Semantic Audio Engineering and Live Sound Reinforcement

Ryan Stables, DMT Lab
Birmingham City University
Overview...

- Problem(s) definition
- Semantic Audio
  - Semantic mixing systems
  - Semantic audio effects
  - Live sound solutions
- The SAFE Project
  - In-DAW data retrieval
  - Initial findings
  - Plug-in tutorial
- Live sound reinforcement
  - Midas Pro 2c experiments
- The future...
Semantic Audio

Musical Semantics
- Extracting *meaning* from musical signals

- This covers a lot of different areas (e.g.)
  - *Speech Recognition*: can we decode the vocalist’s intention?
  - *Music Informatics*: can we extract the composer’s intention via some musical abstraction.
  - *Semantic Web*: can musical data be packaged in a transferable, searchable and comparable format.
  - *Signal Separation*: can we decompose signals into meaningful subsets?
Semantic Audio

A few applications...

• Music similarity searching and recommender systems
• Cover song identification
• Music transcription and score alignment
• Automated remixing/reproduction systems
• Digital archiving and retrieval
• Performance analysis: tutorials, simulation, etc...
Musical Semantics and Audio Engineering

1. Engineers generally talk in a language that is hard to define computationally.
   - *The bass needs to be tighter, the toms are ringing, the vocals need to be more prominent in the mix.*

2. Parameters of music production systems generally address low-level attributes of music processing
   - Compressor threshold, relative gains, filter parameters.

There is a complex, non-linear relationship between the language in (1) and the parameters in (2).
Semantic Audio

- Context dependent/External Influences

Male Lead Vocal - Rock

Male Lead Vocal - Pop

Electric Guitar

Acoustic Guitar

Bass.1

Bass

Drums.1

Drums
Specifically: Live Sound Reinforcement

• Can we provide abstractions to make interfaces more intuitive?

• Can we make the music production workflow more efficient?

• *Can we extract/generate useful metadata during the process?*
Semantic Audio Engineering
Semantic Mixing

Automatic Mixing Systems

- Dugan, D. (1975)
- Perez-Gonzalez, E. & Reiss, J. (2009)
- Ward, D. et al. (2012)

**Concept**: Automatically balance the faders to produce an intelligible mix across all active channels.
Semantic Mixing


- Cross-adaptively optimizes the loudness ratios between each track
- Uses inter-channel dependencies
- Capable of running in real-time, works well for live environments

- Improvements to cross-channel intelligibility based on complex hearing models
- Optimization required for real-time application due to computational overhead
Semantic Mixing

Semantic Mixing Systems
- Scott, J. & Kim, Y. (2009)
- De Man, B. & Reiss, J. (2013)

**Concept**: Automatically manipulate the faders based on some embedded knowledge of the audio signals.
Semantic Mixing

Live Sound Tools

• Terrell, M. et al. (2010) – Automatic noise gating
• Clifford, A. & Reiss, J. (2011) – Proximity effect detection
• Clifford, A. & Reiss, J. (2013) - Comb filtering reduction

**Concept**: Automatically mitigate/address issues that arise in the live sound environment.
Semantic Mixing

Semantic Mixing

Semantic Audio Effects
• Sabin, A. & Pardo, B. (2008) – 2DEQ
• Cartwright, M. & Pardo, B. (2013) – Social EQ
• Seetharaman, P. & Pardo, B. (2014) - Reverbalize
• Stables et al. (2014) – Semantic multi-effects
• Ma, Z. et al. (2015) – Intelligent dynamic range compression

**Concept:** Adaptively or automatically adjust parameters of predetermined audio effects to achieve a desired context-dependent result, often via some abstraction of the parameter space.
Semantic Mixing

Semantic Mixing

The SAFE Project
SAFE: Semantic Audio Feature Extraction
- Semantic Media fund [EPSRC]

Research Question
- Can we collect representative semantic Audio Engineering data on a large scale?

Objectives:
- Extract semantically annotated metadata during the mixing process.
- Use this data to make the Audio Engineer’s workflow more intuitive.
Why?

