Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney, 2002
C. Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick (Eds.). Adelaide: Causal Productions.

A COMPUTATIONAL MODEL OF MELODIC SIMILARITY BASED ON MULTIPLE REPRESENTATIONS AND SELF-ORGANIZING MAPS

Petri Toiviainen & Tuomas Eerola

University of Jyväskylä, Department of Music

{ptoiviai, ptee}@jyu.fi

ABSTRACT

Perceiving similarity relationships in melodies is a fundamental musical process effortlessly performed by the cognitive system of listeners. For this reason, computationally heavy methods such as string-matching algorithms may not be plausible as perceptually oriented computational models of this process. Instead of direct comparison of musical events, similarity can be thought of as a higher order emergent property that has been abstracted by learning, as a result of which similarity relations are based on more compressed representations, such as prototypes. A computational model of melodic similarity is proposed that is based on multiple representations and feature maps formed by unsupervised learning algorithms (self-organizing maps, SOM). Several commonly used components of melodic similarity, including melodic contour and distributions of tones, intervals and durations, are represented as vectors in multidimensional spaces and similarity of these features as distance in metric space. The SOMs yield clustered projections of these features, thus providing reduced representations in terms of both dimensionality and number of feature vectors. The SOMs can be used as a general tool for examining similarity relationships of the features. The current work makes the following contributions. First, several melodic features are extracted in parallel, each feature being represented in a distributed fashion. This corresponds to current understanding of the functioning of the perceptual system (e.g. multiple representations in the visual domain). Second, the model simulates the learning of melodic schemata through self-organization. Third, the model can be applied to musical data mining. This talk will demonstrate how the proposed computational model of melodic similarity is constructed and how it can be used in musical data mining. Representations of melodic features and similarity relationships generated by the model are discussed in terms of perceptual relevance.

1. BACKGROUND

Empirical studies on the perception of melodic similarity have focused on various aspects. Contour information has been found to be one of the main factors [1]. Besides contour, studies on melodic similarity have focused on melodic archetypes, hierarchical structure, themes, motifs, presence of scalar or non-scalar tones, transposed melodies, and pitch range. Commonly, rhythm has been considered as a separate entity, except in [2], where both rhythm and tonal dimensions are considered as constituents of similarity.

Theoretical models of melodic similarity include Cambouropoulos' [4,5] formal definition of similarity based on the number of coinciding attributes of melodies. In [6], similarity is considered as sets of properties on different hierarchical levels, the properties ranging from pitch-class sets to tempo and dynamic descriptions of atonal music. Discussion about theoretical models of melodic similarity was recently supplemented by [7]. Models that deal with contour and interval information of the melodies include [8-10].

Most models of melodic similarity presented to date are based on computationally heavy serial methods such as string-matching algorithms. This is inconsistent with the general knowledge about the functioning of the perceptual system. In visual domain, for instance, neurobiological studies suggest that the visual system makes use of multiple parallel representations of various features present in the visual signal [11,12]. These features are related to location, colors, contrast, shape, lines, orientation, motion, binocular disparity etc., and are mapped onto specific locations on the visual cortex. These maps are processed in a parallel fashion and simultaneously, but jointly represent the stimulus being observed and comprise the initial cognitive map of the visual environment.

Cross-cultural studies on melodic expectancy [13,14] have found that expectations on melodic continuation are to some extent culture-dependent. One could expect that the same applies to melodic similarity as well. A plausible model of melodic similarity should therefore account for learning. In other words, similarity could be thought of as a higher order emergent property that has been abstracted by learning, as a result of which similarity relations are based on more compressed representations, such as prototypes. It is commonly believed that such protopes and representations of their interrelationships emerge by self-organization [15].

Research on music cognition and learning has demonstrated the effect of statistical information for learning and perception by means of both cross-cultural studies [13,14,16,17] and studies using melodies in which the statistical properties of music have been intentionally manipulated [18]. Furthermore, statistical features have been found to predict perceived similarity of melodies [19]. In light of this evidence, it seems that statistical properties of melodies could provide a basis for classification of musical styles in terms of their perceptual similarity. This approach has in fact been used in [20-23]. It is worth mentioning that the predecessors of these methods in music were conceived in ethnomusicology, where musical styles were classified according to the statistical distribution of different intervals, rhythmic patterns, or pitches [24,25].

