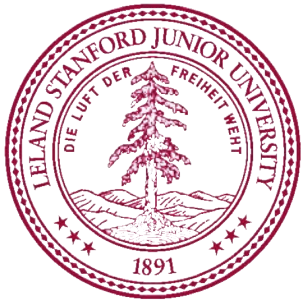
ISSE – An Interactive Source Separation Editor, Part I

Nicholas J. Bryan
Stanford University



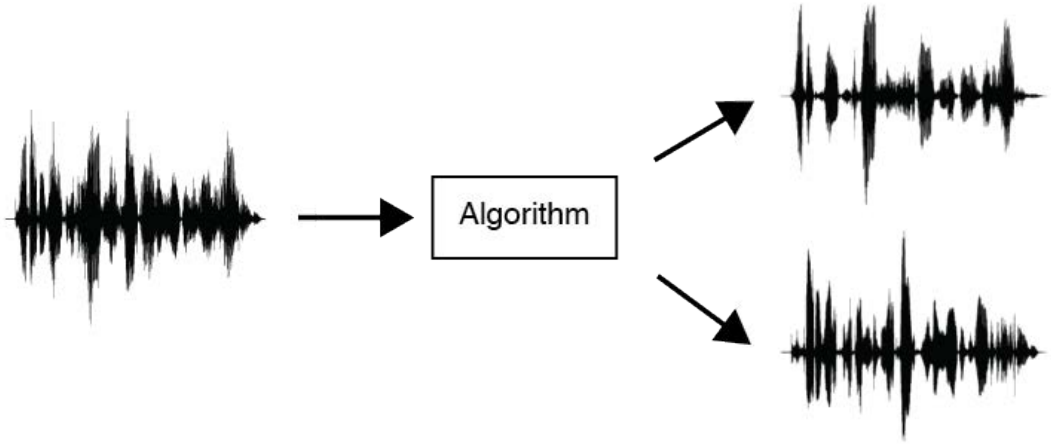
Sound Check

Overview

- Introduction
- Demonstration
- Software
- Perspective
- Algorithm
- Download!

Single-Channel Source Separation

- Generally, a very difficult problem
- Large underdetermined system of equations
- Many strategies to overcome this problem
- We will employ interactive feedback



Denoising

- Live broadcast/TV
- Live recordings (e.g. orchestra cough)
- Studio recordings

Spatial Audio and Upmixing

- Consumer electronics

- Home stereo systems

Audio Post-Production and Remastering

- Extracting dialogue from old movies
- Redo music and sound fx
- Removing noisy hum, hiss, etc. in old tape recordings

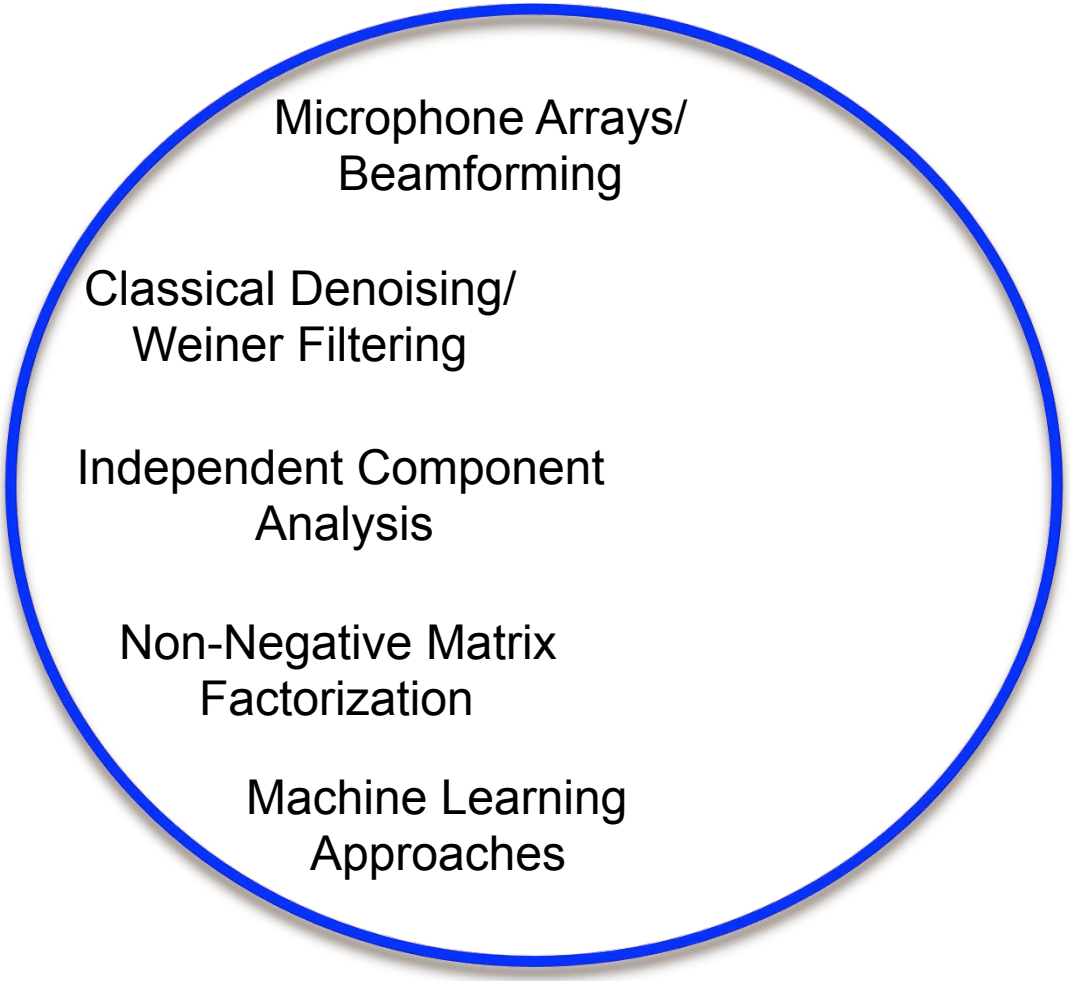
Music Information Retrieval

- Automatic speech recognition
- Tempo estimation
- Key detection
- Melody extraction