Audio Engineering

Semantic Layer

Parameter abstraction
- Intuitive interfaces
- Intelligent Music Production

Data analytics
- Timbre Perception
- Computational Musicology
- Linked data
SAFE Plug-ins

- 4 x plug-ins:
  - Parametric EQ
  - Overdrive
  - Compressor
  - Algorithmic Reverb

- Analytics
  - Semantic descriptors
  - Audio feature data
  - Parameter space
  - Metadata
SAFE: Topology

Server

Descriptor name...

Load... Save...

Analysis...

Natural Language Processing
Dimensionality Reduction
Etc...

Save...

Semantic Descriptor
Parameter Space
Feature Set Pre/Post Gain

Pre/Post Gain

Semantic Descriptor
Parameter Space
Feature Set

Analysis...

Natural Language Processing
Dimensionality Reduction
Etc...
SAFE: Audio Effects

1. **Equaliser**: a five band parametric EQ with three peaking filters and two shelving filters.
2. **Distortion**: Amplitude distortion with tone control (LPF).
3. **Compressor**: a dynamic range compressor with attack and release parameters
4. **Reverb**: an algorithmic reverb based on the figure-of-eight technique proposed by Dattorro (1997)

Effects 1-3 are based on Reiss & McPherson (2014)
SAFE: Feature Extraction Architecture

- Audio features extracted per channel, per frame, per upload, before and after processing, deltas computed on server.
- Metadata extracted per-user-upload.
- 1 x Descriptor string per-upload.
- Parameter set per-upload
SAFE: Feature Extraction Architecture

- 100+ audio features per 20ms frame of audio
- Extracted using the Libxtract C-library: Bullock (2007)
- The next version will support the VAMP framework.
SAFE: Feature Extraction Architecture

- Metadata is encouraged on start-up and available through a UI panel.
- Some fields are stored from previous sessions using a local xml file.
- Users can ignore the panel and fields are left blank.
SAFE: Saving Data

- Free text field, max 500 chars.
- Terms are sent to a MYSQL database via cURL.
- Local copies can be stored in XML format.
- Audio needs to be playing in order to send data.
SAFE: Loading Data

- An incentive for plug-in users to provide data is the load descriptor functionality.
- A descriptor-list is updated in near real-time from the database.
SAFE: Code

- All Plug-ins are written in JUCE, available in VST & AU formats.
- Dedicated SAFE JUCE module, available at: https://github.com/semanticaudio
SAFE: How To...

http://www.semanticaudio.co.uk/documentation/
SAFE Plug-in Reception

The SAFE project teaches computers to understand your musical vocab

The technology can use descriptive music to be explored, giving users a sense of what it means, and is a step closer to understanding music in a way that is similar to human perception. This means that computers can understand and interpret music in a way that is more akin to human cognition.

University launches new software training computers to understand language of musicians

New software launched today by researchers at University of Birmingham aims to reduce the need for expensive training and expensive equipment required to make music, while also improving precision and stability of musical instruments and the music they produce.

The developed software, named 'TuneStation', trains computers to understand the language of musicians by mapping sounds to their names.

The software (the SAFE Project) uses artificial intelligence to allow computers to understand sounds like a human being. The development of the software was motivated by the need for advanced computer-aided training systems in music production, aimed at musicians.

So you want to be a music producer? Just switch on your laptop

A new software, called 'Plug-in Reception', is designed to help music producers create professional-quality recordings of their work. The software, which is available for free, provides users with a range of tools to help them create high-quality music.

Building and customizing music studios will soon be a reality for music producers, with new software making it easier than ever to create professional-quality recordings. The software, called 'Plug-in Reception', allows users to add high-quality audio effects, including reverb, delay, and equalization, to their music productions.

Training computers to understand the language of music

By Michael Cym

The recent push for artificial intelligence is enabling computers to understand the language of music. This is a significant step forward in the field of music production, as it means that computers can now understand and interpret music in a more human-like way.