2. AIMS

We propose a computational model of melodic similarity that is based on parallel representations of multiple musical features and maps formed from these features by an unsupervised learning algorithm (self-organizing map, SOM, [15]). The SOMs yield clustered projections of these features, thus providing reduced representations in terms of both dimensionality and number of feature vectors. These SOMs are fed to a higher hierarchical level that represents the similarity relations of the whole set of melodies. The configuration of the SOMs depends on the distribution of the features in the set of melodies used for training. Therefore, the model can be used in simulating effects of musico-cultural background on similarity perception.

3. MAIN CONTRIBUTION

The first stage of the model consists of the extraction of a number of musical features from a corpus of melodies that is represented symbolically (e.g., in MIDI format). These features include (1) melodic contour, (2) statistical features of pitch structure, including the distribution of pitch-classes, intervals, pitch transitions, and interval transitions; and (3) statistical features of rhythmic structure, including the distribution of note durations and note duration transitions. Each of these features is represented as a vector in a multidimensional metric space and the degree of similarity of these features as distance in this space. It is assumed that all melodies are transposed to a common key before the statistical features are extracted from each melody separately. Figure 1 displays an example of features extracted from a melodic phrase.

To obtain feature maps, each set of features is used for training a SOM. The SOM is an artificial neural network that simulates the process of self-organization in the central nervous system with a simple, yet effective, numerical algorithm. It consists of a two-dimensional planar array of simple processing units, each of which is associated with a reference vector. The dimensionality of these reference vectors is equal to that of the vectors used as input. During the learning session, the SOM performs clustering, that is, represents the training set by a fewer number of prototypes, and forms a non-linear projection that optimally approximates the distribution of the data in which similar vectors are mapped near each other. The non-linearity of the mapping enables the SOM to identify the most salient dimensions of the input set. Furthermore, the mapping has a variable magnification so that regions of input space with a high density of input vectors are represented with a higher resolution than other regions, a characteristic that models the dependence of the accuracy of mental representations on the environmental exposure. Figure 2 depicts schematically the principles of the mapping provided by the SOM.

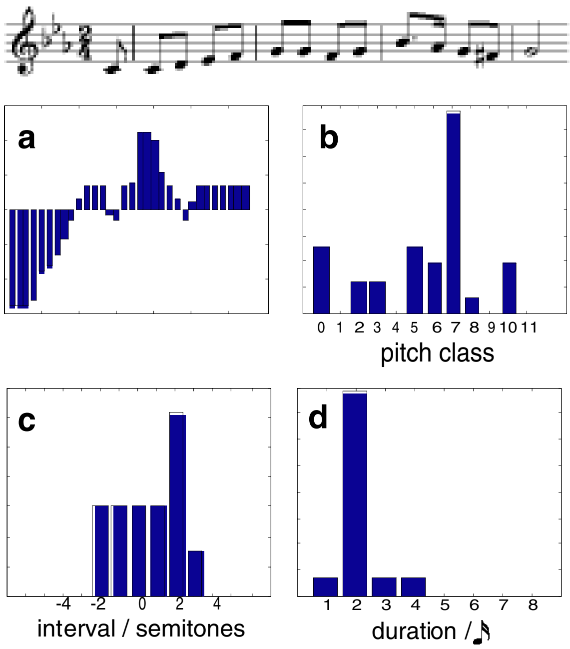

Fig. 1. Notation of a melodic phrase and its (a) contour representation, (b) pitch class distribution, (c) interval distribution, and (d) note duration distribution.