Music Remixing and Content Creation

- DJs
- Sound artists
- Electronic musicians
- Audio/recording engineers

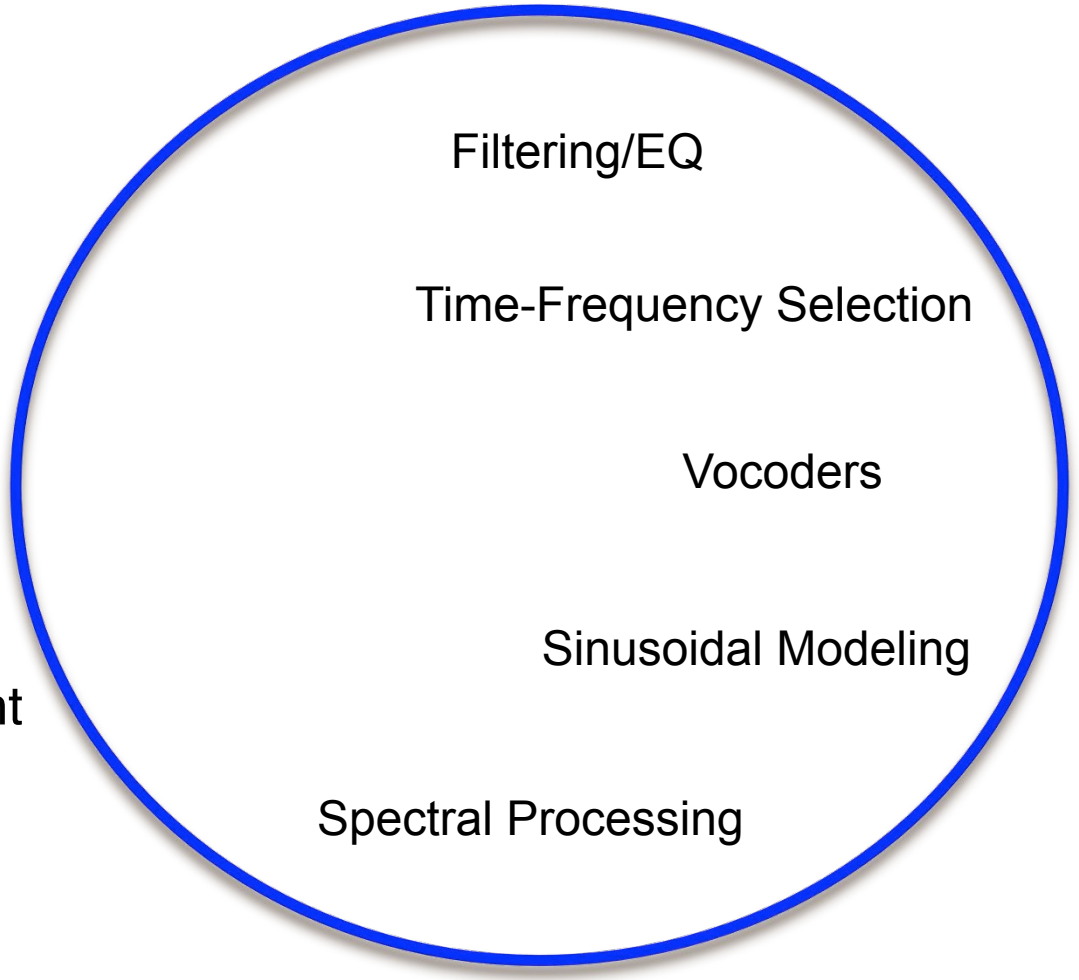
Automatic



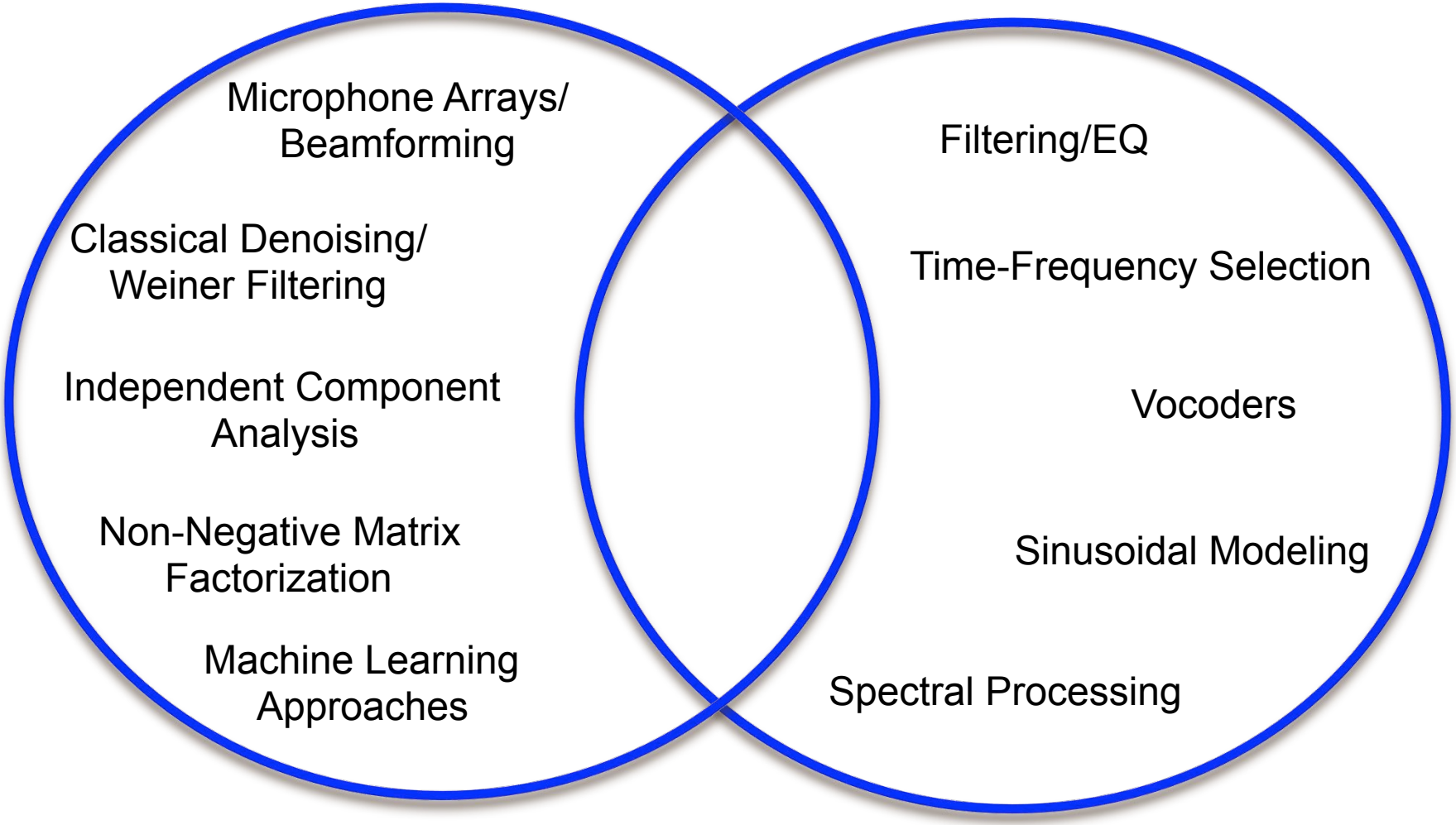
- Powerful
- Operate independently
- Operate based on the geometry or statistics
- Can breakdown/fail
- No way to correct for errors

Manual

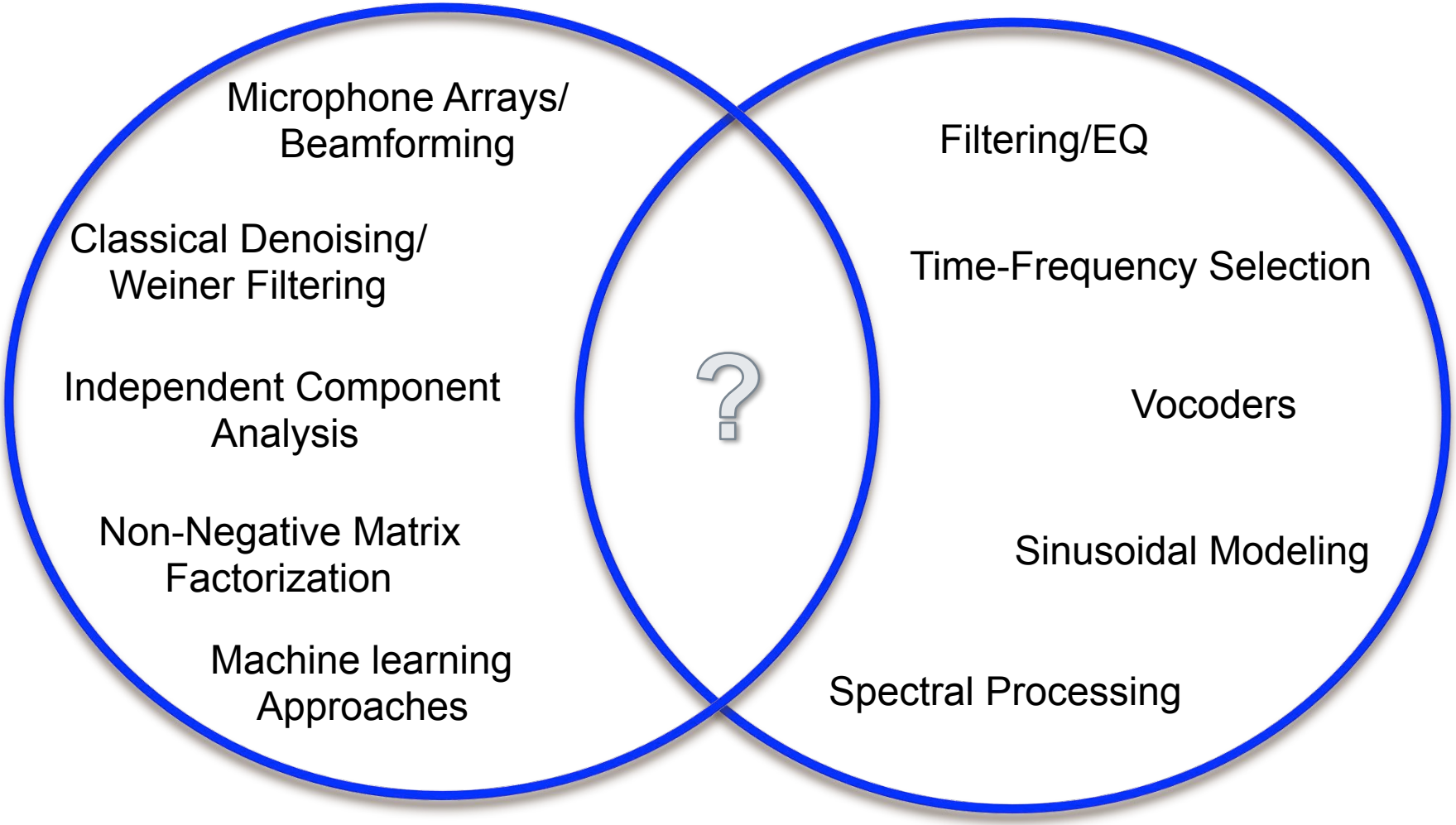
- Tedious, manual work
- Work on simple cases
- Cannot achieve quality of automatic approaches
- Allow adjustment refinement



Automatic vs. Manual



Automatic and Manual



User's Perspective

- DJs
- Sound artists
- Electronic musicians
- Audio/recording engineers

User's Perspective

- How does an audio/recording engineer perform source separation?

Audio editing, plugin, processing

- For what purpose?

Music remixing and content creation

- What skill level?

Advanced knowledge and skill

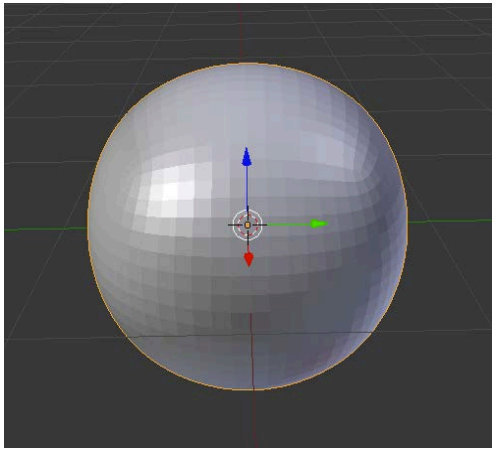
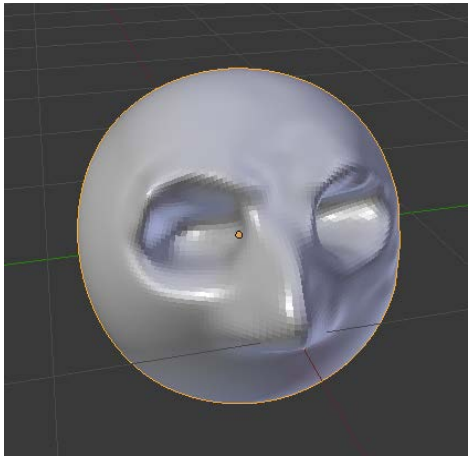
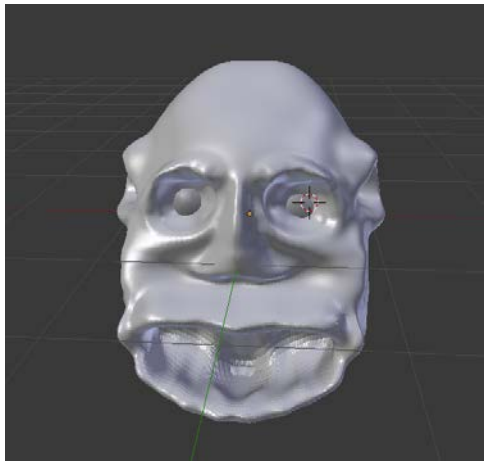
Analogy

- Photoshop “layers”



…

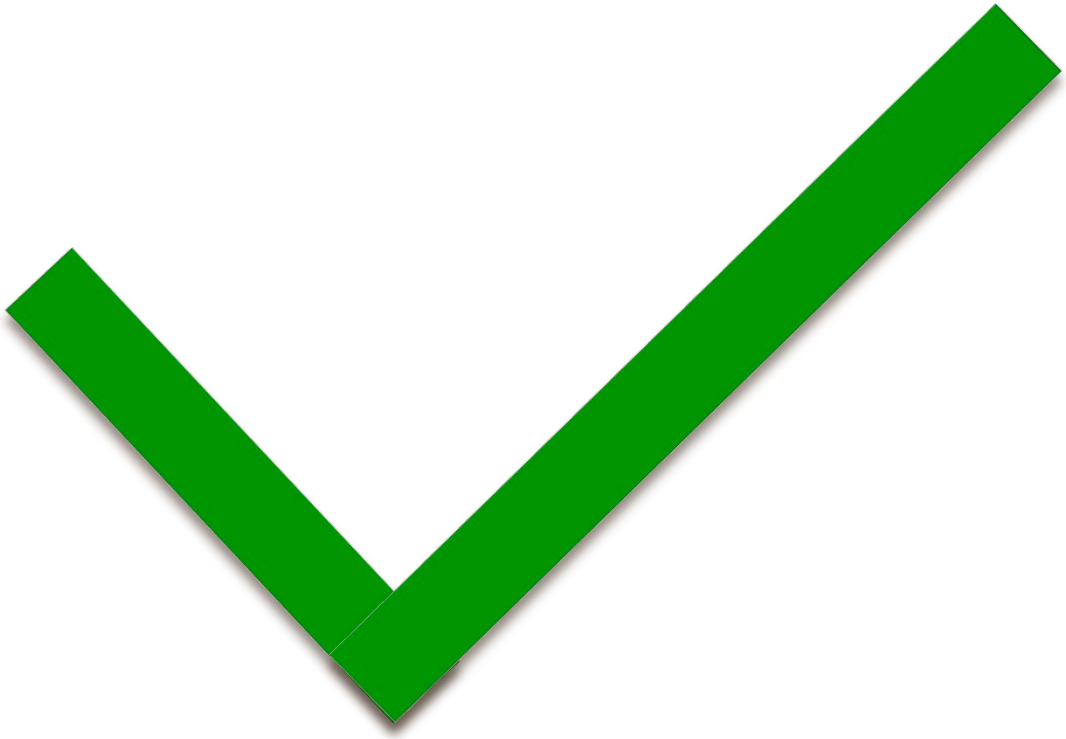
- 3D Sculpting


…

…


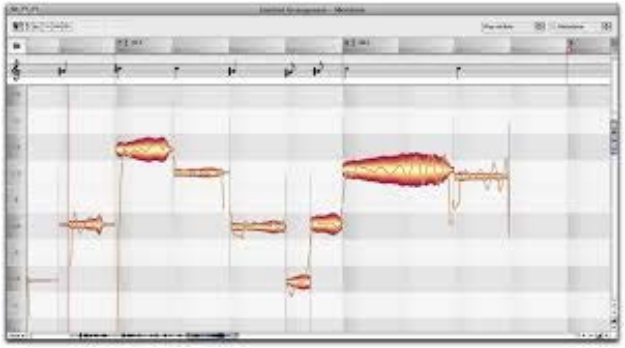
- User-feedback is key!