We often describe human minds as being "wired" and "streaming", but do these words mean anything to a computer?
SAFE Plug-in Reception

“I deleted your plugins. They cant do anything that any stock-plugin from any DAW couldnt (sic) do better. Another compressor, eq, reverb and so on ... yawn.” – Anonymous Gearslutz user

The acronym is terrible, it sounds like an anti-bullying campaign – M. Destrade, Irish Times.

This idea is just as ridiculous as the idea of DJs being redundant is ridiculous. - Anonymous Engadget commenter
Plug-in Reception

Initial release date

British Science Festival

Pageviews
30,077

Users
6,129

<table>
<thead>
<tr>
<th>Country</th>
<th>Sessions</th>
<th>% Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>1,334</td>
<td>16.65%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1,244</td>
<td>15.53%</td>
</tr>
<tr>
<td>France</td>
<td>888</td>
<td>11.08%</td>
</tr>
<tr>
<td>Russia</td>
<td>541</td>
<td>6.75%</td>
</tr>
<tr>
<td>Germany</td>
<td>465</td>
<td>5.80%</td>
</tr>
<tr>
<td>Canada</td>
<td>305</td>
<td>3.81%</td>
</tr>
<tr>
<td>Spain</td>
<td>290</td>
<td>3.62%</td>
</tr>
<tr>
<td>South Korea</td>
<td>272</td>
<td>3.40%</td>
</tr>
<tr>
<td>Italy</td>
<td>263</td>
<td>3.28%</td>
</tr>
<tr>
<td>Japan</td>
<td>242</td>
<td>3.02%</td>
</tr>
</tbody>
</table>
SAFE: Word Frequency Analysis

**Compressor**
- Unique terms: 137
- Syllables-per-term: 1.64
- Term length: 6.03 chars
- Top 3: Punchy, Smooth, Vocal

<table>
<thead>
<tr>
<th>Word</th>
<th>Occurrences</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>punchy</td>
<td>15</td>
<td>5.9%</td>
</tr>
<tr>
<td>smooth</td>
<td>14</td>
<td>5.5%</td>
</tr>
<tr>
<td>vocal</td>
<td>13</td>
<td>5.1%</td>
</tr>
<tr>
<td>nice</td>
<td>12</td>
<td>4.7%</td>
</tr>
<tr>
<td>squashed</td>
<td>6</td>
<td>2.4%</td>
</tr>
<tr>
<td>warm</td>
<td>6</td>
<td>2.4%</td>
</tr>
<tr>
<td>soft</td>
<td>5</td>
<td>2%</td>
</tr>
<tr>
<td>comp</td>
<td>5</td>
<td>2%</td>
</tr>
<tr>
<td>crushed</td>
<td>5</td>
<td>2%</td>
</tr>
<tr>
<td>master</td>
<td>4</td>
<td>1.6%</td>
</tr>
</tbody>
</table>

**Overdrive**
- Unique terms: 106
- Syllables-per-term: 1.63
- Term length: 6.31 chars
- Top 3: Warm, Crunch, Crunchy

<table>
<thead>
<tr>
<th>Word</th>
<th>Occurrences</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>warm</td>
<td>18</td>
<td>9.7%</td>
</tr>
<tr>
<td>crunch</td>
<td>9</td>
<td>4.9%</td>
</tr>
<tr>
<td>crunchy</td>
<td>7</td>
<td>3.8%</td>
</tr>
<tr>
<td>fuzz</td>
<td>7</td>
<td>3.8%</td>
</tr>
<tr>
<td>fuzzy</td>
<td>6</td>
<td>3.2%</td>
</tr>
<tr>
<td>raspy</td>
<td>4</td>
<td>2.2%</td>
</tr>
<tr>
<td>cha</td>
<td>3</td>
<td>1.6%</td>
</tr>
<tr>
<td>harsh</td>
<td>3</td>
<td>1.6%</td>
</tr>
<tr>
<td>bass</td>
<td>3</td>
<td>1.6%</td>
</tr>
<tr>
<td>smooth</td>
<td>3</td>
<td>1.6%</td>
</tr>
</tbody>
</table>
## SAFE: Word Frequency Analysis

