A METHOD FOR COMPARATIVE ANALYSIS OF FOLK MUSIC BASED ON MUSICAL FEATURE EXTRACTION AND NEURAL NETWORKS

Petri Toiviainen & Tuomas Eerola
University of Jyväskylä
Finland

Introduction

A common problem in comparative musicology and ethnomusicology is that large collections of music are difficult to classify and visualize. Therefore, a tool which could be applied to either acoustic signals or symbolic representations would be useful. The choice of the features that are extracted from a collection of music and subsequently used by the tool should be psychologically relevant for the task. This study presents a simple data-mining tool for data-bases that use a symbolic representation of melodic information. The statistical distributions of melodic events are considered as a suitable features for several reasons. Firstly, the distributions are relatively straightforward to analyze computationally. Secondly, it has been shown that listeners are sensitive to pitch distributional information (Kessler et al. 1984; Oram & Cuddy, 1995; Krumhansl et al, 1999) and they can be used to predict similarity relationships between melodies (Eerola et al, 2001). It is also noteworthy that ethno-musicology has a long tradition in using statistical information to classify music (Freeman and Merriam, 1956; Lomax, 1968) and that there has been more recent attempts to classify musical styles according to their statistical features (Järvinen, Toiviainen, & Louhivuori, 1999).

Feature extraction and visualization (SOM)

The method is based on first extracting the common statistical measures of music. These consist of the distributions of pitches, intervals and durations as well as the distributions of pitch, interval, and duration transitions (Figure 1). It is assumed that all melodies are transposed to a common key before the statistical features are extracted from each melody separately.

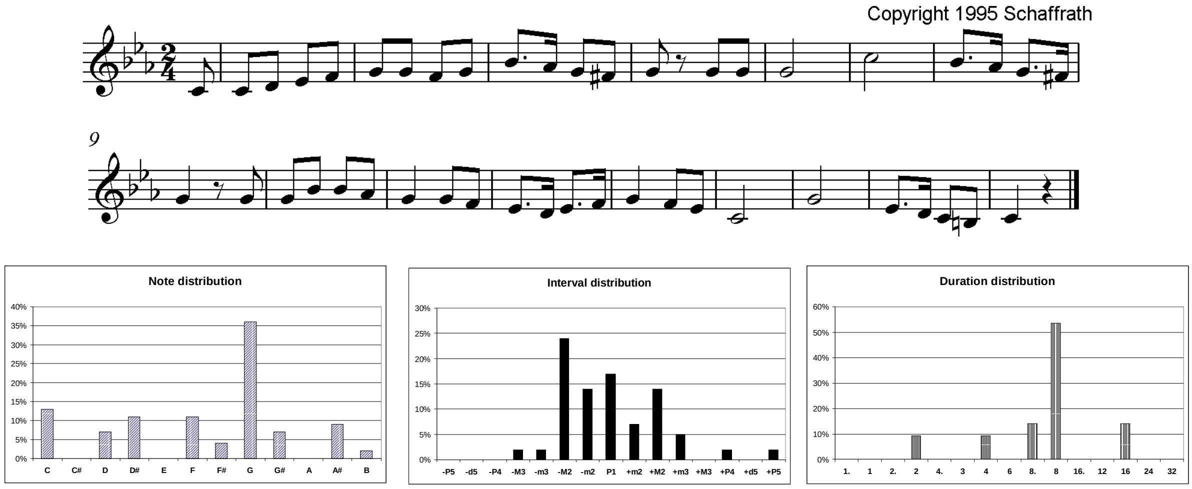

Figure 1. The distribution of pitches, intervals, and durations extracted from the melody *Och riddaren han gångar sig till hafsjöstranden ned*.

The representations of the musical materials obtained with the statistical analysis constitute a set of vectors with a large number of components. Because of the high dimensionality of the data, the mutual relations of the data items can be difficult to determine. Therefore, we used the self-organizing map (SOM) for the visualization of these mutual relations (Kohonen, 1997). 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. After being trained with the input vectors, the SOM provides a non-linear topographic mapping from the multidimensional input space to the two-dimensional array. In other words, each input vector is mapped to some unit in the array, and vectors that are close to each other in the input space are mapped near each other. In addition, the SOM identifies the most salient features of the input set by detecting in each part of the input vector distribution the dimensions with the highest variance. Figure 2 depicts schematically the principles of the mapping provided by the SOM.