Automatic and Manual

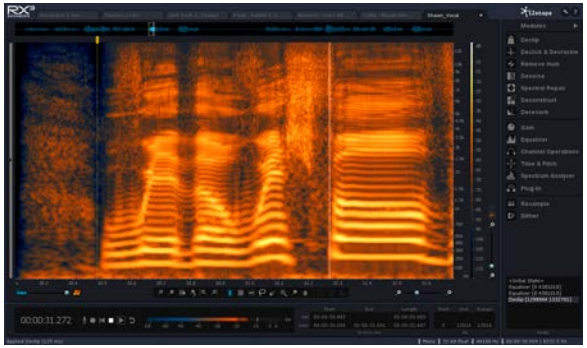
- Balance between machine and the human user



Automatic and Manual



- Very exciting products over the past decade
- Melodyne/Celemony



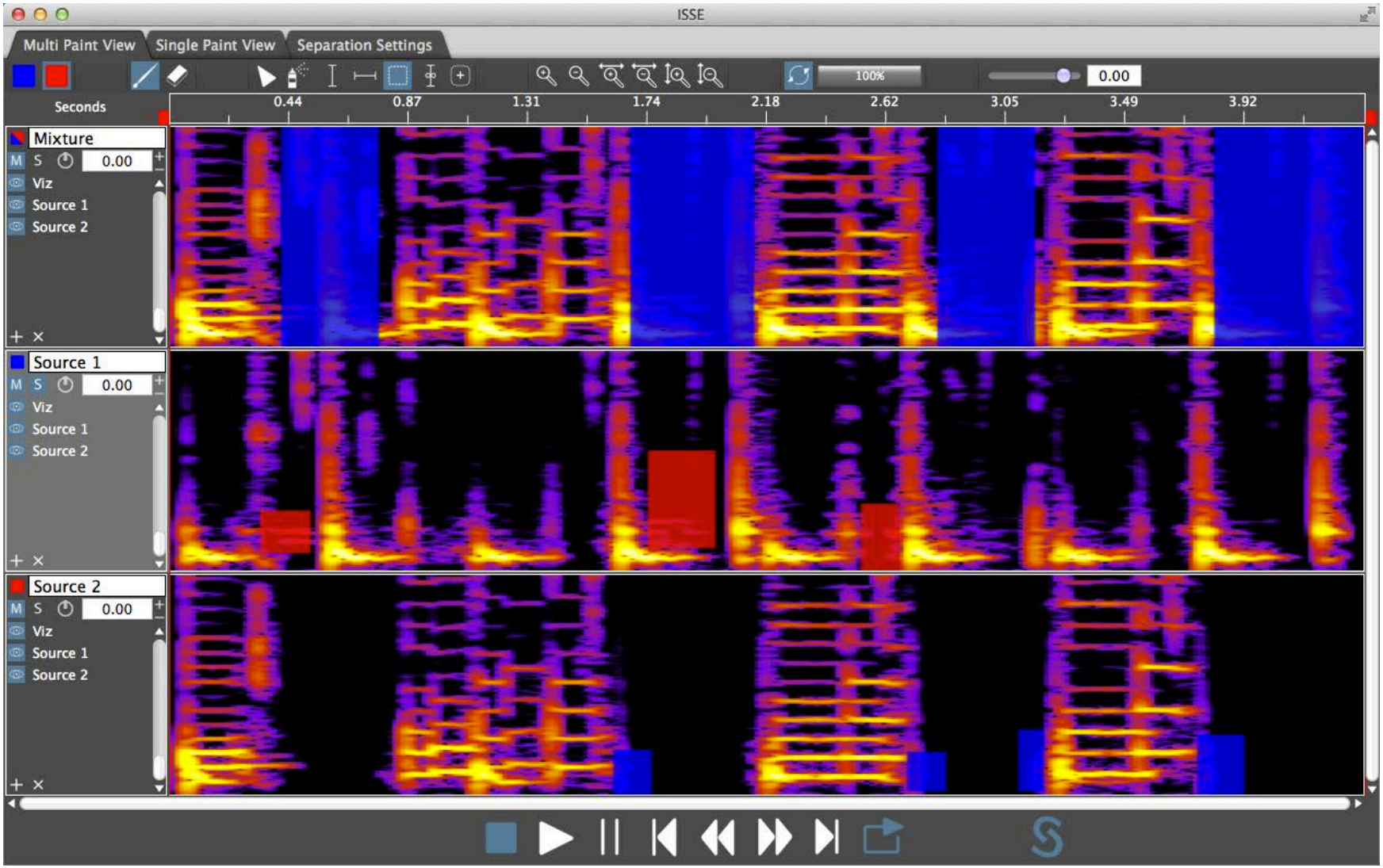
- iZotope RX
- Adobe Audition

- Many others (please shout out!)

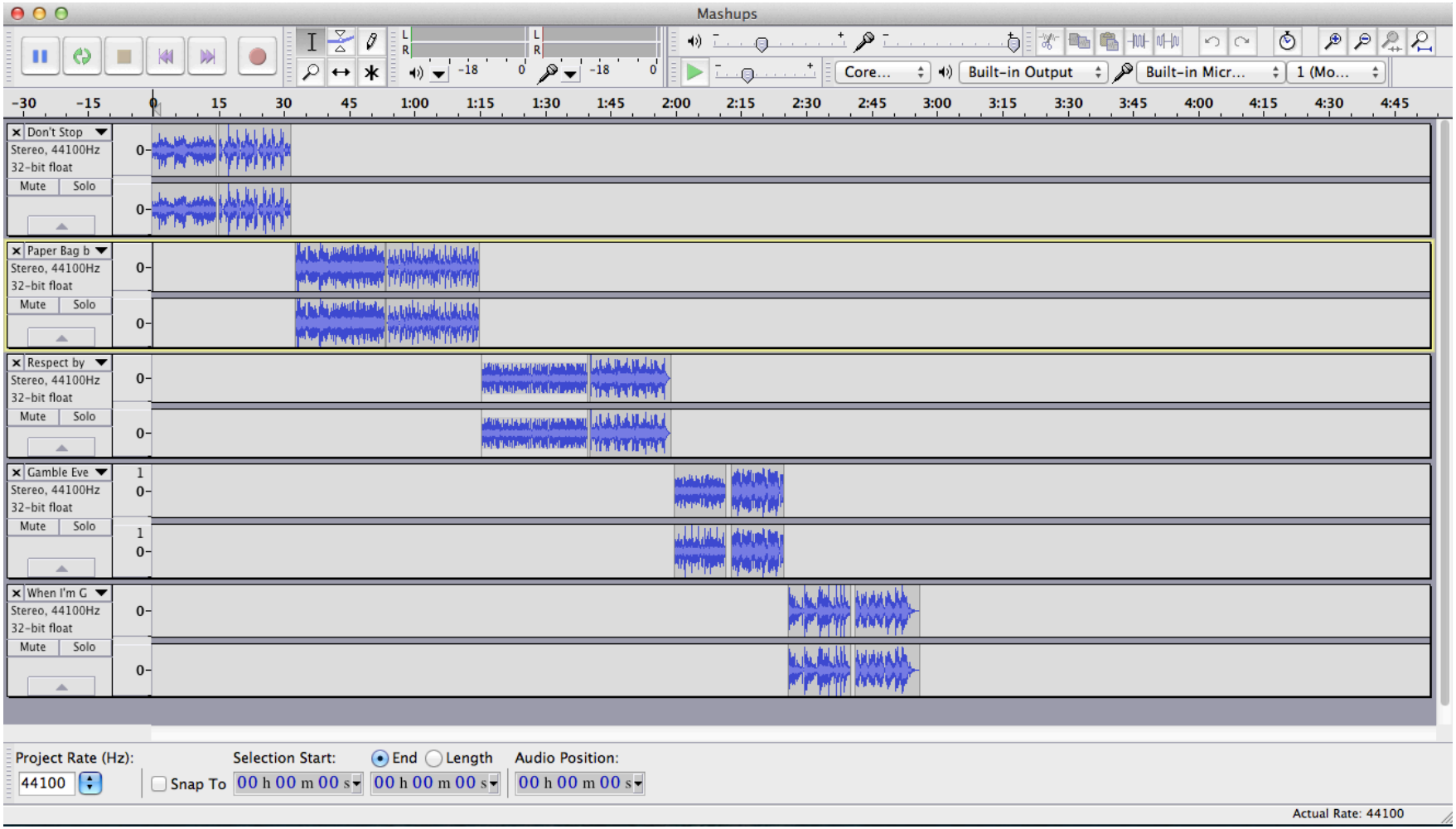
Overview

- Introduction
- Demonstration
- Software
- Perspective
- Algorithm
- Download!

ISSE – An interactive Source Separation Editor



Vocal Extraction Examples



Overview

- Introduction
- Demonstration
- Software
- Perspective
- Algorithm
- Download!