### Equaliser
- Unique terms: 188
- Syllables-per-term: 1.48
- Term length: 6.55 chars
- Top 3: Warm, Bright, Clear

<table>
<thead>
<tr>
<th>Word</th>
<th>Occurrences</th>
<th>Frequency</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>warm</td>
<td>48</td>
<td>11%</td>
<td>1</td>
</tr>
<tr>
<td>bright</td>
<td>30</td>
<td>8.7%</td>
<td>2</td>
</tr>
<tr>
<td>clear</td>
<td>11</td>
<td>2.5%</td>
<td>3</td>
</tr>
<tr>
<td>thin</td>
<td>10</td>
<td>2.3%</td>
<td>4</td>
</tr>
<tr>
<td>boomy</td>
<td>8</td>
<td>1.8%</td>
<td>5</td>
</tr>
<tr>
<td>clean</td>
<td>8</td>
<td>1.8%</td>
<td>5</td>
</tr>
<tr>
<td>and</td>
<td>8</td>
<td>1.8%</td>
<td>5</td>
</tr>
<tr>
<td>airy</td>
<td>6</td>
<td>1.4%</td>
<td>6</td>
</tr>
<tr>
<td>tinny</td>
<td>6</td>
<td>1.4%</td>
<td>6</td>
</tr>
</tbody>
</table>

### Reverb
- Unique terms: 117
- Syllables-per-term: 1.73
- Term length: 5.60 chars
- Top 3: Room, Small, Hall

<table>
<thead>
<tr>
<th>Word</th>
<th>Occurrences</th>
<th>Frequency</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>room</td>
<td>14</td>
<td>5.9%</td>
<td>1</td>
</tr>
<tr>
<td>small</td>
<td>7</td>
<td>3%</td>
<td>2</td>
</tr>
<tr>
<td>hall</td>
<td>7</td>
<td>3%</td>
<td>2</td>
</tr>
<tr>
<td>subtle</td>
<td>7</td>
<td>3%</td>
<td>2</td>
</tr>
<tr>
<td>dreamy</td>
<td>6</td>
<td>2.5%</td>
<td>3</td>
</tr>
<tr>
<td>natural</td>
<td>6</td>
<td>2.5%</td>
<td>3</td>
</tr>
<tr>
<td>drum</td>
<td>6</td>
<td>2.5%</td>
<td>3</td>
</tr>
<tr>
<td>damp</td>
<td>6</td>
<td>2.5%</td>
<td>3</td>
</tr>
<tr>
<td>echoy</td>
<td>5</td>
<td>2.1%</td>
<td>4</td>
</tr>
<tr>
<td>roomy</td>
<td>5</td>
<td>2.1%</td>
<td>4</td>
</tr>
</tbody>
</table>
**SAFE: Word Frequency Analysis**

<table>
<thead>
<tr>
<th>Word</th>
<th>Occurrences</th>
<th>Frequency</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>warm</td>
<td>45</td>
<td>5.5%</td>
<td>1</td>
</tr>
<tr>
<td>bright</td>
<td>33</td>
<td>4.1%</td>
<td>2</td>
</tr>
<tr>
<td>smooth</td>
<td>17</td>
<td>2.1%</td>
<td>3</td>
</tr>
<tr>
<td>room</td>
<td>17</td>
<td>2.1%</td>
<td>3</td>
</tr>
<tr>
<td>punchy</td>
<td>15</td>
<td>1.8%</td>
<td>4</td>
</tr>
<tr>
<td>vocal</td>
<td>15</td>
<td>1.8%</td>
<td>4</td>
</tr>
<tr>
<td>nice</td>
<td>15</td>
<td>1.8%</td>
<td>4</td>
</tr>
<tr>
<td>subtle</td>
<td>11</td>
<td>1.4%</td>
<td>5</td>
</tr>
<tr>
<td>clear</td>
<td>10</td>
<td>1.2%</td>
<td>6</td>
</tr>
<tr>
<td>clean</td>
<td>9</td>
<td>1.1%</td>
<td>7</td>
</tr>
</tbody>
</table>

