INTERACTIVE REFINEMENT OF SUPERVISED AND SEMI-SUPERVISED SOUND SOURCE SEPARATION ESTIMATES

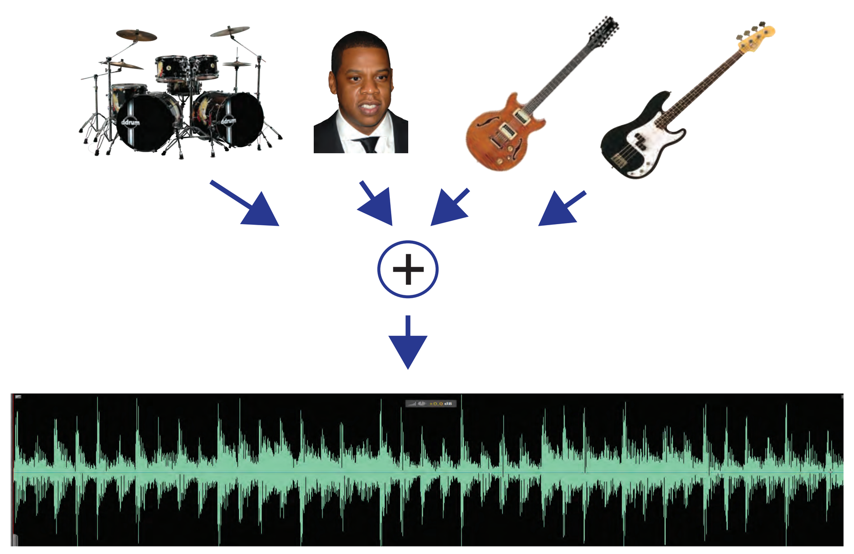
Nicholas J. Bryan[1] & Gautham J. Mysore[2]

[1]Center for Computer Research in Music and Acoustics, Stanford University

[2]Adobe Research

Introduction

- Real world sounds are mixtures of many individual sounds



- It's useful to separate a mixture into its respective sources

music transcription **audio denoising**

audio-based forensics **music remixing**

- Current supervised and semi-supervised NMF-based methods can perform well, but:

  -may also yield poor results

  -are typically a one-shot process

  -have no user-feedback or method of refinement

Idea

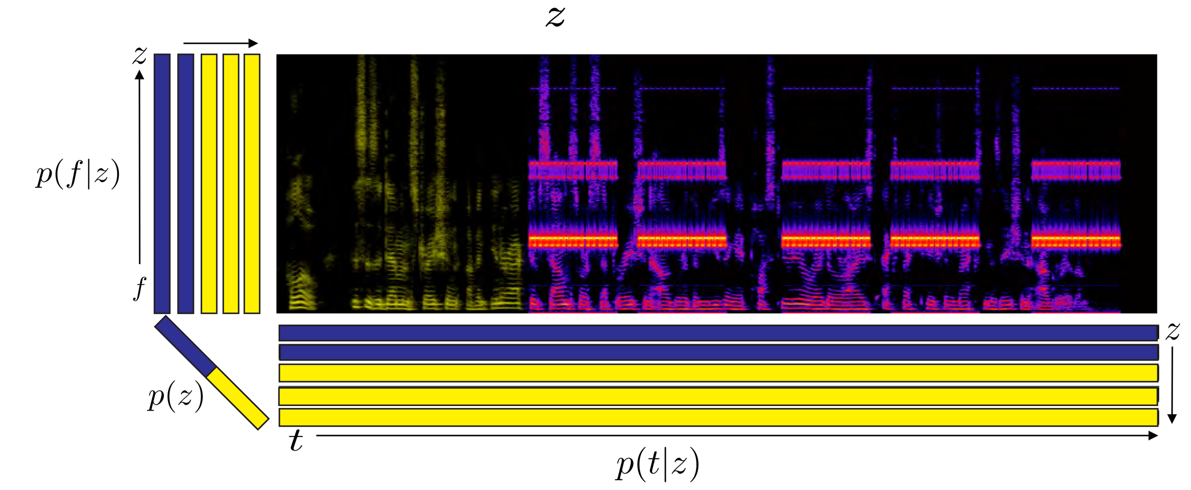
- A layers-sculpting-like environment for audio