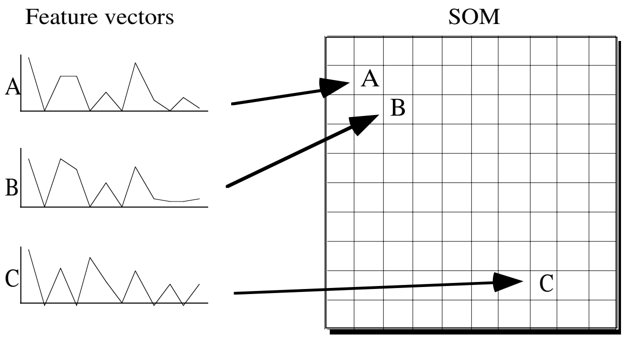

Fig. 2. A schematic presentation of the self-organizing map. Similar vectors, such as A and B, are mapped to mutually adjacent units. Dissimilar vectors, such as B and C, are mapped to mutually distant units.

The outputs of the feature maps are used as input to a super-SOM to obtain a global representation of similarity relationship. More specifically, for each melody, the coordinates of the unit where the melody is mapped on each SOM are concatenated and subsequently used as the input vector for the super-SOM. This yields a two-dimensional map on which melodies with similar features are proximally located. In other words, melodies that display similar properties in terms of contour structure as well as distributions of pitch, interval and duration structure are located at adjacent positions on this map. The structure of the model is schematically depicted in Figure 3.

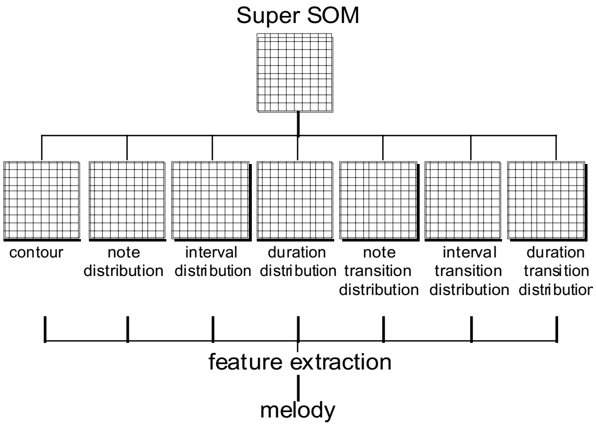

Fig. 3. A schematic representation of the model.

4. IMPLICATIONS

The proposed model aims at perceptual plausibility by adopting multiple parallel representations that are processed simultaneously. These representations are formed through self-organization, and can be regarded as schemata representing typical melodic features abstracted from the musical environment. This property of the model enables the simulation of the development of culture-specific differences in the perception of melodic similarity. Besides such simulations, the model can be used as a general tool for examining similarity of features in corpora of melodies. Another potential application is data mining for hypothesis formation in comparative ethnomusicology. The model has been applied to a corpus of melodies that consists of about 8,500 folk songs taken mainly from the Essen collection [26]. The obtained feature maps can be examined on the Internet at www.jyu.fi/~musica/essen.

5. DISCUSSION

The proposed model uses a distributed representation of multiple features to represent melodic similarity, which corresponds to current understanding of the functioning of the perceptual system. Furthermore, it simulates the learning of melodic schemata through self-organization. The similarity relationships between melodies are represented by the metric similarities between the schemata that correspond to the particular melodies. Therefore, the model offers a method for simulating cross-cultural differences in the perception of melodic similarity.

The selection of features used in the model to represent the melodies is not meant to be exhaustive: it is probable that a more extensive set of features is used by the perceptual system to represent melodic structure. To assess the validity of the proposed model, it should be tested against empirical data. This would include, for instance, the comparison of similarity judgments against the mapping distances on the SOMs. This comparison could shed light on the relative importance of each feature in similarity formation among listeners in a given musical culture. Moreover, intercultural differences in similarity ratings could be simulated by maps trained with collections consisting of various mixtures of extra- and intracultural musical material.