**All Plug-ins**
Unique terms: 362  syllables-per-term: 1.65  
Total terms: 848  Top 3: Warm, Bright, Smooth
SAFE: Term Highlights

**Compressor**
- “tickled”, “phat”, “kissing”, “die hard”
- “you need to write something in the box, fool!”

**Distortion**
- “beastly”, “destroyed”, “underwater”, “menacing”
- “dramatic tangs”, “el cruncho”, “caramel”
SAFE: Term Highlights

Reverb
- “DP-SAR-1”, “DIDI 1”, “fluffy”, “creamy”,
- “wholesome”, “150215”, “J”

Equaliser
- “bastard”, “drum 1”, “drum 2”, “drum 3”
- “pianombient”, “sticks out in the mix a bit more”
- “we call it bass”
Experiments in:
Live Sound
Session Analytics

- Collaboration with Music Group Research (Midas, Klark-Technik, Behringer, etc)
- Data captured in real-time from a modified Midas console with SAFE architecture.
Session Analytics

Midas Pro2C
- 56 mic/line-ins
- Assignable faders and groups.
- 12 multichannel FX engines
- Dedicated effects parameter space with digital display.

Modified to print session data to console and receive state commands.
15 live UK music festival recordings
- 32 channels per each recording
- All bands from a given stage across a full day of music
- Recordings include preliminary material and sound checks

30+ Engineers asked to mix as if they were at the festival
Objectives

1. **Automatic metadata derivation**
   - Via console-parameter analysis
   - Via audio-feature analysis

2. **Workflow optimization**
   - Provide intuitive interface abstractions
1. Metadata Derivation

This can be derived from:
- parameter information
- Audio feature data

Instruments/Genre/etc...
Engineer background/

Microphones being tested
End of Soundcheck
Feedback caused by microphone 1
Compression applied
Using machine learning, we are able to estimate instrument classes:

- Essid et al. (2006): SVM, octave-band intensities [93%]
- Tjoa & Liu (2010): NMF-derived temporal features and SVM [92%]
- And genre tags...
  - Tzanetakis & Cook (2002): Low/high level audio features with GMM/K-NN [61%]
  - Ezzaidi & Rouat (2006): MFCCs with GMMs [73-99%]
Metadata Derivation

- It’s relatively easy to estimate a user’s location based on their IP address, and the physical location of their ISP.
- Language is slightly more complex, but can be estimated using a language model and IP.

Genre, Instrument, Current Location, Language, Experience, Age
Production experience classifiers are less common.

- This information is useful for: weighting descriptor terms, partitioning data, user analytics.
- We need to arrange specific experiments due to anonymous data.
- We can use both audio feature data and parameter-space data
Parameter Tracking

- Parameter modulation data was gathered during a series of live-simulation production sessions.
- Subjects with varying production experience were given recorded mixing tasks to perform.
- Tasks were selected based on creative and corrective procedures.
- e.g. sec 1. make channel n warmer/brighter. Sec 2. route input channels to stereo bus.
Parameter Tracking
Modulation Patterns

- Initial findings show range of common basis functions.
- These tend to be correlated with level of production (defined through listening tests).
We extract temporal features from the tracks, including modulation spectra features, and DTW-features.

Various modulation periods are used:

- single adjustments,
- descriptor-wise adjustments
- track-wise adjustments

Clustering is then applied to identify regions of similarity.
Clustering

- Agglomerative/hierarchical clustering applied to modulation patterns
- Clusters compared to labeled targets
Workflow Optimization

How do engineers use the consoles in a live setting?

Using data can we develop novel/intuitive interfaces for live mixing?
Workflow Optimization
Workflow Optimization
Workflow Optimization: Parameter Abstraction

- Can we control complex audio effects with low-dimensional spaces?
- Can we use both audio feature data and parameter space data.