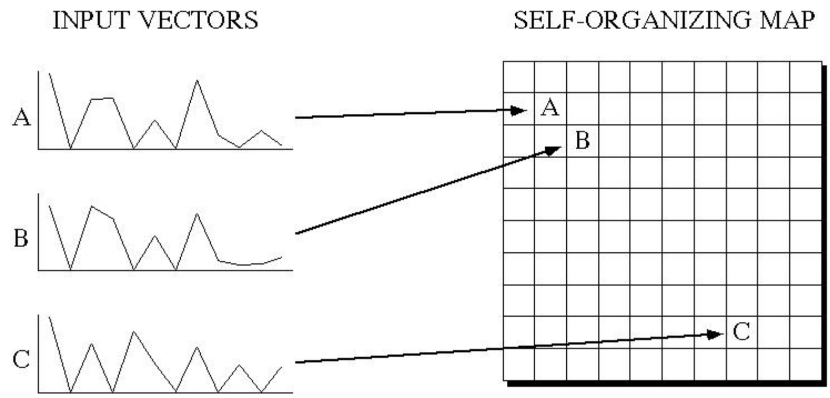

Figure 2. A schematic presentation of the self-organizing map. Similar vectors, such as A and B, are mapped near each other. Dissimilar vectors, such as B and C, are mapped far away from each other.

After the extraction of statistical features, each feature is for the trained of a SOM. The maps thus obtained can be used separately for visualization. Furthermore, a Super SOM can be trained with the vectors consisting of the outputs of these SOMs (see Figure 3). This yields a two-dimensional Supermap on which melodies with similar features are proximally located. In other words, melodies that display similar statistical properties in terms of pitch, interval and duration distributions and their transitions are located at adjacent positions on the Supermap. Increased perceptual validity is obtained if a weighting scheme corresponding to empirical findings of the importance of each feature for listeners' similarity formations is used in the teaching of the Super SOM.

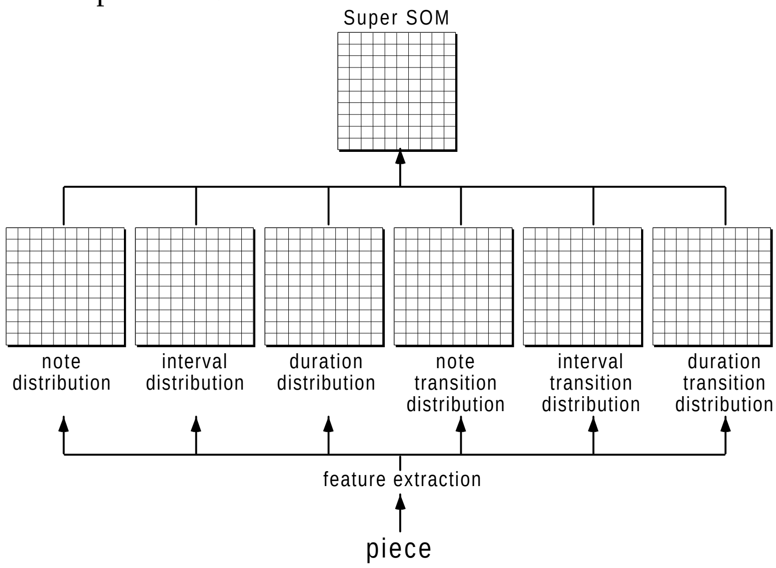

Figure 3. Overview of the method.

The method applied to a large folk song collection

The method was applied to a corpus of melodies that consists of 6,252 folk songs from the Essen collection (Schaffrath, 1995) and 2,226 Chinese folk songs. The songs in the Essen collection are mainly from Germanic regions and the Chinese songs are from the northern part and border region of Ningxia and Shanxi. One of the advantages of using this particular corpus was that all songs are encoded in symbolic format (*Humdrum **kern* format, Huron, 1994) and an electronic version of the database is published and distributed by the Center for Computer Assisted Research in the Humanities (CCARH). Another benefit of this corpus is that all transcriptions include the definition of the genre, geographical region, rhythm type, key, and a free description of the content and context in the form keywords. This additional information can be used as search criteria in the visualization tool and thus extend the utility of the corpus.