ISSE Overview

- Alpha release, open-source, freely available project.
- Cross-platform (OSX, Windows, Linux), C++.
- Useful for applications such as:

Music remixing/upmixing

Audio-based forensics

Audio denoising

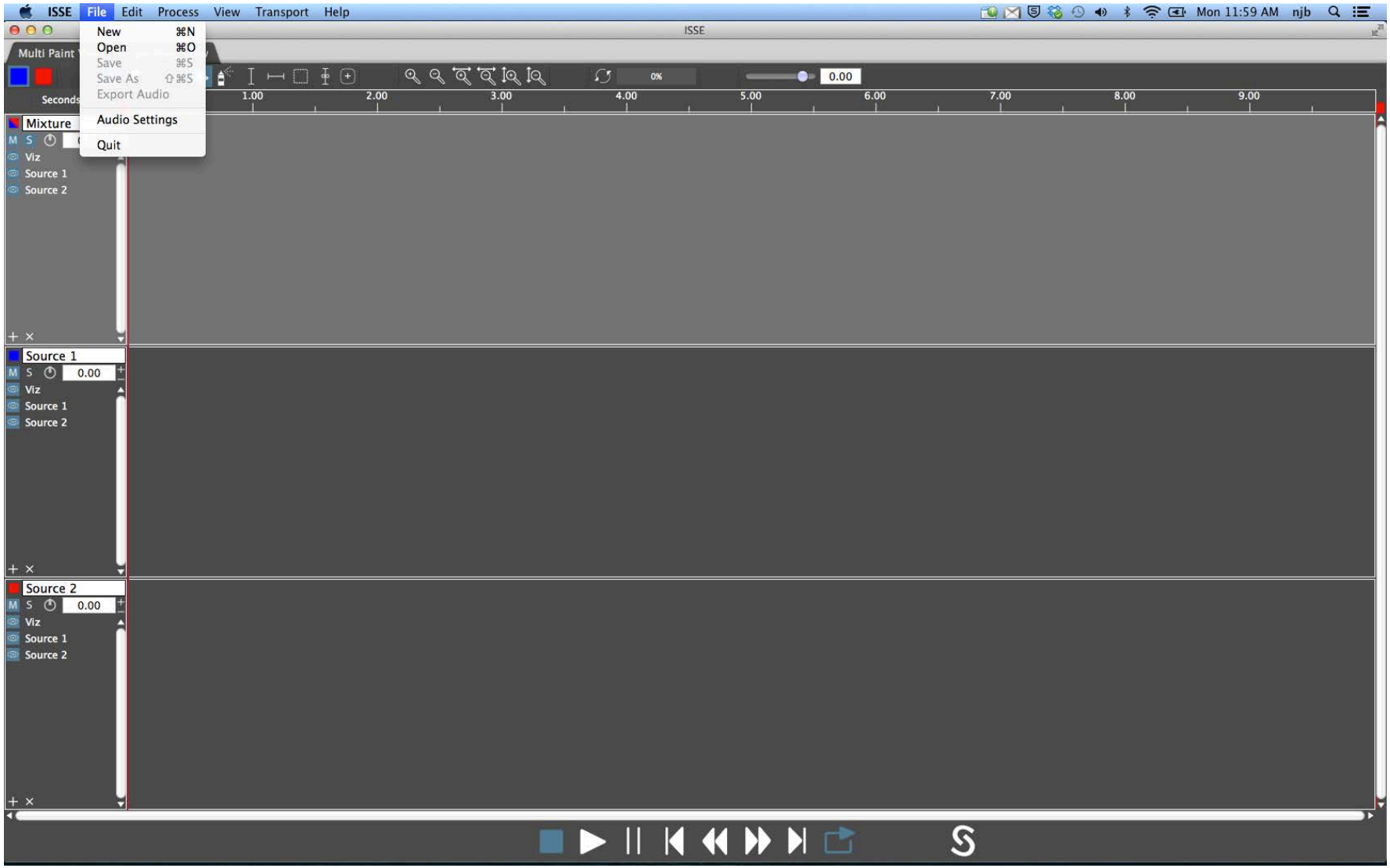
Dialogue extraction

- Released in collaboration with:
  - Adobe Research
  - Stanford University

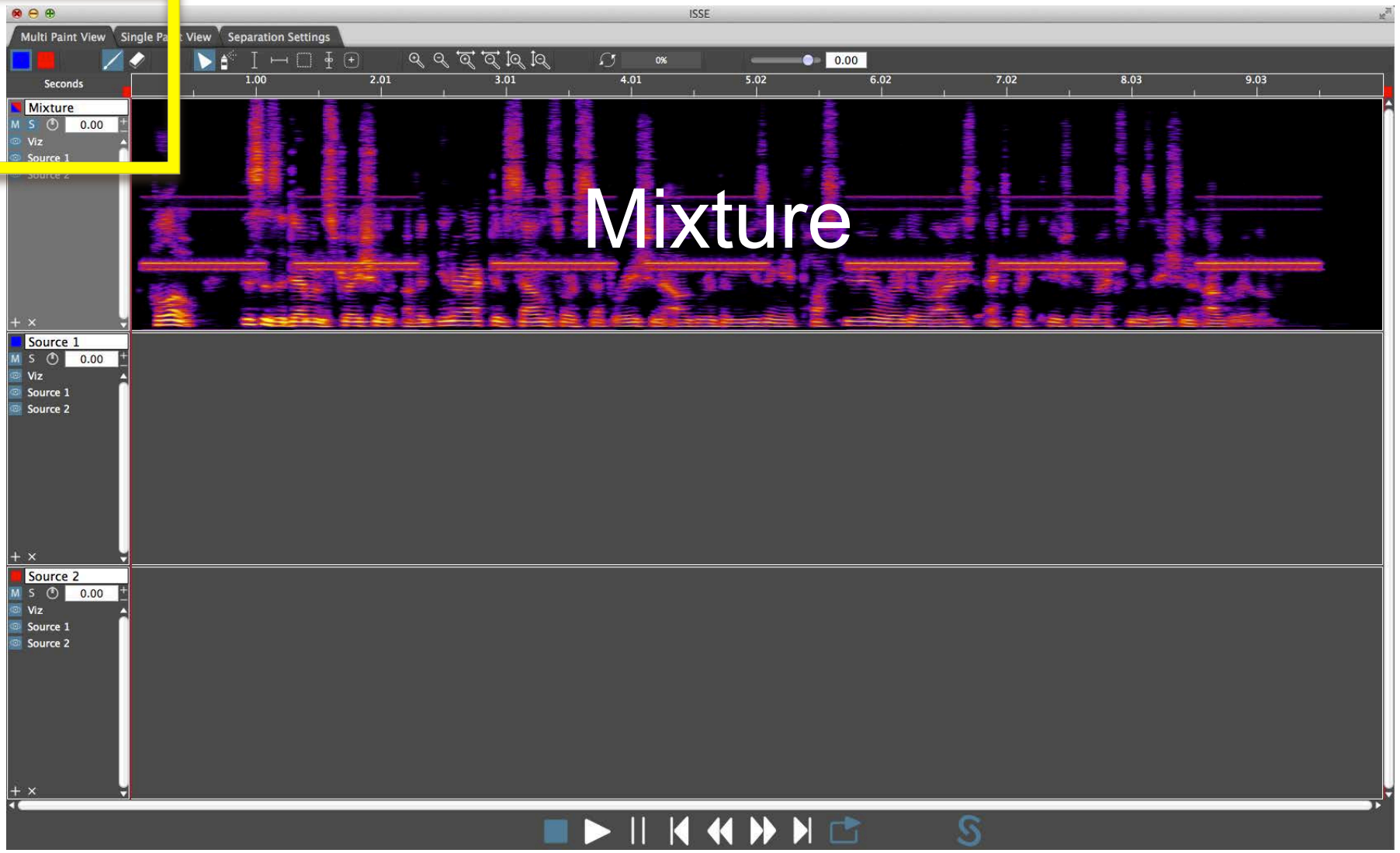
Third-Party Dependencies

- JUCE (juce.com)
  - An extensive, cross-platform C++ toolkit
  - GUI windowing, widgets, audio playback
  - Open-source with GPL license
- Eigen (eigen.tuxfamily.org)
  - C++ template library for linear algebra and matrix library
  - Mozilla Public License 2
- FFTW (fftw.org)
  - Cross-platform C library for fast Fourier transforms
  - GPL license

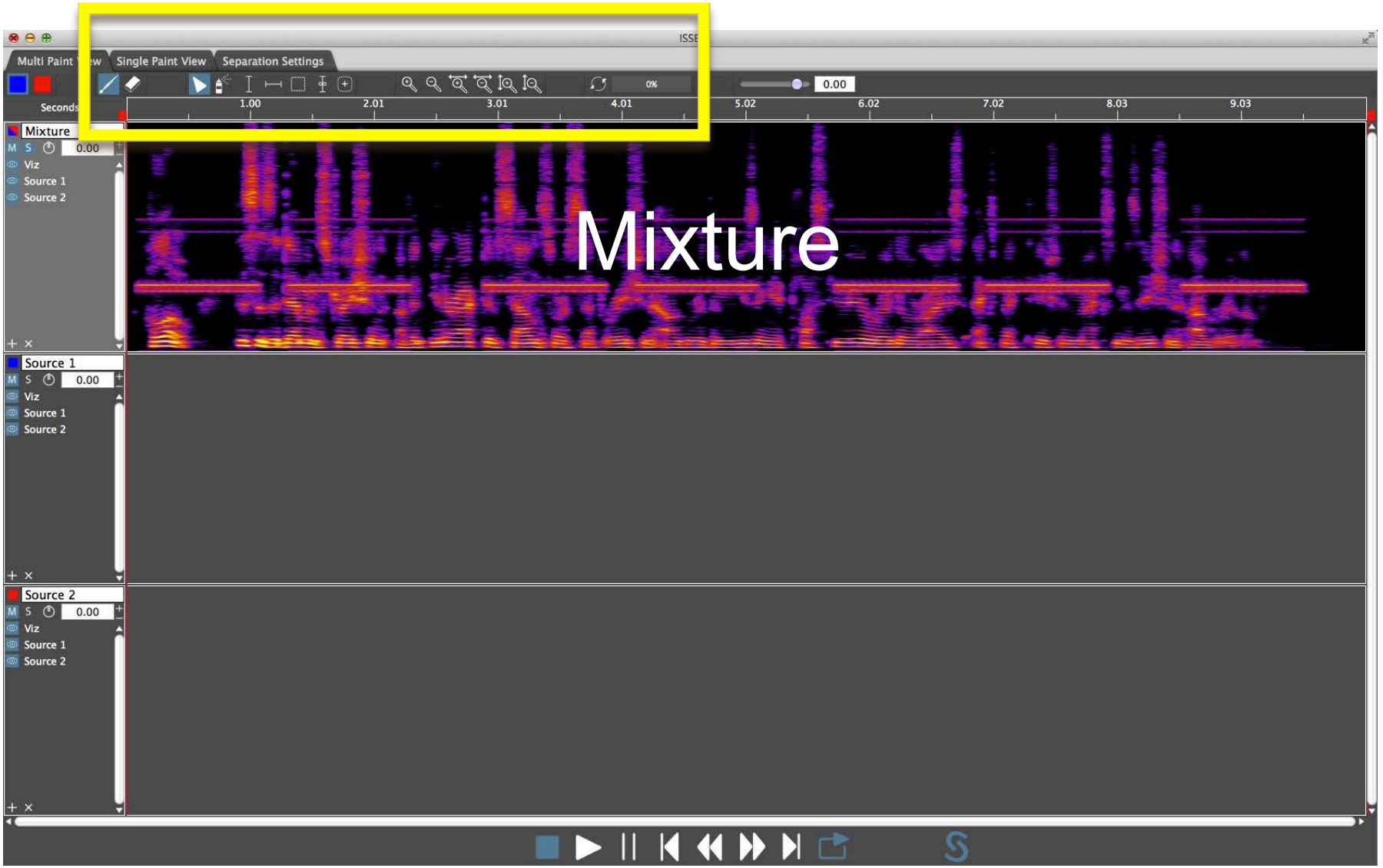
Load in a short 5-30 second recording (a chorus, verse, etc.)



