2 shows the distribution of normalized distance values per measure, for each of the pairwise comparisons of phrases. The first thing to notice from figure 2 is the difference in similarity assessments at the note-level on the one hand, and the interval, direction and I/R-levels on the other hand. Whereas the distance distributions of the last three measures are more spread across the spectrum with several peaks, the note level measure has its values concentrated around one value. This suggests that the note-level measure has a low discriminatory power. We can validate this by computing the Shannon entropy of the distributions as a measure of discriminatory power. The results are shown in the top row of table 1.

We have also investigated the ability of the measures to identify 'close variation' phrases from the same song (e.g. A1 and A2 phrases). Ideally, comparisons of such phrases should yield values in a distinctly lower range than comparisons of other phrases. We assessed this ability by comparing for each measure, the distribution of distances between *variation-phrase* distances with the distribution of *non variation-phrase* distances. The comparison of these distributions was done by computing the Jensen-Shannon divergence (JSD) between them. The ratings are shown in the bottom row of table 1. It turns out that the note and I/R representations perform better at identifying variation phrases.

It might be argued that the interval and direction representations are too abstract to give a good impression of the melodies, since any rhythmical information is absent. The divergence between variation comparisons and non-variation comparisons might be improved by adding durational information in the representation. To check this, we also assessed the melodic distances using these extended interval and direction representations. The entropy and JSD values are shown in the last two columns of table 1. As can be seen, adding durational information indeed improves the ability to distinguish variation comparisons from non variation comparisons, but it comes at the cost of a decrease in discriminative power on the whole data set.

	Note	Interval	Direction	I/R	Interval +duration	Direction +duration
Entropy	4.41	5.27	5.12	4.91	5.04	4.81
JSD(var.-phrase, non var.-phrase)	0.67	0.62	0.58	0.64	0.68	0.69

Table 1. Comparison of the entropy of the distance measures for the data set as a whole (top row), and the Jensen-Shannon divergence between the variation-phrase and non variation-phrase comparisons for each of the measures (bottom row).

4 Conclusions and future work

In this paper we have proposed a new way of assessing melodic similarity and compared it with existing methods for melodic similarity assessment. The discriminatory power (using an entropy based definition) on the whole data set was highest for the (most abstract) contour and I/R level measures and lowest for the note level measure. This suggests that abstract melodic representations serve better to differentiate between phrases that are not near-identical (e.g. phrases belonging to different musical styles) than very concrete representations such as the note representation. The experimentation also showed that the note and I/R level measures were better at clustering phrases from the same song than the contour (i.e. interval and direction) level measures. This was shown to be due to the fact that rhythmical information is missing in the contour level measures. Taking into account this information in the contour level measures improved their ability to separate *variation-phrase* comparisons from *non variation-phrase* comparisons, at the cost of discriminatory power on the whole data set.

The interval measure with additional rhythmical information performed slightly better than the I/R measure. Still, the I/R measure seems to be a good compromise between very concrete and very abstract melodic representations. It incorporates rhythmic information in an implicit way (through *closure*), allowing the measure to separate *variation-phrase* comparisons from *non variation-phrase* comparisons, while maintaining its discriminative power on assessments that involve more diverse musical phrases. Additional tuning of the I/R measure, might even improve its performance.

In the future, we wish to investigate the usefulness of the similarity measures to identify/cluster phrases from the same musical style. Some initial tests indicated that in particular the contour and I/R measures separated bebop style phrases from ballads. Possibly, further categorizations can also be made. However, for definitive conclusions in this direction, more research (with explicitly labeled data) is needed.

References

[1] David Cope, *Computers and Musical Style*, Oxford University Press, 1991.

[2] W. J. Dowling, 'Scale and contour: Two components of a theory of memory for melodies', *Psychological Review*, **85**(4), 341–354, (1978).

[3] D. Hörnel and W. Menzel, 'Learning musical structure and style with neural networks', *Computer Music Journal*, **22 (4)**, 44–62, (1999).

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

[5] Kjell Lemström and Sami Perttu, 'Semex - an efficient music retrieval prototype', in *First International Symposium on Music Information Retrieval (ISMIR'2000)*, Plymouth, Massachusetts, (October 23-25 2000).

[6] M. Mongeau and D. Sankoff, 'Comparison of musical sequences', *Computers and the Humanities*, **24**, 161–175, (1990).

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

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

[9] P.Y. Rolland, 'Discovering patterns in musical sequences', *Journal of New Music Research*, **28 (4)**, 334–350, (1999).