- Remove burden of being perfect the first time
- Focus on the professional

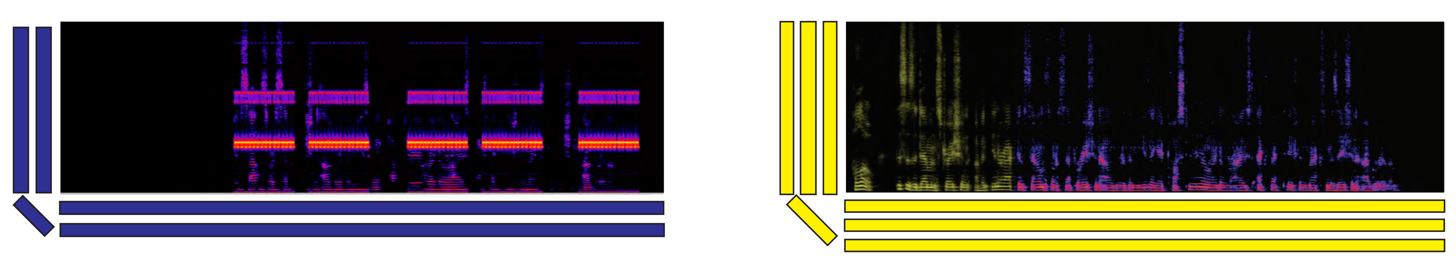
Proposed Method

- Probabilistic latent component analysis model

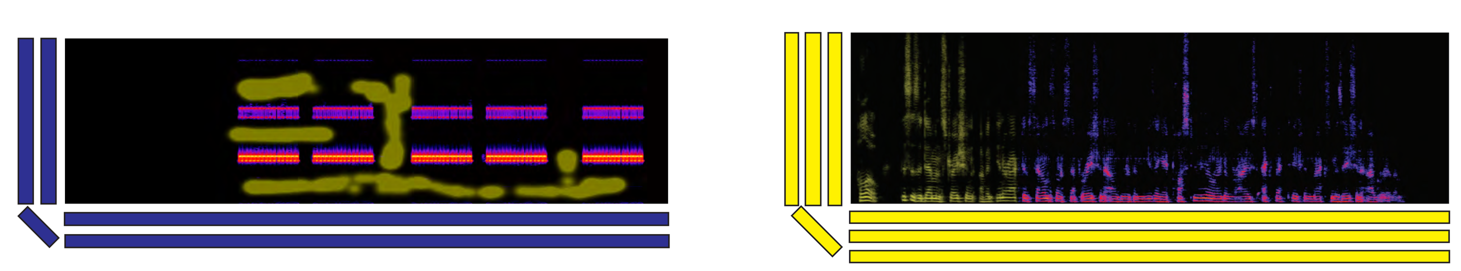
$$p(f,t) = \sum_z p(z)p(f|z)p(t|z)$$



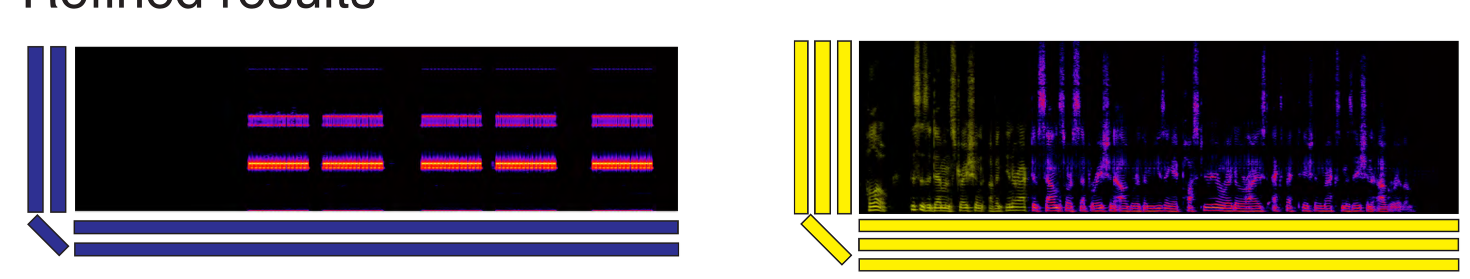
- Initial separation (semi-supervised)



