Master Thesis: Query for Sound

Author: Eric Lehmann

>

Introduction {15}

Working with audio in post production {2}

This work shall investigate current practices in audio post-production. We focus on the special case of searching for audio clips in sound design and similar tasks.

The soundtrack of a movie is a complex collage of multiple layer of audio, each consiting of layers itself, containing several thousands of single audio files arranged, manipulated and mixed to fit into the global "view". [1]

- Recording Production Sound, Foley Artists, Music Effects etc.
- Synthesizing Generating formerly unknown Sounds, Sounds for Layering, creating patches and virtual Instruments - related to Music Effects.
- Collecting / Library Usage Bridge to main topic: Workig with existing soundclips, as recording and synthesizing are very expensive in time and money. Aim is to find appropriate audio material in the least possible time:
  - How do you organize a huge amount of audio clips?
  - How are the Libraries sorted?
  - Short overview of heterogeniousity in reality.

Metadata Formats {4}

Introdution only of relevant Data Formats (BWAV, AIFF, mp3, ogg, Flac, aac, iXML, aXML)

Accessible Fields {4}

Overwiew of details about the usage in two or three formats (description, mic channel, author etc.)

Functional Overview of Sound Library Software {5}

List of functionalities in available tools at the market that provide searching functionality specialized on audio clips. (Soundminer, Basehead, Netmix etc.) (* cited from creativefieldrecording.com *)

Searching methods for audio clips {20}

The basic query methods

Introduction to Music Information Retrieval / Audio Content Analysis {4}

Text search {3}

- plain search {1}
- boolean search {1}
- thesaurus {1}

Content search {6}

- overview of audio features (superficial) {6}
    - low-level (time-, spectrum based)
    - mid-level (rhythm, tonality, hardness, brightness)
    - high-level (timbre, similarity, segmentation, genre, tempo, key)

Query by example {4}

- exemplary model of a distance based clustering

Statistical methods {6}

- methods for classification
- methods for clustering

Possible visualisations {4}

List view

- "classic mode" - interactive list view with sorting

Tags

- automated labeling of audio clips

Graphs

- per file views like magnitude spectrum, waveform etc.
- relationship visualizations (dendrogram, multy-layer cake diagram of labels)
- distribution visualizations (based on audio features)
- Dimensionality reduction techniques

The Lab {38}

The Gap {6}

- Why is it probably the state of the art (what works well)?
- What is quite impossible with a blank text search?
- Which sounds are particulary difficult to find?
- What other representation than text and visualized spectral or energy views (e.g. Waveforms) are possible?
- Some examples.

Research {4}

Overview of a few papers and prototypes and Institutes.

Existing Prototypes {4}

Leading to Freesound Explorer and its user concepts.

the concept {6}

How to enhance this particular one tool to make it feasible for production? Why did I choose the Freesound-Explorer?

Freesound-Explorer as a possible Solution {4}

Author, freesound, papers

technical background {8}

- How does it work in general?
- Freesound API, Client-Server, Model-View-Controller
- t-SNE clustering
- what to modify, how to "plug in"

Review of the existing functionality {4}

- derived from "The Gap" chapter - what is missing?
- Where are my modifications placed within the source code

Proof of Concept {16}

- Implemented additional functionality
- Pictures and descriptions

The classic list view {6}

- interactive list synchronized with the map view

Batch download of selection {4}

- including legal information

Semantic Zones {6}

- e.g. Highlighting frequent tags > Frequent pattern mining

Discussion of Results {14}

Evaluation? {6}

How to show, that the concept works? How to measure speed improvements in the workflow?

Review {4}

Why this is actually less mature for production: Only Online Only freesound

Further Work {4}

see evernote notes...

References

1. the overall intended impression↩