Choose What to Separate

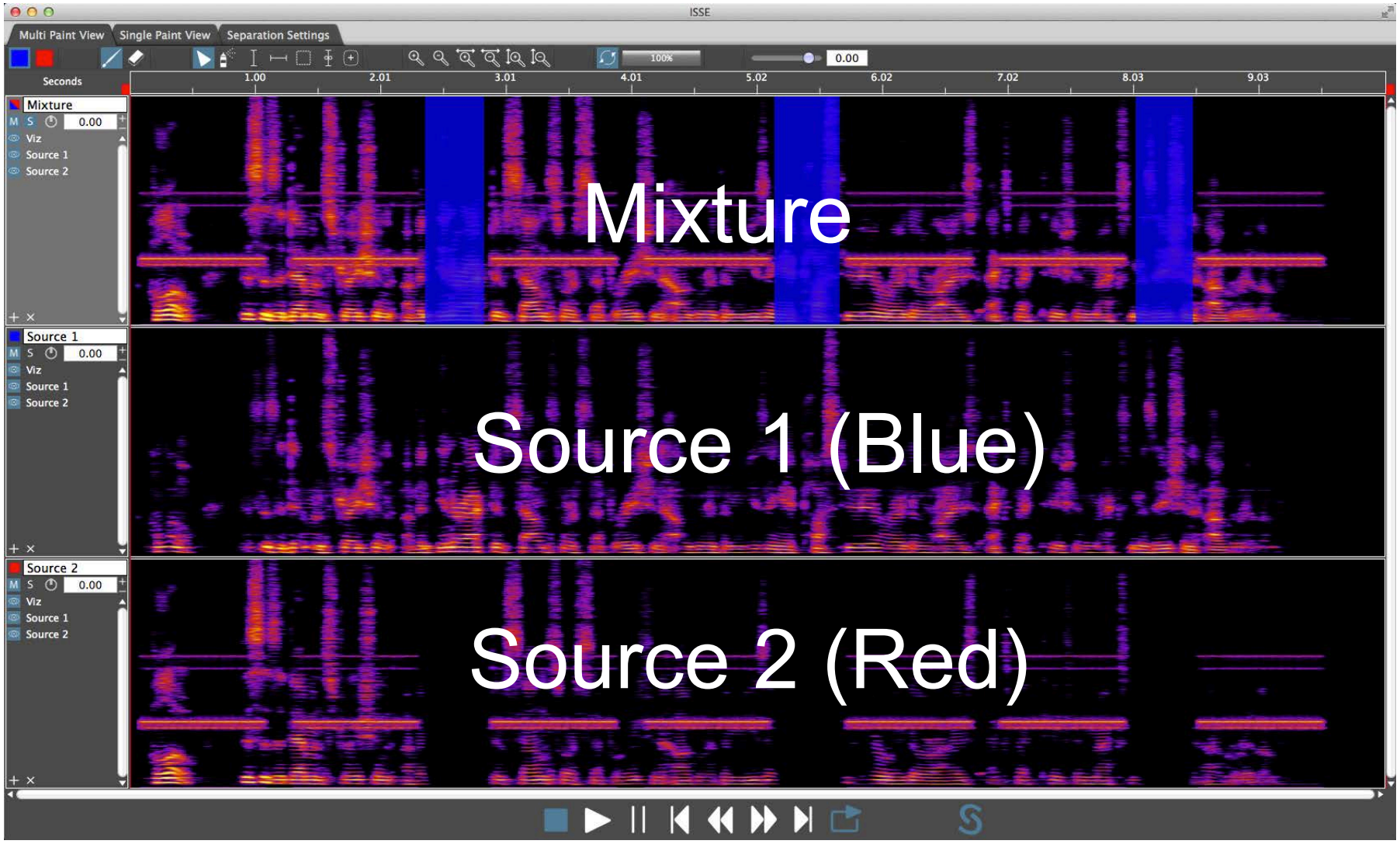


Start Painting, Processing, & Listening



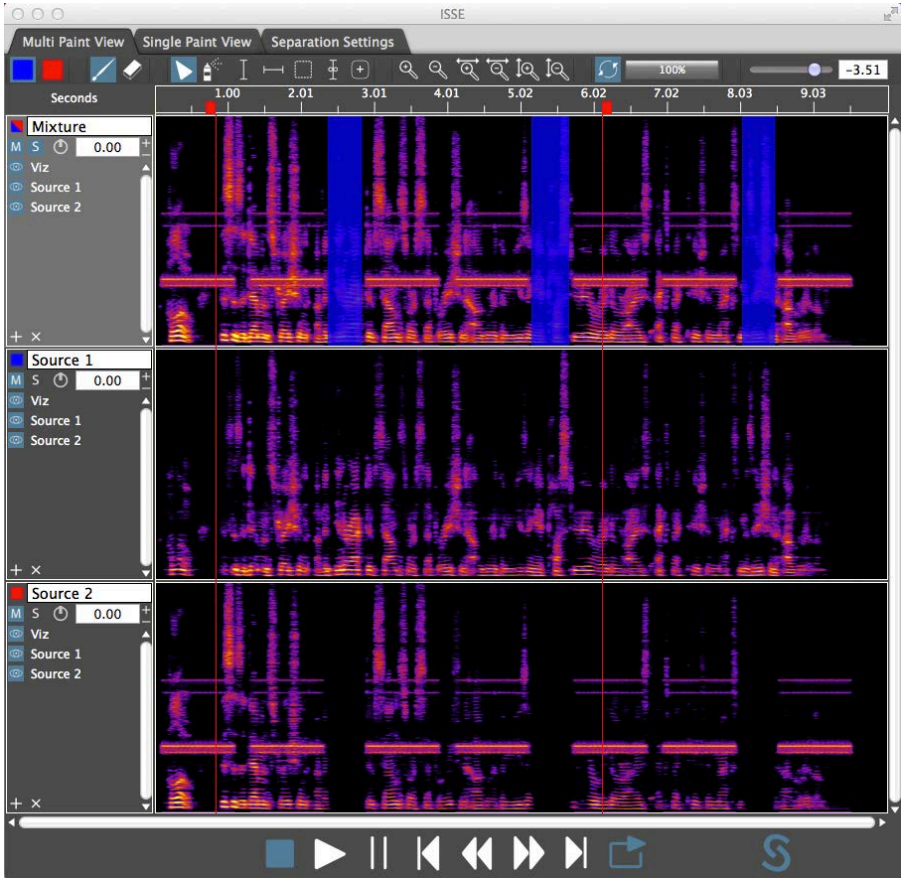
Automatically updates processing after annotations when processing is on

