

Applying the Independent Component Analysis for the estimation of an upper bound of the expressiveness of basic audio features commonly used in algorithms for computing music similarity

Research Proposal

Study Problem

Many different methods for automatic retrieval of music similarity have been proposed. There are a number of low-level descriptors that are widely used as a basis for comparison algorithms. Although the different authors usually have different sets of songs to evaluate the algorithms and there usually is no exhaustive user evaluation, which makes the comparison between the different algorithms difficult, it seems that the expressiveness of the algorithms is somehow limited. The question that arises is, if this upper bound exists, or if the set of basic descriptors are sufficient to describe music exhaustively for computing similarity between different pieces of music.

Relevance Statement

If it comes out that the basic descriptors are sufficiently expressive, the research can focus on optimally combining them. If not, more research is to be done to find useful low-level descriptors, and it even might be that music cannot be described sufficiently by only looking at the audio content.

Literature Review

[1] gives a list of low-level audio descriptors commonly used in MIR. Among them are Mel-Frequency Cepstral Coefficients (MFCCs), which describe the envelope (shape) of the frequency spectrum. For example, MFCCs are used as a first step in [2] and [3] before clustering and mapping to classes is done respectively. Other low-level features include several loudness measures and flatness measures.

[4] and [5] target the problem of comparing different music similarity measures; in [6] the authors state that an objective comparison is difficult because there exists no large-scale ground truth to evaluate the different measures on.

Method

Given a set of music pieces and a set of low-level descriptors, for each piece separately the low-level descriptors are determined, so each song is represented as a sequence (vector) of feature data. The goal is to assess how useful this whole set of feature data can maximally be in describing the songs with regard to their similarity. So as a criterion, information about the musical similarity between each pair of songs in the set of given songs is needed. (This information can be approximated by supposing that songs by the same artist, on the same album, or in the same genre are more similar than songs that do not have one of these things in common; e.g. [2])

The crucial step is to find an optimal combination / weighting of the features; when this is achieved, it can be attempted to calculate the given similarities with the weighted combination of features. The goodness of the result reflects the quality of the features. (The optimisation process is done on the same set of data as the final evaluation, so that features that are optimally capable of describing the music would exactly reach the optimisation criterion.)

When searching for the optimal combination of features, it is important to note that the low level features are not necessarily independent (e.g. it is not obvious if the various methods to describe the

power and shape of spectrum or subband are independent, as they describe related properties; also the type of audio material might introduce dependencies). This problem is getting worse when the number of used features increases – and it is most interesting to use as much features as possible. As a solution, the Independent Component Analysis (ICA, e.g. [6]) might be applied to the (uniform-scaled) feature vectors of all songs in the given set, allowing the construction of independent “meta features“. Each of these “meta features“ can be tested separately against the given criterion. Afterwards they can be linear combined according to their significance, as they are independent.

A drawback of the proposed method is that the used low-level features have to be calculated on a data basis that itself is not dependant on the data of the song (e.g. when clustering a song and then taking mean, weight and covariance of each cluster as feature data for the song). As a solution, the basis for the data basis can be calculated for all songs (e.g. clustering all songs together, and thus fixing the cluster centroids that are allowed).

References

- [1] **A multiple feature model for musical similarity retrieval**, *Allamanche, E., Herre, J., Hellmuth, O., Kastner, T., and Ertel, C.*, Fourth International Conference on Music Information Retrieval (ISMIR 2003), October 2003
- [2] **A Music Similarity Function Based on Signal Analysis**, *Beth Logan and Ariel Salomon*, IEEE International Conference on Multimedia and Expo (ICME), August 2001
- [3] **Anchor space for classification and similarity measurement of music**, *A. Berenzweig, D.P.W. Ellis, and S. Lawrence*, in Proc IEEE Intl Conf on Multimedia and Expo (ICME), Baltimore, MD, 2003.
- [4] **A Large-Scale Evaluation of Acoustic and Subjective Music Similarity Measures**, *Adam Berenzweig, Beth Logan, Daniel P. W. Ellis and Brian Whitman*, Fourth International Conference on Music Information Retrieval (ISMIR 2003), October 2003
- [5] **On the Evaluation of Perceptual Similarity Measures for Music**, *Elias Pampalk, Simon Dixon, Gerhard Widmer*, 6th Int. Conference on Digital Audio Effects (DAFx-03), London, UK, September 2003
- [6] **Redundancy Reduction for Computational Audition, a Unifying Approach**. *Paris Smaragdis*, Massachusetts Institute of Technology, Media Laboratory. Dissertation, May 2001