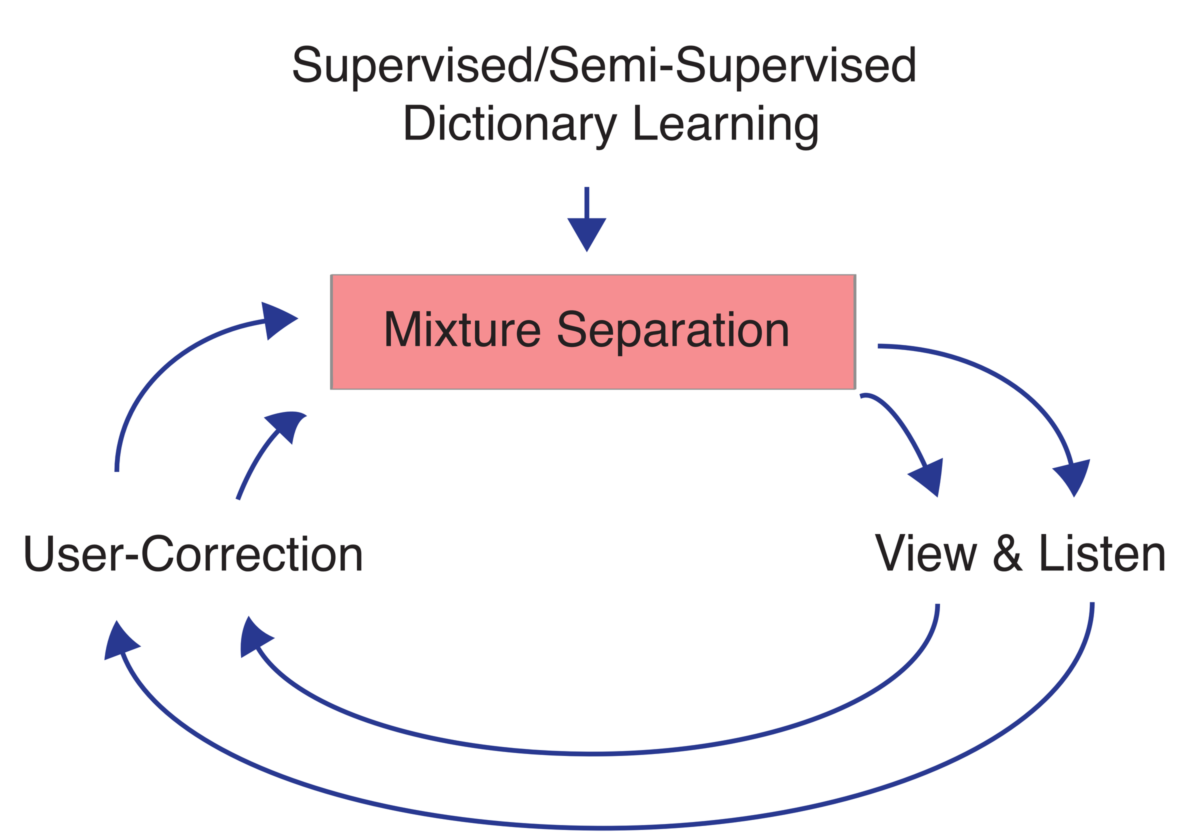
- User correction



- Refined results



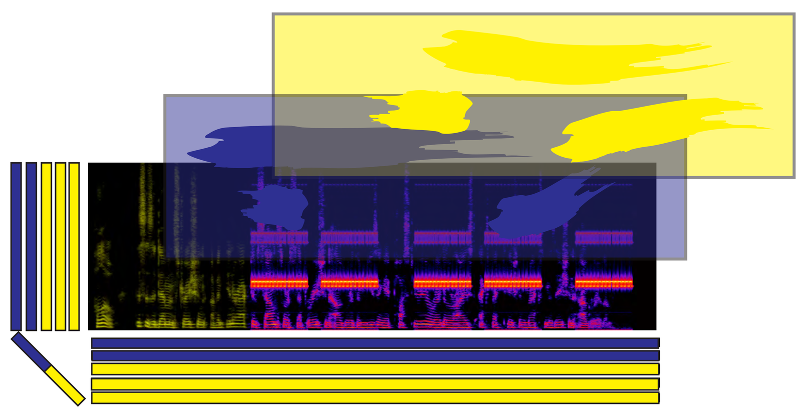
Block Diagram



Posterior Regularization

- Incorporate painting annotations as penalty contraints
- Difficult to encode time-frequency-source constraints via priors