Start Painting, Processing, & Listening

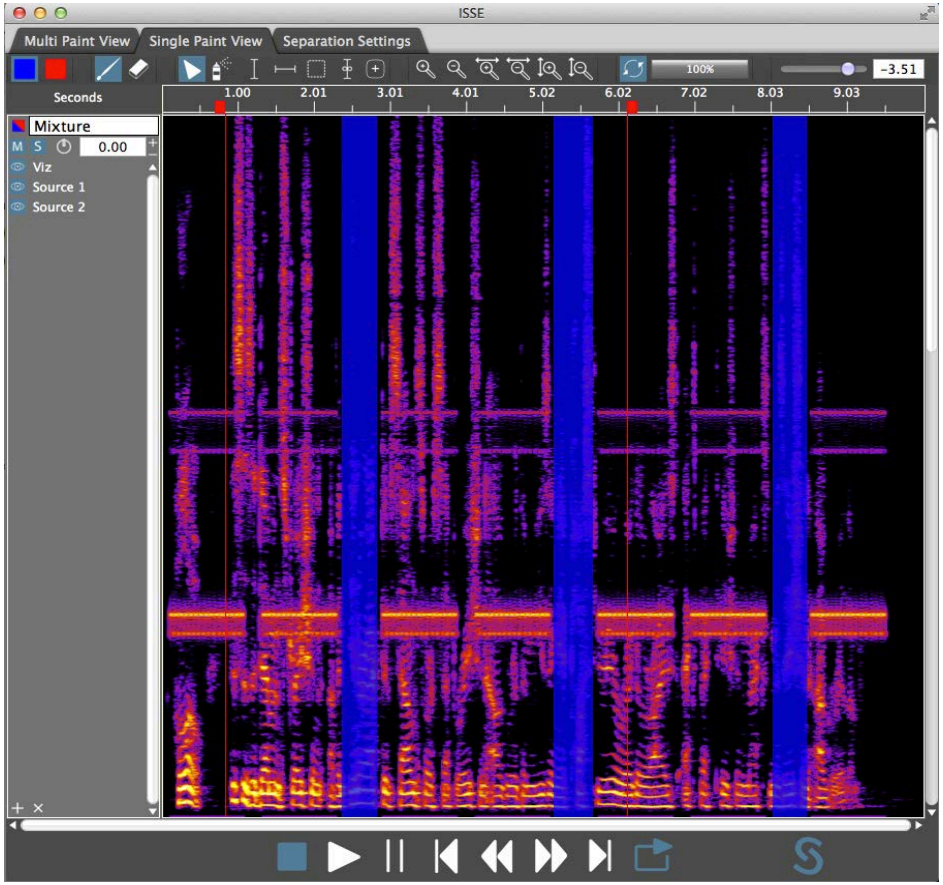


Display Controls I

Multi Paint View

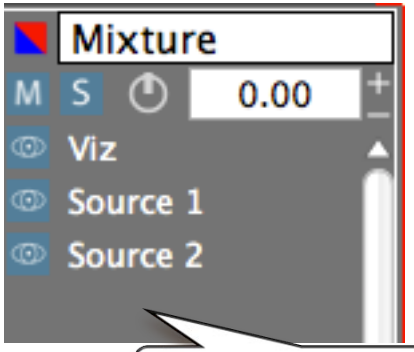


Single Paint View

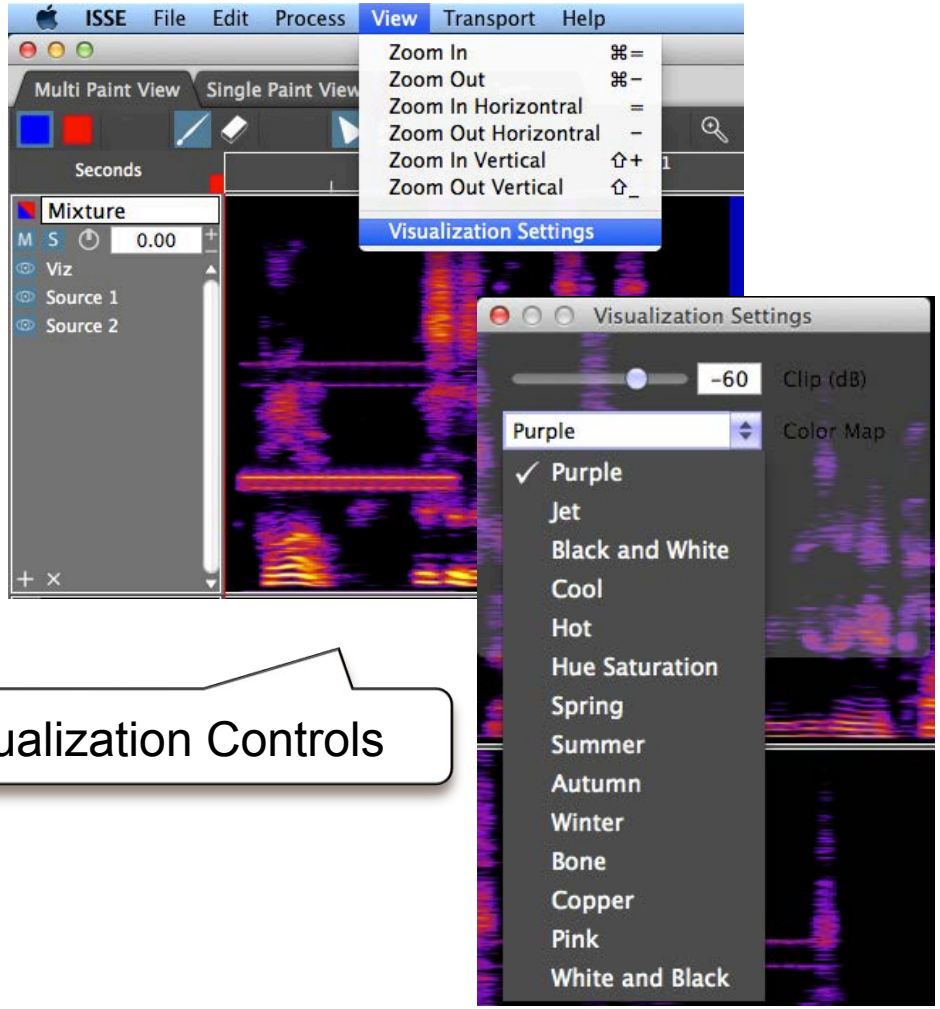


Display Controls II

Zoom controls



Layer visibility



Visualization Controls

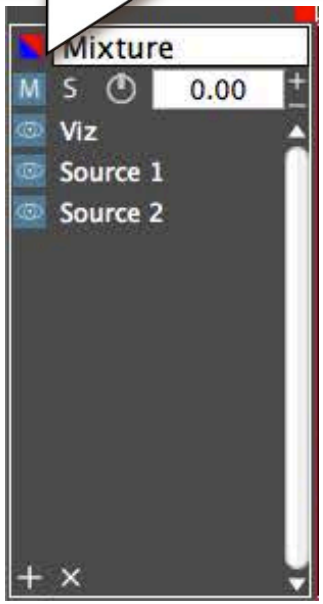
Transport Control

Master volume

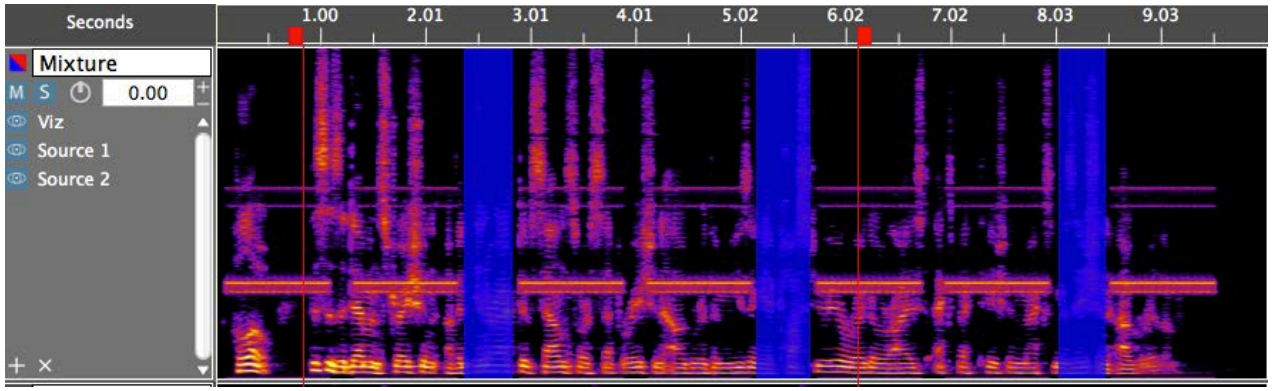
Play, pause, stop, loop, etc.

Auto Solo

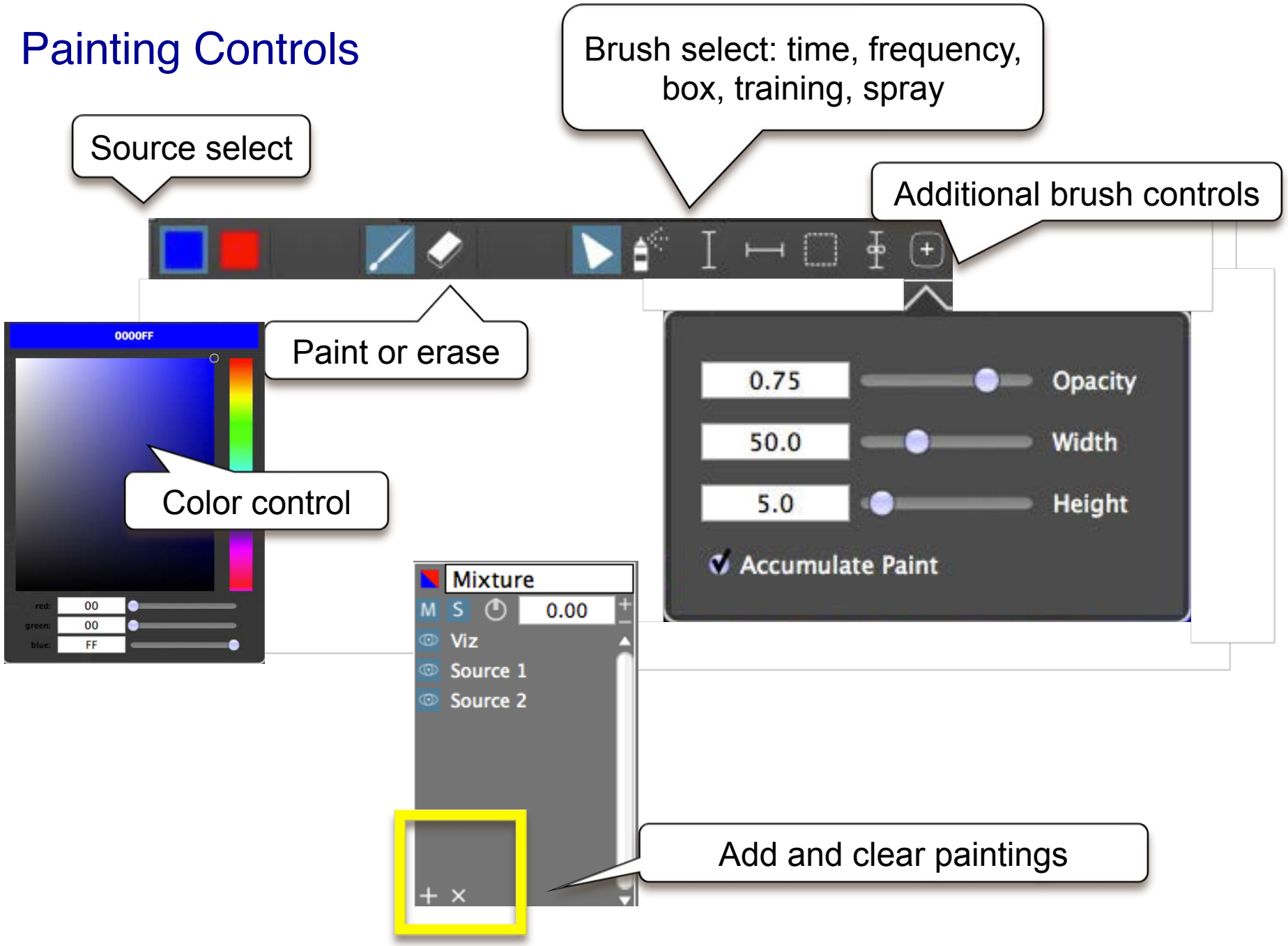
Mute, solo, volume



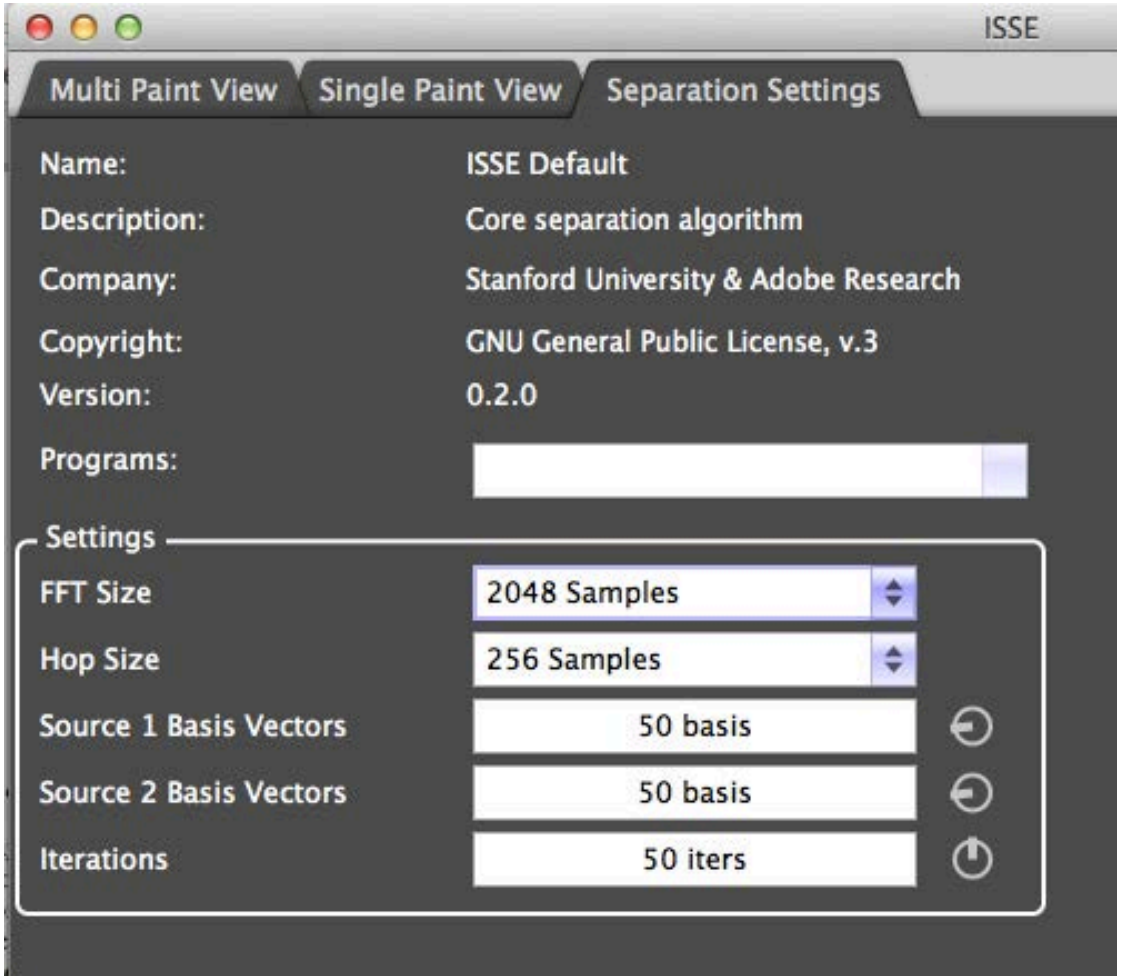
Loop markers, playback cursor placement



Painting Controls



Processing Control



On/Off, progress

FFT and HOP size
Changing clears paintings

Iterations, number
of basis vectors

Parameters all effect processing speed

Save, Loading, Undo, Redo

- Project saving and loading
- Undo/redo
  - Cleared on save
  - Currently implementation is memory intensive

Overview

- Introduction
- Demonstration
- Software
- Perspective
- Algorithm
- Download!

