An Efficient Posterior Regularized Latent Variable Model for Interactive Sound Source Separation

Nicholas J. Bryan & Gautham J. Mysore

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 non-negative matrix factorization and related probabilistic models methods can perform well, but:
  - require training data
  - may also yield poor results
  - are typically a one-shot process w/no user-feedback

Analogy

- A layers-sculpting-like environment for audio
- Remove burden of being perfect the first time and interact w/algorithm

Proposed Method

- Probabilistic model of audio spectrogram data
  \[ P(f, t) = \sum_z P(z)P(f|z)P(t|z) \]
- Interactively constrain/regularize the model via painting annotations
- Parameter estimation via expectation-maximization
- No explicit training data needed

Posterior Regularization

- Incorporate painting annotations as penalty constraints
- Difficult to encode time-frequency-source constraints via priors
- Use framework of posterior regularization for EM algorithms
- Constraints on the posterior (E step) as oppose to standard priors (M step)
  \[ Q^{n+1} = \arg\min_{Q}\sum_{f} K L(Q|Q_{n}) \]
  \[ Q^{n+1} = \arg\min_{Q}\sum_{t} K L(Q|Q_{n}) \]
- Map painting annotations to linear grouping expectation constraints
- Within a single E step, solve for each time-frequency point:
  \[ \lambda = [\alpha, \beta, \gamma, \delta] \]

Evaluation

- Use SDR, SAR, SIR evaluation metrics for comparison
- Tested on a variety of sounds: cell phone + speech (C), drum + bass (D), orchestra + cough (O), piano + wrong note (P), siren + speech (S), vocals + background music (S1, S2, S3, S4)
- On the examples tested, the proposed method outperformed prior work
  - Relatively insensitive to the number of latent components (if large enough)

Conclusions

- Source separation algorithm that allows:
  - time-frequency constraints via posterior regularization
  - efficient, interactive algorithm
  - improved results over prior work
- For audio and video demonstrations, please see https://ccrma.stanford.edu/~njb/research/iss