Music Information Retrieval

Alicja A. Wieczorkowska
Polish-Japanese Institute of Information Technology, Poland

INTRODUCTION

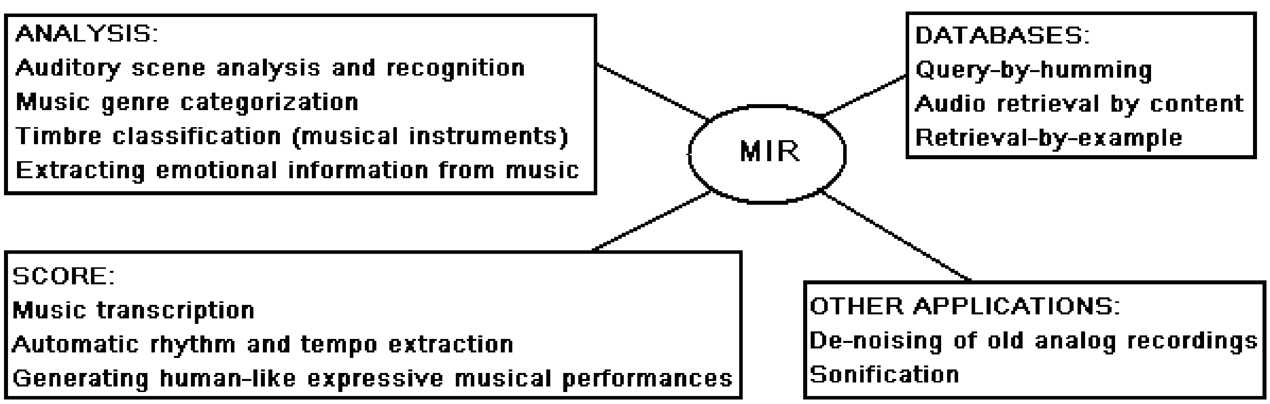
Music information retrieval (MIR) is a multi-disciplinary research on retrieving information from music, see Fig. 1. This research involves scientists from traditional, music and digital libraries, information science, computer science, law, business, engineering, musicology, cognitive psychology and education (Downie, 2001).

BACKGROUND

Huge amount of audio resources, including music data, is becoming available in various forms, both analog and digital. Notes, CDs, and digital resources of the World Wide Web are constantly growing in amount, but the value of music information depends on how easy it can be found, retrieved, accessed, filtered and managed. MIR consists in quick and efficient searching for various types of audio data of interest to the user, and filtering them in order to receive only the data items which satisfy the user's preferences (Fingerhut, 1997), (International Organization for Standardization, 2003), (Wieczorkowska & Ras, 2003). MIR applications include retrieval of music pieces from huge audio databases, and automatic production of music score on the basis of the presented input. The MIR topics are interrelated, since similar techniques can be applied for various purposes, e.g., source separation is applied in auditory scene analysis, music transcription, and even for restoring (de-noising) of all recordings. Generally, the research within MIR domain is focused on: harmonic structure analysis, note extraction, melody and rhythm tracking, timbre and instrument recognition, classification of type of the signal (speech, music, pitched vs. non-pitched), etc.

The research basically uses digital audio recordings, where sound waveform is digitally stored as a sequence of discrete samples representing the sound intensity at given time instant, and MIDI files, storing information on parameters of electronically synthesized sounds (voice, note on, note off, pitch bend etc.). Sound analysis and data mining tools are used to extract information from music files, in order to provide the data that meet user's needs (Wieczorkowska & Ras, 2001).

Figure 1. Topics of MIR



MAIN THRUST OF THE CHAPTER

Various types of music data are investigated in MIR research. Basic techniques of digital sound analysis come from speech processing, focused on automatic speech recognition and speaker identification (Foote, 1999). The obtained sound descriptors can be applied to facilitate content-based searching of music databases.

The issue of representation of music and multimedia information in a form that allows interpretation of the information's meaning is addressed by MPEG-7 standard, named Multimedia Content Description Interface. MPEG-7 provides a rich set of standardized tools to describe multimedia content through metadata, i.e. data about data, and music information description has been also taken into account in this standard (International Organization for Standardization, 2003).

The following topics are investigated within MIR domain:

- Auditory scene analysis and recognition (Rosenthal & Okuno, 1998), which focuses on various aspects of music, like timbre description, sound harmonicity, spatial origin, source separation, etc. (Bregman, 1990), based on signal processing techniques. Timbre is defined subjectively as this feature of sound that distinguishes two sounds of the same pitch, loudness and duration. Therefore, subjective listening tests are often performed in this research. One of the main topics of computational auditory scene analysis is automatic separation of individual sound sources from a mixture. It is difficult with mixtures of harmonic instrument sounds, where spectra overlap. However, assuming time-frequency smoothness of the signal, sound separation can be performed, and when sound changes in time are observed, onset, offset, amplitude and frequency modulation have similar shapes for all frequencies in the spectrum, thus a de-mixing matrix can be estimated for them (Virtanen, 2003), (Viste & Evangelista, 2003). Audio source separation techniques can also be used to source localization for auditory scene analysis. These techniques, like independent component analysis (ICA), originate from speech recognition in cocktail party environment, where many sound sources are present. ICA is used for finding underlying components from multidimensional statistical data, and it looks for components that are statistically independent (Vincent, Rodet, Röbel, Févotte, Carpentier, Gribonval, Benaroya, & Bimbot, 2003). Also, computational auditory scene recognition may aim at classifying auditory scenes into predefined classes, using audio information only. Examples of auditory scenes are various outside and inside environments, like streets, restaurants, offices, homes, cars etc. Statistical and nearest neighbor algorithms can be applied for this purpose. In the nearest neighbor algorithm the class (type of auditory scene in this case) is assigned on the basis of the distance of the investigated sample to the nearest sample, for which the class membership is known. Various acoustic features, based on Fourier spectral analysis (i.e. mathematic transform, decomposing the signal into frequency components), can be applied to parameterize the auditory scene for classification purposes. Effectiveness of this research approaches 70% correctness for about 20 auditory scenes (Peltonen, Tuomi, Klapuri, Huopaniemi, & Sorsa, 2002).
- Music genre categorization is aimed at automatic classification of music into various genres. This can be especially useful for large audio collections, if they are not manually labelled (Guaus & Herrera, 2006).
- Audio retrieval-by-example for orchestral music aims at searching for acoustic similarity in an audio collection, based on analysis of the audio signal. Given an example audio document, other documents in a collection can be ranked by similarity on the basis of long-term structure, specifically the variation of soft and louder passages, determined from envelope of audio energy versus time in one or more frequency bands (Foote, 2000). This research is a branch of audio retrieval by content. Audio query-by-example search can be also performed within a single document, when searching for sounds similar to the selected sound event. Such a system for content-based audio retrieval can be based on a self-organizing feature map, i.e. a special kind of a neural network, designed by analogy with a simplified model of the neural connections in the brain, and trained to find relationships in the data. Perceptual similarity can be assessed on the basis of spectral evolution, in order to find sounds of similar timbre (Spevak

Related Content