Audio signal interpolation using optimal transportation of spectrograms

David Valdivia*, Marien Renaud†, Elsa Cazelles* and Cédric Févotte*

*IRIT, CNRS, Université de Toulouse
†IMB, Université de Bordeaux

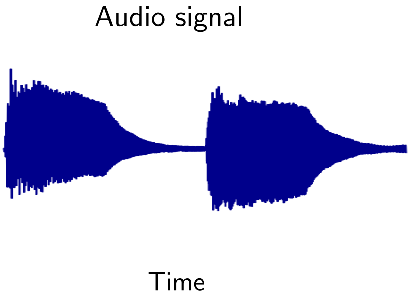
GdR IASIS
February 17, 2025

Problem setting

- **Problem**: Given two audio signals $\mathbf{y}^s$ (source) and $\mathbf{y}^t$ (target), generate an interpolant $\mathbf{y}^\alpha$ with $\alpha \in [0, 1]$, using optimal transport (OT).

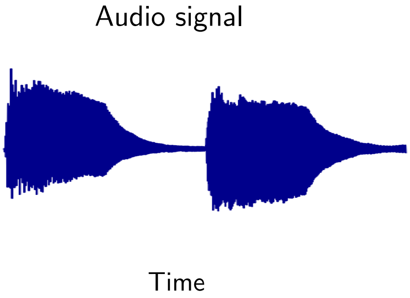
Problem setting

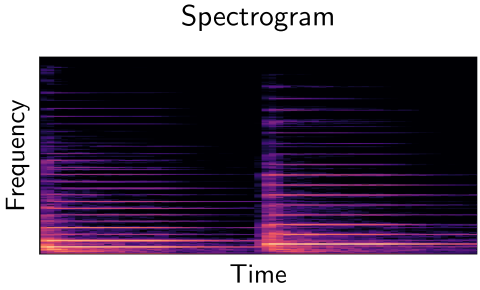
- **Problem**: Given two audio signals $\mathbf{y}^s$ (source) and $\mathbf{y}^t$ (target), generate an interpolant $\mathbf{y}^\alpha$ with $\alpha \in [0, 1]$, using optimal transport (OT).



Problem setting

- **Problem**: Given two audio signals $\mathbf{y}^s$ (source) and $\mathbf{y}^t$ (target), generate an interpolant $\mathbf{y}^\alpha$ with $\alpha \in [0, 1]$, using optimal transport (OT).





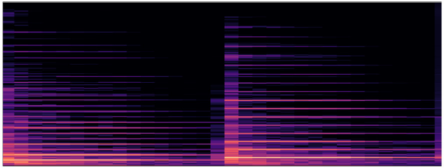
Problem setting

- **Problem**: Given two audio signals $\mathbf{y}^s$ (source) and $\mathbf{y}^t$ (target), generate an interpolant $\mathbf{y}^\alpha$ with $\alpha \in [0, 1]$, using optimal transport (OT).
- **Idea**: Treat the normalized spectrogram $\mathbf{X} \in \mathbb{R}_+^{M \times N}$ as a **discrete probability distribution** μ:

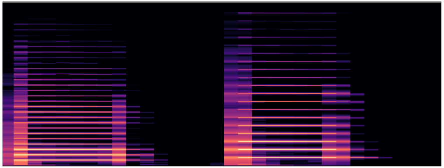
$$\mu := \sum_{m=1}^{M} \sum_{n=1}^{N} X_{mn} \delta_{(f_m, t_n)}.$$

Baseline approach with OT

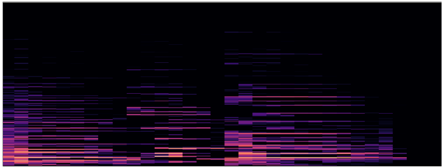
Piano (source)



Guitar (target)

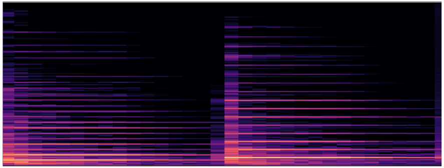


Wasserstein barycenter

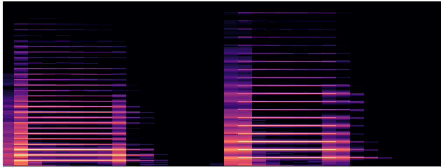


Limit temporal transportation with structured cost matrix and UOT

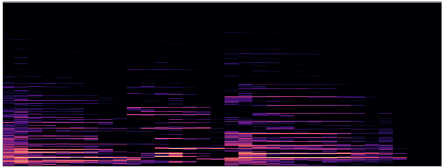
Piano (source)



Guitar (target)



Wasserstein barycenter



Structured cost matrix and UOT