Perspective I

- Machine learning approach that constantly **adapts** to user annotations
- User is **training** the separation algorithm
- Separation algorithm is **learning** what to separate
- A single, **local** annotation can have a **global** effect on separation quality

Perspective II

- **Reduces** manual **effort** required by the user
- **Not copying** the pixel data underneath the annotations
- Improves upon automatic approaches via user-**refinement**
- Indirectly incorporate a **perceptual objective** into the separation algorithm

Strategy

- High-quality separation will take **time** (separate 100+, 1000+ times)
- Work with **short** sound files
- Paint on regions **large-to-small**
- **Annotate**, **listen**, and **refine**
- Each **iteration** is (hopefully) an improvement. If not, undo and continue.

Strategy: Large-to-Small Annotations

1. Time regions (time brush, training brush, box brush)

2. Frequency regions (frequency brush, box brush)

3. Large time-frequency regions (box brush, spray brush)

4. Harmonics, fricatives, transients, and detail (box brush, spray brush)

Limitations

- Significant learning curve
  - Difficult at first
  - You will improve
- Will not work well on all sounds
  - Very dense or noisy sounds
  - Highly orchestrated music
- Bugs
  - Slow and/or crashes with long files (try 5-30 seconds)
  - A few issues with older OS…

But!

- We’re constantly working on upgrades
- Interested in gaining community developers
- Will have a third-party plugin format

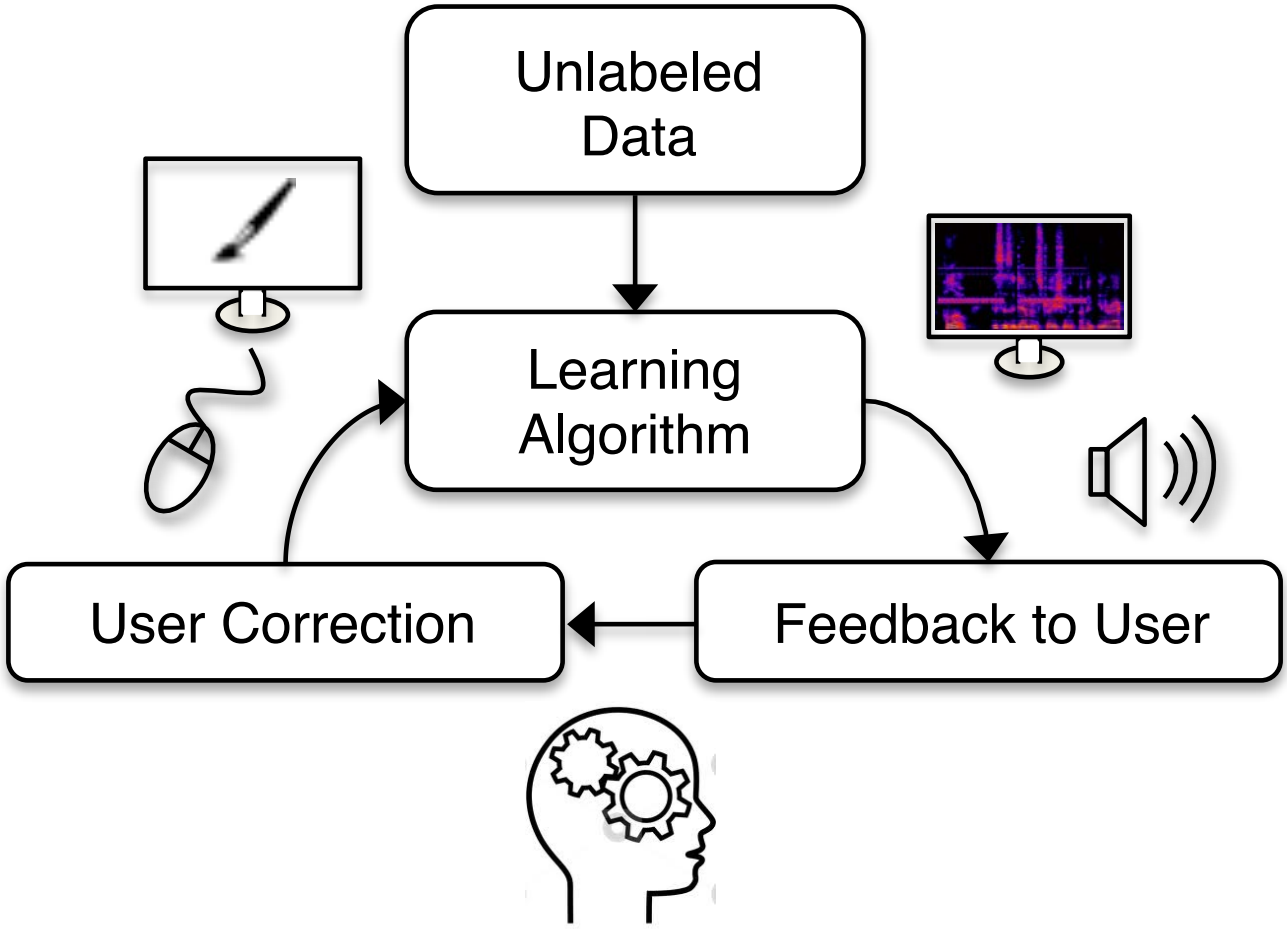
Overview

- Introduction
- Demonstration
- Software
- Perspective
- Algorithm
- Download!

How does it work?

- A model-based approach (NMF)
- Build a model of each sound source that together explain the mixture
- Learn the parameters of the model from data
- Leverage the user to help with the parameter estimation
- Interactive machine learning

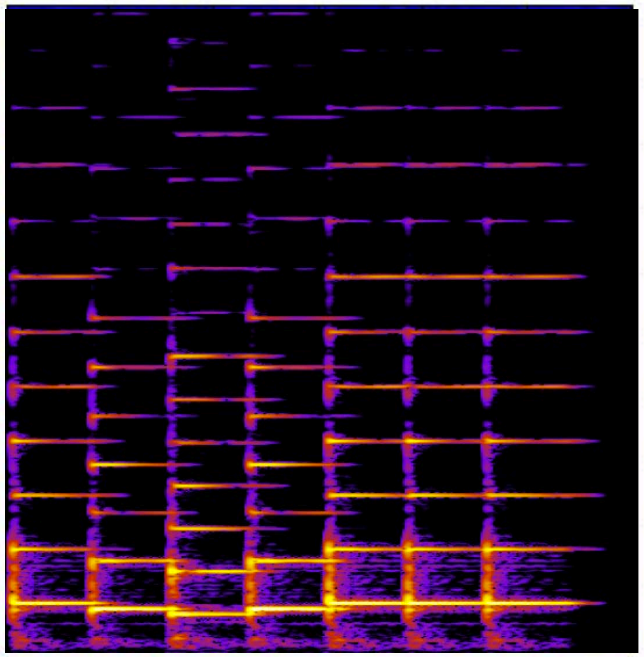
Block Diagram



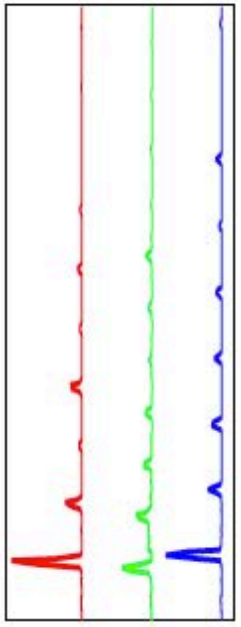
Non-negative matrix factorization

- Non-negative matrix factorization & probabilistic models
  - [Lee & Seung, 2001]
  - [Smaragdis & Brown 2003]
  - [Raj & Smaragdis 2005, Smaragdis et al., 2006]
  - [Virtanen et al. 2007]
  - [Févotte et al. 2009]
  - Mysore, Fitzgerald, and many others