6. REFERENCES

1. Dowling, J. W. "Recognition of inversions of melodies and melodic contours", *Perception & Psychophysics,* 9, 348-349, 1971.
2. Dowling, W. J. "Scale and contour: two components of a theory of memory for melodies", *Psychological Review,* 85, 341-354, 1978.
3. Monahan, C. B., & Carterette, E. C. "Pitch and duration as determinants of musical space", *Music Perception,* 3, 1-32, 1985.
4. Cambouropoulos, E. & Smaill A. "A computational model for the Discovery of Parallel Melodic Passages", *Proc. of the 11 Colloquio di Informatica Musicale,* Bologna, 1995.
5. Cambouropoulos, E. "Melodic cue abstraction, similarity and category formation: A computational approach", *Music Perception,* 18, 347-370, 2001.
6. Anagnostopoulou, C. & Smaill, A. "Similarity and categorisation in Boulez' Parenthese from the Third Piano Sonata: A formal analysis", *Proc. of the 6th ICMPC,* Keele, 2001.
7. Hewlett, W. B. & Selfridge-Field. E. (Eds.), *Melodic similarity: Concepts, procedures, and applications,* Cambridge, Massachusetts: MIT Press, 1998.
8. Deutsch, D. & Feroe, J. "The internal representation of pitch sequences in tonal music", *Psychological Review,* 88, 503-522, 1981.
9. Ó Maidín, D. "A geometrical algorithm for melodic difference", in W. B. Hewlett & E. Selfridge-Field (Eds.). *Melodic similarity: concepts, procedures, and applications,* Cambridge, Massachusetts: MIT Press, 65-72, 1998.
10. Hofmann-Engl L. & Parncutt R. "Computational modeling of melodic similarity judgments: Two experiments on isochronous melodic fragments", http://freespace.virgin.net/ludger.hofmann-engl/similarity.html, 1998.
11. Marr, D., *Vision,* New York, W. H. Freeman, 1982.
12. Zeki, S., *A vision of the brain,* London: Blackwell, 1993.
13. Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. "Melodic expectation in Finnish folk hymns: Convergence of statistical,

ISBN 1 876346 39 6

behavioral, and computational approaches", *Music Perception,* 17(2), 151-196, 1999.

14. Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T. & Louhivuori, J. "Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks", *Cognition,* 76, 13-58, 2000.

15. Kohonen, T., *Self-organizing maps (2nd ed.),* Berlin: Springer, 1997.

16. Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. "Tonal hierarchies in the music of North India", *Journal of Experimental Psychology: General,* 113, 394-412, 1984.

17. Kessler, E. J., Hansen, C., & Shepard, R. N. "Tonal schemata in the perception of music in Bali and the west", *Music Perception,* 2, 131-65, 1984.

18. Oram, N., & Cuddy, L. L. "Responsiveness of Western adults to pitch-distributional information in melodic sequences", *Psychological Research,* 57, 103-118, 1995.

19. Eerola, T., Järvinen, T., Louhivuori, J. & Toiviainen, P. "Statistical features and perceived similarity of folk melodies", *Music Perception,* 18, 275-296, 2001.

20. Järvinen, T., Toiviainen, P., & Louhivuori, J. "Classification and categorization of musical styles with statistical analysis and self-organizing maps", in A. Patrizio, G. A. Wiggins & H. Pain (Eds.), *Proc. of the AISB'99 Symposium on Musical Creativity,* Edinburgh: The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 54-57, 1999.

21. Smaill, A., & Westhead, M. D. "Automatic characterisation of musical style", *Music Education: An artificial intelligence approach,* Berlin: Springer-Verlag, 157-170, 1993.

22. Atalay, B. V., & Placek, R. "Machine versus human: Responding to the task of identifying eras for keyboard pieces", in A. Gabrielsson (Ed.), *Proc. of the Third Triennial ESCOM conference,* Uppsala: Uppsala University, 521-526, 1997.

23. Hörnel, D. "A multi-scale neural-network model for learning and reproducing choral variations", in W. B. Hewlett & E. Selfridge-Field (Eds.), *Melodic Similarity: Concepts, Procedures, and Applications, Cambridge*, Massachusetts: MIT Press, 141-157, 1998.

24. Freeman, L. C. and Merriam, A. P. "Statistical classification in anthropology: An application to ethnomusicology", *American Anthropologist,* 58, 464-472, 1956.

25. Lomax, A., *Folk song style and culture,* Washington, D.C.: American Association for the Advancement of Science, 1968.

26. Schaffrath, H., *The Essen Folksong Collection in Kern Format,* [computer database], D. Huron (Ed.). Menlo Park, CA: Center for Computer Assisted Research in the Humanities, 1995.