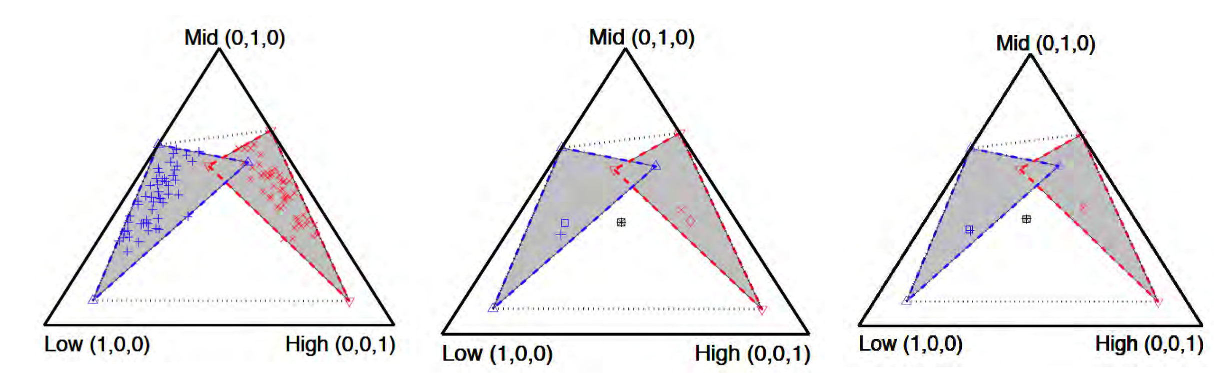
$$p(f,t) = \sum_z \tilde{p}(z)\tilde{p}(f|z)\tilde{p}(t|z)$$



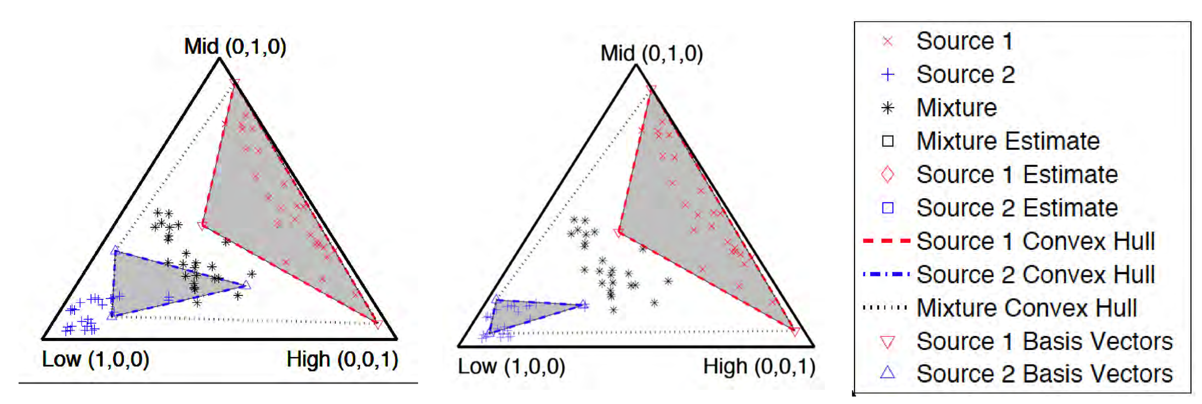
- Use framework of posterior regularization for EM algorithms
- Applies contraints on the posterior (E step) that would otherwise not be possible via stadard priors (M step)

Geometric Perspective

Supervised Separation



Semi-Supervised Separation



Conclusions

- Source separation algorithm that allows:

  -time-frequency constraints via posterior regularization

  -interactive refinement

  -improved results over baseline methods

- For audio and video demonstrations, please see https://ccrma.stanford.edu/~njb/research/iss