

# SAFE: A System for the Extraction and Retrieval of Semantic Audio Descriptors

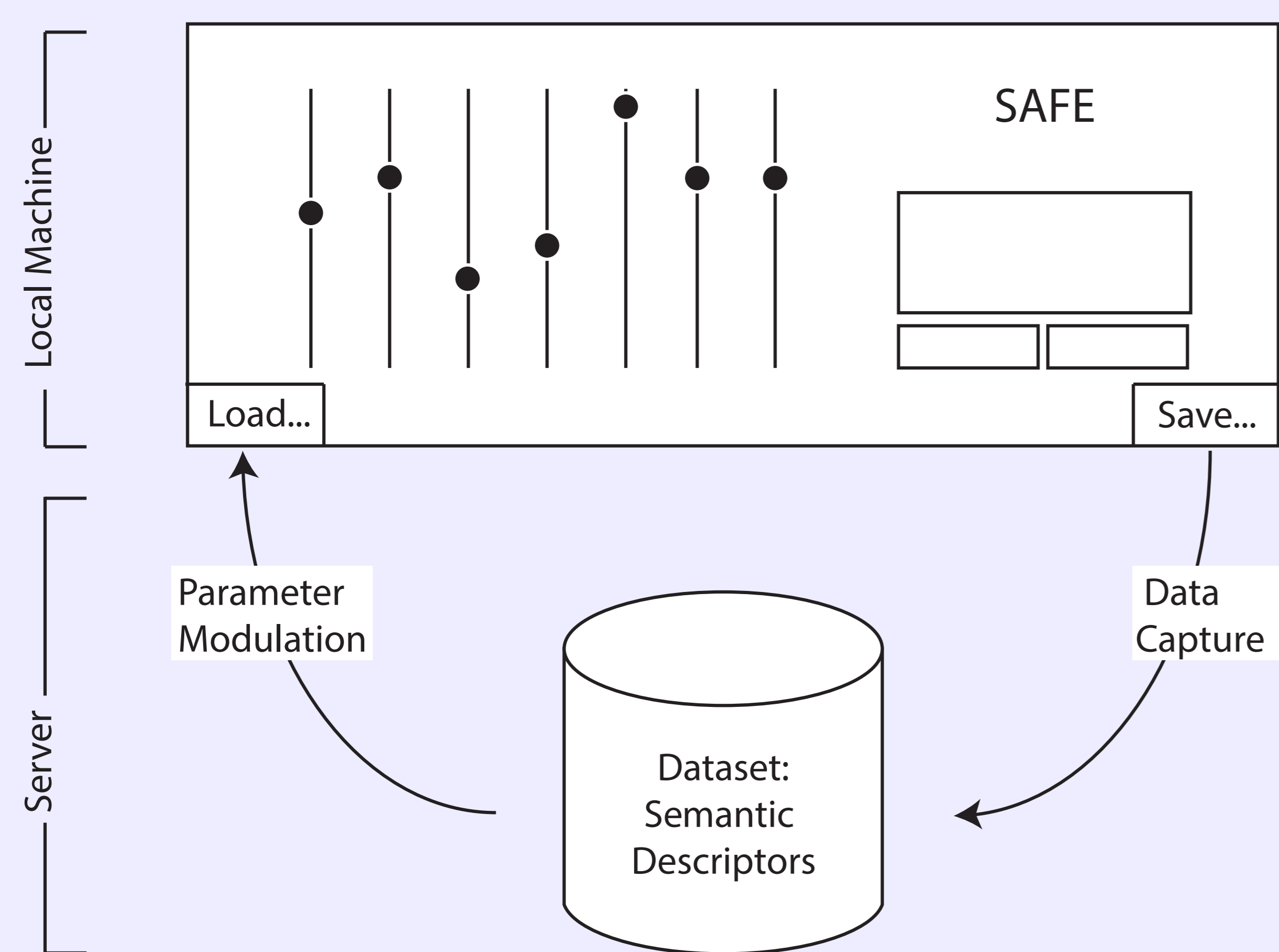
Ryan Stables<sup>1</sup>, Sean Enderby<sup>1</sup>, Brecht De Man<sup>2</sup>, György Fazekas<sup>2</sup>, and Joshua D. Reiss<sup>2</sup>  
ryan.stables@bcu.ac.uk, sean.enderby@mail.bcu.ac.uk  
{b.deman, g.fazekas, joshua.reiss}@qmul.ac.uk

## 1 Introduction

- Perceptual, semantic or high-level descriptors (*bright*, *warm* or *fuzzy*) are not (always) mathematically defined, and often have a nonlinear relationship to the audio effect's parameter space
- Audio effects are controlled using parameters that interface with statistical, low-level properties of the audio signal
- Musical semantics commonly collected through controlled tests: pre-defined samples and descriptors and small sample sizes
- We map the relationship between low-level parameter space and high-level musical timbre descriptions, designing audio effects that can be controlled with semantically understandable parameters
- We compile a dataset of timbral transformation data, complete with reliable semantic annotation

## 2 The SAFE Project

Suite of Digital Audio Workstation (DAW) plug-ins, encouraging free-text annotation of parameter states during production process. Extracts and retrieves semantic audio descriptors, low-level audio features and parameters from within music production workflow



Downloadable from [www.semanticaudio.co.uk](http://www.semanticaudio.co.uk)

## 3 System overview

- **Digital Audio Effects**
  - VST and Audio Unit formats
  - Amplitude distortion effect with tone control, algorithmic reverb, dynamic range compressor, and 5 band parametric EQ
- **Analysis Framework**
  - Stores  $N \times M$  matrix of audio features (number of frames  $N$  and number of audio features  $M$ )
  - Uses LibXtract library for feature extraction [1]
  - Stores  $1 \times P$  parameter vector (number of UI parameters  $P$ )
  - Optional metadata: user age, location, production experience, song genre, and track musical instrument
- **Parameter Modulation**
  - Semantic profiles updated in real-time
  - Terms hierarchically partitioned into metadata categories (instrument, genre and location-specific terms)
  - Transformations from nonlinear effects applied relative to signal RMS
- **Missing Data Approximation**
  - User's location approximated by geolocation data (IP address)
  - Musical instrument and genre tags estimated using unsupervised machine learning algorithm, applied to reduced-dimensionality of audio feature set

## 4 Acknowledgements

**EPSRC**  
Engineering and Physical Sciences  
Research Council

Funded by the Engineering and Physical Sciences Research Council (EPSRC) grant (EP/J010375/1): "Semantic Media"

[1] J. Bullock, "Libxtract: A lightweight library for audio feature extraction," in *Proceedings of the International Computer Music Conference*, vol. 43, 2007.