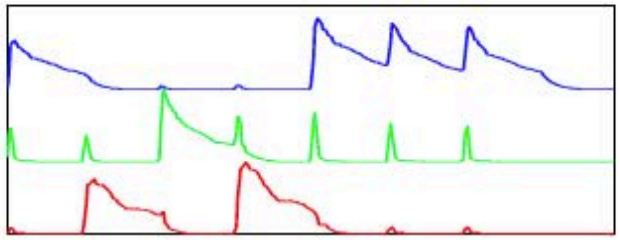
NMF With Spectrogram Data



Spectrogram



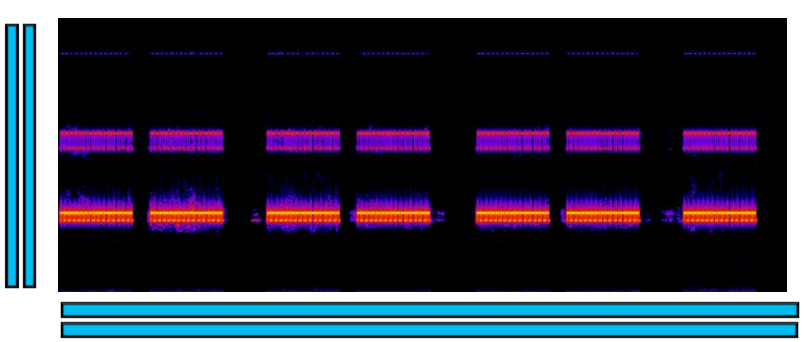
Frequency Model

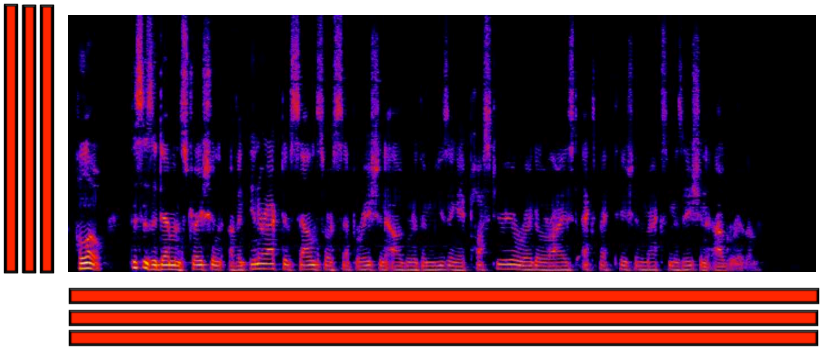


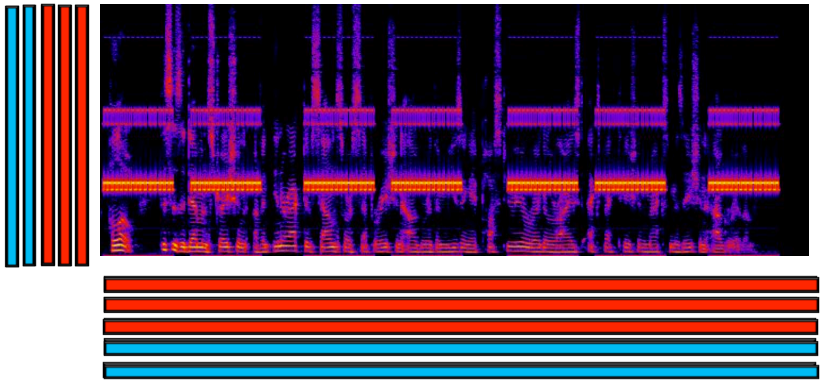
Gains/Weights

- Frequency model
- Weights or gains of frequency

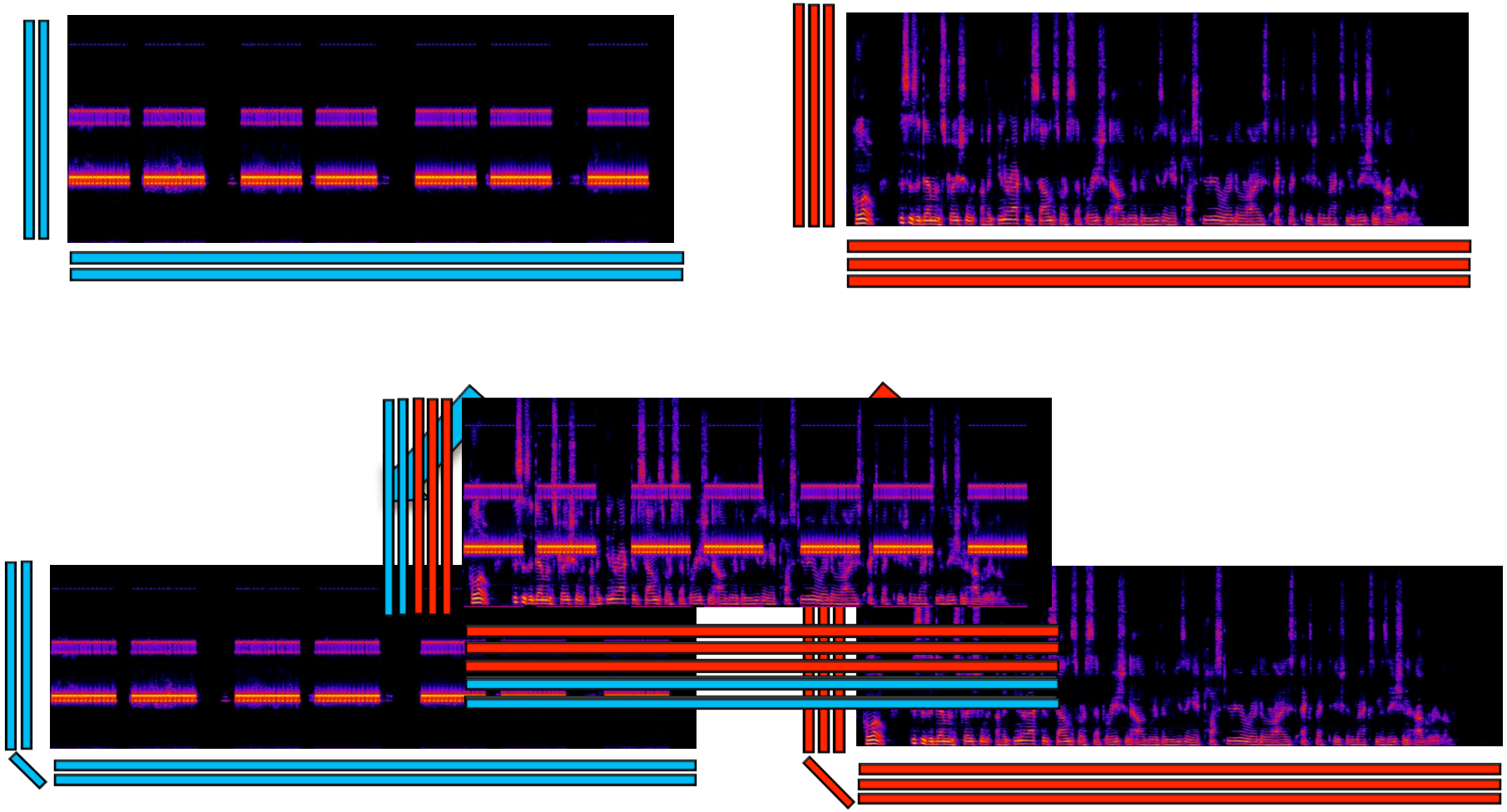
Traditional Approach



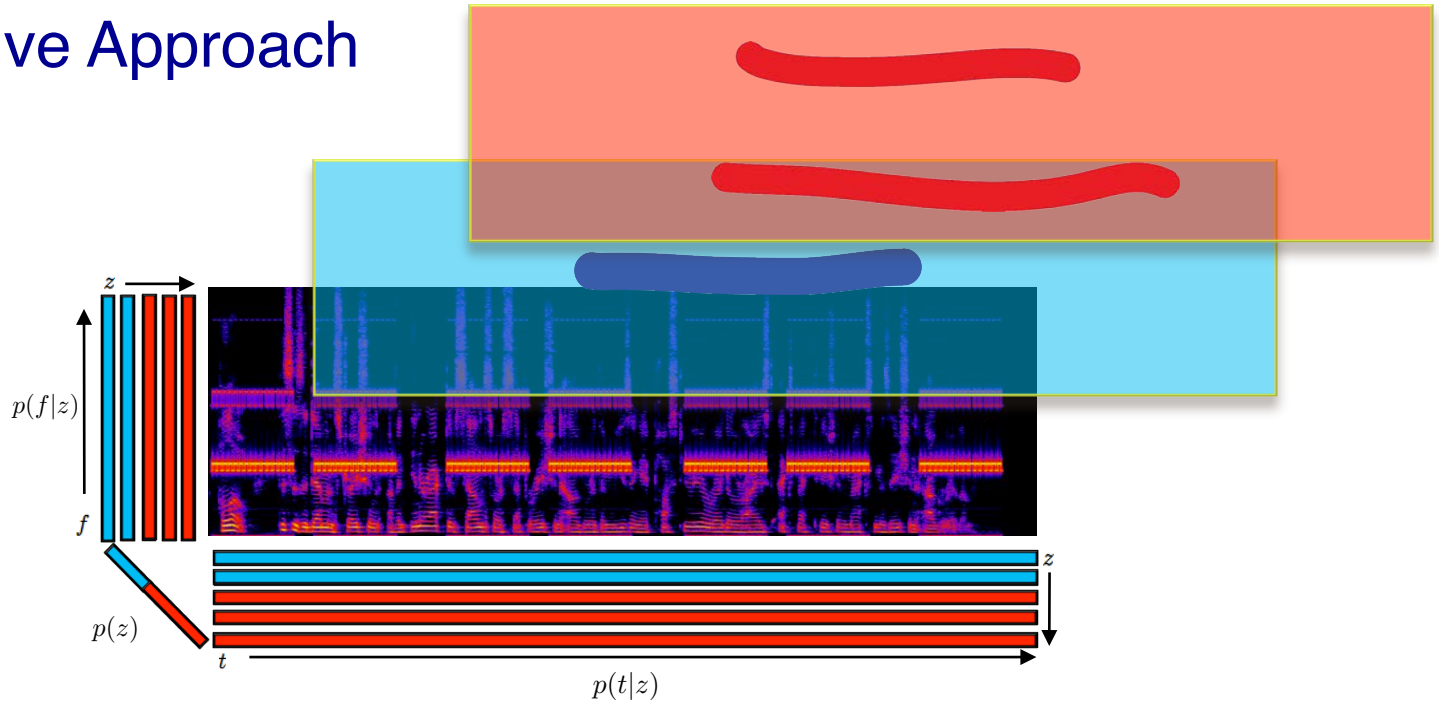


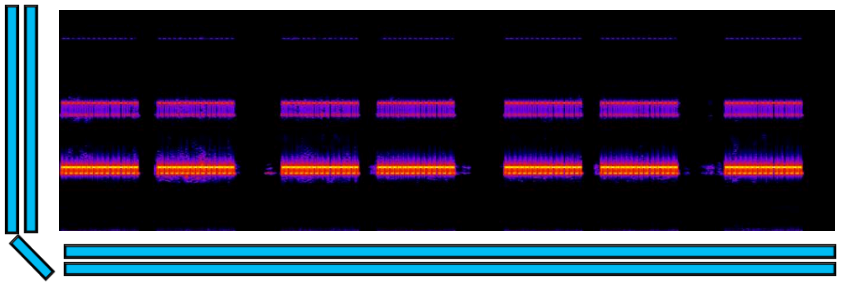


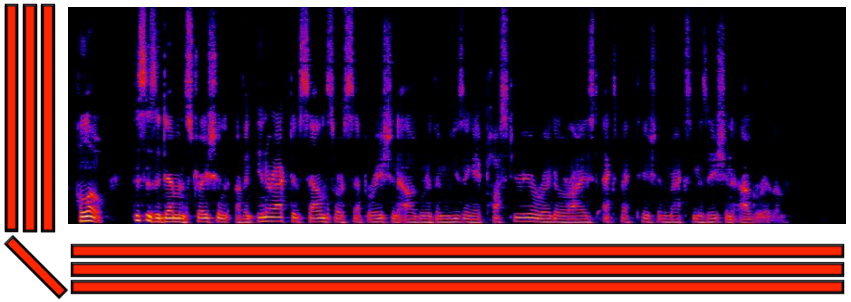
Traditional Approach



Interactive Approach







Overview

- Introduction
- Demonstration
- Software
- Perspective
- Algorithm
- Download!

http://isse.sourceforge.net

Thank you!