

STATEMENT OF RESEARCH

George Tzanetakis

Music is a universal activity shared among every human culture on this planet. From the early days of humanity significant amounts of time and effort are spent playing and listening to organized collections of sounds produced by an extraordinary variety of objects and methods of sound production. For most people, music seems to be able to evoke emotions, thoughts, images, impressions and more generally our internal states as we think and experience the world. The invention of the phonograph, made the recording and preservation of arbitrary sounds possible and created the music industry of today. Every month about 4000 compact discs (CDs) are created in Western countries and an estimated 4 million of tracks have been recorded. Music is pervasive in all human activities and is heard everywhere from grocery stores to exercise rooms and from movie theaters to concert halls.

Current advances in compression technology, processor speed, storage capacity, and network bandwidth have made possible the creation of large collections of music. A large part of current internet traffic consists of audio data and it is likely that in the near future all of recorded music will be available digitally. In order to fully take advantage of this great possibility it is necessary to provide tools to organize, structure and explore these vast amounts of musical data. Unlike current audio software tools which are agnostic to audio content and treat it as a monolithic block of digital samples, new software tools must be able to automatically extract content information and have some “understanding” of the music. Examples of such “understanding” are: locating the saxophone solo in a jazz recording, finding the refrain of a song, recognizing the genre and singer of a particular song and searching for similar songs in a database.

My main research focus is the creation, development and evaluation of algorithms and tools that extract information from complex audio signals such as music and the design of novel user interfaces that utilize the extracted information to assist and enhance the browsing and retrieval of audio signals and collections. This work falls under the general research area of Computer Audition which is a term used to describe any audio information extraction process (similarly to Computer Vision). More specifically it belongs to the growing area of Music Information Retrieval (MIR) which deals specifically with the retrieval and analysis of musical signals. My research is inherently interdisciplinary and draws inspiration from the areas of Signal Processing, Machine Learning, Music Cognition, Information Visualization and Human Computer Interaction. Because of the complexity of music, automatic music analysis provides an excellent testbed for the creation, development, use and evaluation of techniques that are potentially useful in all these different areas. Although my main focus has been musical signals, the techniques I have developed are also applicable to other types of audio signals as well as multidimensional time series data in general. Some examples that I have worked on are: sound effects, isolated musical instrument tones and motion capture data. In addition to Music Information Retrieval (MIR) large collections of audio signals are utilized in many research areas such as Auditory Display, Bioacoustics, Computer Music, Music Cognition, Psychoacoustics, and Virtual Reality research.

In my research, I have worked on most aspects and stages of designing and building software systems for manipulation, analysis and retrieval of audio signals from large collections. Specific contributions of my thesis include: extraction of timbral texture features directly from mp3 compressed data, beat analysis using the Discrete Wavelet Transform, and automatic musical genre classification and query-by-example content-based retrieval based on features that characterize the spectral, pitch, and rhythmic content of music. Based on these information extraction techniques I have proposed and developed novel content and context aware user interfaces for editing and browsing audio signals and collections.

In my work I have tried to follow several general guidelines that differentiate it from the majority of existing Computer Audition research. These are: the combination and integration of multiple different audio analysis techniques, the use of Statistics and Machine Learning techniques in addition to Signal Processing methods, the use of expressive 2D and 3D graphical user interfaces that enable the user to be part of the system, the emphasis on working with large collections of real-world complex audio signals rather than individual synthetic “toy” examples, and the careful evaluation of the proposed techniques using statistical methods such as cross-validation as well as user studies and experiments.

The developed techniques and interfaces are all integrated under MARSYAS, a free software framework for rapid prototyping of computer audition research applications developed as part of my research. In addition MARSYAS, supports many of the existing published techniques and facilitates the design and integration of new ones. MARSYAS follows a client-server architecture, in which the client, written in JAVA, is the graphical user interface and the server, written in C++, contains all the numerically intensive Signal Processing and Machine Learning code. The framework has been downloaded by more than 1700 different hosts and has been used for a variety of different projects.

In general I am interested in content-based analysis of multimedia signals and user interface design for retrieval and browsing of multimedia data. As a summer intern, I have worked on the development of user interfaces and data structures for video browsing at SRI International, and designed a robust fast content-based audio fingerprinting algorithm for Moodlogic Inc. Currently, I am working as a PostDoctoral Fellow at the Computer Science Department of Carnegie Mellon University working on a variety of projects mostly related to music information retrieval. Some examples are: the design and evaluation of query-by-humming systems, automatically generated continuous audio feedback music browsing interfaces, and adding support for content-based music search to peer-to-peer networks.

There are many exciting directions for future computer audition research that I would like to explore and in some cases have already started doing so. Some examples are: instrument tracking in complex recordings, classification and segmentation of non-western music, clustering and classification of bioacoustic signals (for example bird or whale songs), singer identification, instrumentation recognition, audio thumbnailing, music generating query-interfaces for similarity retrieval, and clustering of sound effects. In addition to audio signals it is my belief that the techniques described in my thesis can potentially be applied to other complex time-varying signals such as video, DNA, biological monitor signals, motion capture data and weather data. For my future research I would like to explore these possible connections and I am looking forward to collaborate with researchers in these areas.

To summarize, my broad research goal is the creation of digital libraries of music that support various ways of interaction that take advantage of automatic music “understanding”. All of my work can be viewed as building the foundations for achieving this goal and I hope to continue working towards achieving it.