- Currently testing:
  - Dimensionality Reduction algorithms
    - PCA, MDS, tSNE, etc
  - Visualisation techniques
    - Voronoi, Lloyd’s relaxation
  - Parameter space representation
    - Linear, n-D, object/path-based
Semantic Representation

• Can we control complex audio processes with minimal parameters?
• Can we use both audio feature data and parameter space data.
• Map between low/high-dimensional space based on semantic representation
Reduced Dimensionality Interfaces

- Visualisations are available online
- Figures are updated in near-real-time using MDS (with the R package: ).
- Two sources: params and audio features.
- Terms are placed on coordinates of 2-D map.
- Term size is determined by confidence score (sum of variance):

\[
Conf = \frac{1}{N} \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} (x_{k,n} - \mu_n)^2
\]
Reduced Dimensionality Interfaces

Distortion...

An MDS plot showing the Parameter space from the SAFEDistortion.

First Dimension
Second Dimension

raspy
fuzzy
warm
decimated
crunch
crunchy
clipped
tinny
loud
twinkly
beastly
metal
fuzz
buzzy
driven
shredded
warm tube
granulated
grainy
fat
boomy
destroyed
subby
heavy
bluesy
distorted
sharp
punch

An MDS plot showing the Differences space from the SAFEDistortion.

First Dimension
Second Dimension

raspy
fuzzy
warm
decimated
crunch
crunchy
clipped
tinny
loud
twinkly
beastly
metal
fuzz
buzzy
driven
shredded
warm tube
granulated
grainy
fat
boomy
destroyed
subby
heavy
bluesy
distorted
sharp
punch
Reduced Dimensionality Interfaces

Compressor...

An MDS plot showing the Parameter space from the SAFECompressor.

First Dimension
Second Dimension

- loud
- crushed
- warm
- nice
- compressed
- quiet
- lv2
- tight
- bass
- kick
- pumping
- tickled
- smooth
- squashed
- smoothed
- choppy
- pinned
- glued
- boost
- max
- present
- soft comp master

An MDS plot showing the Differences space from the SAFECompressor.

First Dimension
Second Dimension
Reduced Dimensionality Interfaces

Reverb...

An MDS plot showing the Parameter space from the SAFEReverb.

First Dimension

Second Dimension

distant
dreamy
slap
slapback
delay
subtle
emphasize
mental
massive
big
cloudy
fluffy
warm
lush
rich
layered
vocals
huge
ghostly

dp
sar
reverb
pad
dark
piano
room
small drum room
big drum room
medium drum room
fat drums
gated
church
large room
hall
small room
room
roomy

An MDS plot showing the Differences space from the SAFEReverb.

First Dimension

Second Dimension

distant
dreamy
slap
slapback
delay
subtle
emphasize
mental
massive
big
cloudy
fluffy
warm
lush
rich
layered
vocals
huge
ghostly

dp
sar
reverb
pad
dark
piano
room
small drum room
big drum room
medium drum room
fat drums
gated
church
large room
hall
small room
room
roomy
P2. PC-EQ-2D

- Real-time param modulation
- Dimension assignment
- Descriptor suggestions
The Future...
P3. Semantic API

- SAFE API:
  - Arbitrary parameter space optimisation
  - Can we extract the same parameter information from different interfaces?

Parameter Space Mapping
P3. Interface Abstraction

- Auto-loading plug-ins
  - what do I need to make something warm?
  - Can we load combinations of audio effects based on production data
P3. Content Reproduction

Content-based semantic music production:
- Parameter learning: *Make my guitar sound like Waterloo Sunset by The Kinks*
- Automatic remixing: *Make the Blowin’ in the Wind sound like it was produced by Rick Rubin*
Anyway...

Please get involved!!!!

Plug-ins and data available from: www.semanticaudio.co.uk

ryan.stables@bcu.ac.uk

@otmiv
References


References


References


References


Thanks...

• Matthew Cheshire, DMT Lab, BCU
• Brecht De Man, C4DM, QMUL
• Sean Enderby, DMT Lab, BCU
• Spyridon Stasis, DMT Lab, BCU

• Cham Athwal, DMT Lab, BCU
• George Fazekas, C4DM, QMUL
• Alessandro Palladini, Music Group Research.
• Josh Reiss, C4DM, QMUL
• Mark Sandler, C4DM, QMUL

• EPSRC Semantic Media Project (semanticmedia.org.uk)