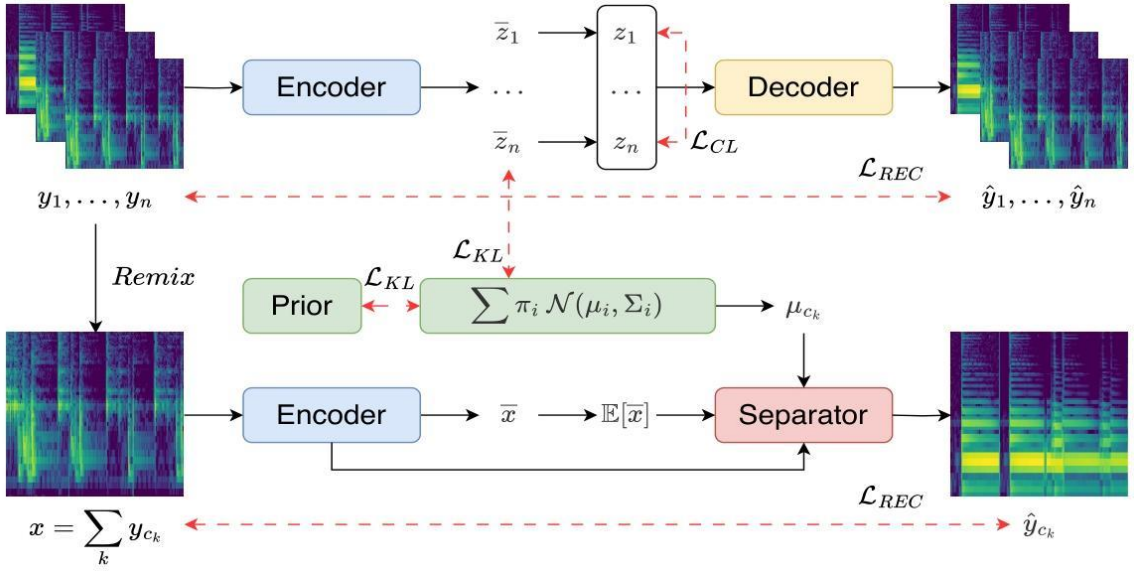
A variational stochastic dirichlet process-based autoencoder model for fine-grained music source separation

Yin ZHU, Jingqi LI, Cong JIN, Qiuqiang KONG, Hongming SHAN, Junping ZHANG

Frontiers of Computer Science, DOI: 10.1007/s11704-025-51043-2

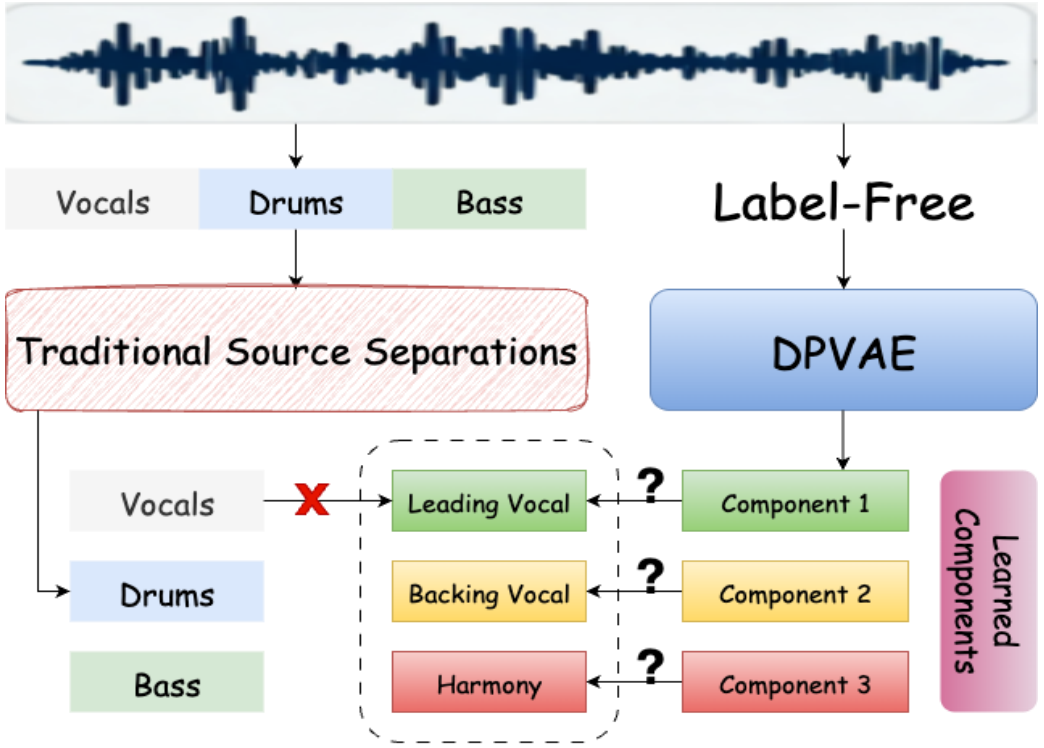
Problems & Ideas

- Music Source Separation (MSS) decomposes mixed audio into constituent sources. Existing methods rely on coarse annotations, while fine-grained annotations need massive resources and lack standards making precise labeling difficult.
- Ideas: We integrate separation with clustering and use the Dirichlet Process to automatically learn the optimal number of source categories from data, avoiding predefining source counts.



Main Contributions

- Contributions:
  - Propose DPVAE, the first variational stochastic process-based MSS framework that eliminates the need for pre-specified component counts.
  - Develop a systematic two-stage training framework (latent space partitioning + component separation) to balance discrete clustering and continuous reconstruction in generative models.
  - Verify the framework's efficacy via synthetic data, hybrid networks, and benchmark comparisons, while clarifying scalability advantages and addressing fine-grained separation challenges.



DPVAE: Automation; Fine-Grained; Scalable