

Auditory Group Theory with Applications to Statistical Basis Methods for Structured Audio

Michael Anthony Casey

Submitted to the Program in Media Arts and Sciences, School of Architecture and
Planning on January 9, 1998, in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy.

Abstract

To date there have been no audio signal representation methods capable of characterizing the everyday sounds that are used for sound effects in film, TV, video games and virtual environments. Examples of these sounds are footsteps, hammering, smashing and spilling. These environmental sounds are generally much harder to characterize than speech and music sounds because they often comprise multiple noisy and textured components, as well as higher-order structural components such as iterations and scatterings. In this thesis we present new methods for approaching the problem of automatically characterizing and extracting features from sound recordings for re-purposing and control in structured media applications.

We first present a novel method for representing sound structures called auditory group theory. Based on the theory of local Lie groups, auditory group theory defines symmetry-preserving transforms that produce alterations of independent features within a sound. By analysis of invariance properties in a range of acoustical systems we propose a set of time-frequency transforms that model underlying physical properties of sound objects such as material, size and shape.

In order to extract features from recorded sounds we have developed new statistical techniques based on independent component analysis (ICA). Using a contrast function defined on cumulant expansions up to fourth order, the ICA transform generates an orthogonal rotation of the basis of a time-frequency distribution; the resulting basis components are as statistically independent as possible. The bases are used in conjunction with auditory group transforms to characterize the structure in sound effect recordings. These characteristic structures are used to specify new sounds with predictable, novel features.

For our results we have implemented auditory group models that are capable of synthesizing multiple sound behaviors from a small set of features. These models characterize event structures such as impacts, bounces, smashes and scraping as well as physical object properties such as material, size and shape. In addition to applications in video and film media, the methods presented herein are directly applicable to the problem of generating real-time sound effects in new media settings such as virtual environments and interactive games, as well as creating new sound synthesis methods for electronic music production and interactive music experiences.

Thesis Advisor: Professor Barry L. Vercoe
Professor of Media Arts and Sciences

This work was performed at the MIT Media Laboratory. Support for this work was provided in part by MERL - A Mitsubishi Electric Research Laboratory.