The demonstration of the method is divided into three tools. Tool 1 provides a coarse overview of features by displaying the organization of each map together with the entropy of each feature. Entropy is a measure of complexity that has been used previously in discriminating musical styles (Knopoff, & Hutchinson, 1983; Snyder, 1990). This tool shows the songs with similar features in proximate areas and can thus be used to investigate the similarity relationships between the songs. It also enables the playback of any chosen song on each SOM. A demonstration of this tool is available on the WWW (*www.jyu.fi/musica/essen*). Tool 2 provides a visualization of the statistical features as represented by the SOMs. Tool 3 combines keyword search with the similarity relations of the features. This tool can be used to find stylistic clusters or specific locations of the songs containing any selected criteria such as "ballads", 3/4 time-signatures, "Tirol" or any combination of these. This facilitates formulating and answering musically and culturally interesting questions from the corpus.

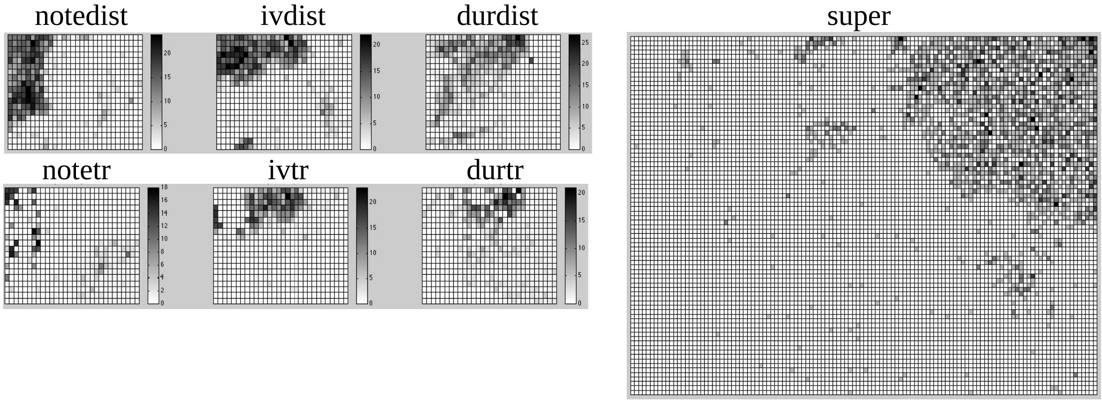

Figure 3. An example of the Tool 3 displaying the mapping locations of the melodies of the Han (A Chinese ethnic group).

Conclusions and future directions

A method for the analysis of large corpus of music and specific practical tools for musical data mining were presented. The method was based on the statistical distribution of symbolic events and subsequent investigation of similarity relationships. Self-organizing neural network (SOM) was used to visualize the feature vectors. Examples included keyword-based investigation of musical features as well as separate topological maps of all the songs for each extracted feature. One possible application of the method is to use it to find stylistic disparities or similarities between materials from distant cultural regions and employ this information when creating hypothesis for cross-cultural comparisons. However, there is currently a lot of room for the improvement of the method itself. For example, taking into account the overall melodic contour, hierarchical reduction of the melodic surface, perceptual weighting of the events according to the metrical position and salience and phrasing would provide more sophistication and increase the perceptual relevance of the method. Further research would be needed to assess the applicability of the present method to audio-based material.

References

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

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

Huron, D. (1994). *UNIX tools for musical research: The humdrum toolkit reference manual*. Stanford, CA: Center for Computer Assisted Research in Humanities.

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

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

Knopoff, L. & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. *Journal of Music Theory*, 27, 75-97.

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

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish folk hymns: Convergence of statistical, behavioral, and computational approaches. *Music Perception, 17(2)*, 151-196.

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

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

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

Snyder, J. L. (1990). Entropy as a measure of musical style: The influence of a priori assumptions. *Music Theory Spectrum, 12*, 121-160.