Audio Optimal Transport: A Generalized Portamento

Constant Bourdrez

IASD Master

PSL RESEARCH UNIVERSITY PARIS

January 2025

Table of Contents

Problem setup

Problem Setup

Goal:

- Develop smooth portamento [1] effects **(Audio Example)**
- Apply techniques to various types of audio or signals
- Enable mass transfer between spectrograms derived from STFT
- Reconstruct the signal interpolated into the time domain

Continuous and discrete Short-Time Fourier Transform (STFT)

$$X(\omega, \tau) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-j\omega t} \, dt \tag{1}$$

$$X[m, k] = \sum_{n=-\infty}^{\infty} x[n] w[n-m] e^{-j2\pi kn/N} \tag{2}$$

Audio Transport

Audio Transport

Note

Audio is sampled at **44.1 kHz**, making supports identical. One frame of spectrograms are treated as **1D vectors**: X and Y

Optimal Transport Plan [2]

Discrete Optimal Transport between two spectrograms:

$$\pi^* = \underset{\pi \geq 0}{\arg\min} \sum_{i,j} \left| \omega_i^X - \omega_j^Y \right|^2 \pi_{ij}, \tag{3}$$

Conservation of mass constraints:

$$\sum_j \pi_{ij} = |X_i|, \quad \sum_i \pi_{ij} = |Y_j|, \tag{4}$$

Unbalanced Formulation

Unbalanced Transport Plan [3]

Allow mass destruction or construction using a penalisation term (Cizarr f-divergence):

$$W_\rho(a, b) = \min_{P \in \mathbb{R}_+^{n \times m}} \langle P, C \rangle + \rho D_\varphi(Pu \mid X) + \rho D_\varphi(P^\top v \mid Y), \quad (5)$$

Kullback-Leibler divergence:

$$D_\varphi(Pu \mid X) = D_{\mathrm{KL}}(P_u \| X) = \sum_i P_u(i) \log \frac{P_u(i)}{X(i)}$$

$$D_\varphi(P^\top v \mid Y) = D_{\mathrm{KL}}(P^\top v \| Y) = \sum_i P(i) \log \frac{P(i) v_i}{Y_i}$$

Going back to the time domain

Use the transport plan π_*:

$$X_{\text{interp}}\left[(1-k)\omega_i^X + k\omega_j^Y\right] = \pi_*[i,j] \tag{6}$$

To be able to go back into the time domain, we need the signal to be coherent. We compute the phase of frequency i at interpolation time t $\phi_i(t)$ using the phase vocoder paradigm [4]:

$$\begin{aligned} \phi_i(t) &= 2\pi f_i \cdot t + \phi_{i-1}(t), \\ \phi_{i,\text{interp}}(t) &= A_i(t) \cdot \sin(\phi_i(t)), \end{aligned} \tag{7}$$

Then, one juste needs to take the reverse STFT.

- We used the algorithm introduced in [1] and a CVXPY implementation using linear programing for both balanced and unbalanced transport

Experiments

Experiments

Monophonic signals:

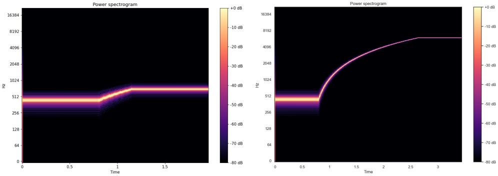


Figure: 440Hz and 660Hz (Left), 440Hz and 6000Hz (Right)

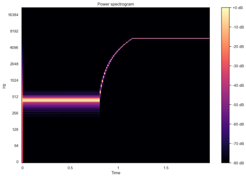
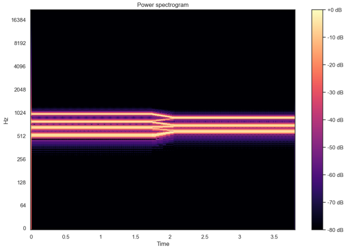


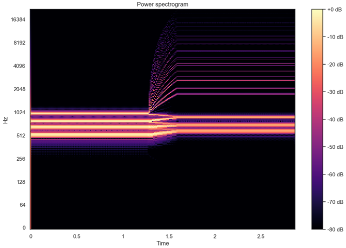
Figure: 440Hz and 6000Hz with $t_{\text{transition}} \leq \frac{1}{\Delta f}$

Experiments

Polyphonic Signals:



Rectangle and Triangle Waves:



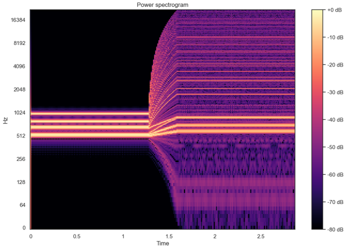


Figure: (Top Left) Transition from sine waves at 523, 659, 784, 988 Hz to 587, 698, 880 Hz.
(Top Right) Sine wave at 523, 659, 784, 988 Hz to triangle wave at 587, 698, 880 Hz.
(Bottom Right) Sine wave at 523, 659, 784, 988 Hz to rectangle wave at 587, 698, 880 Hz.

Experiments

Natural Sounds:

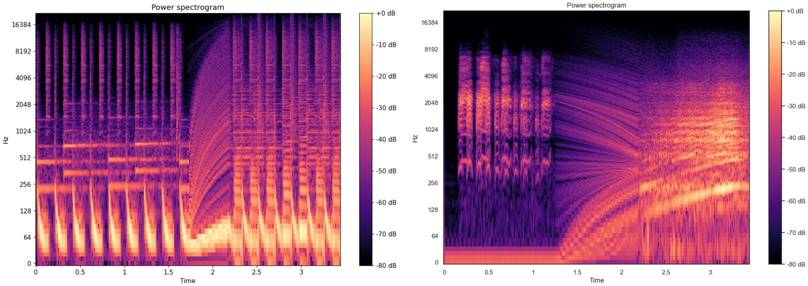


Figure: Left: Transition between two electronic tracks, Right: Transition between duck and car sounds.

Experiments

Unbalanced Transport:

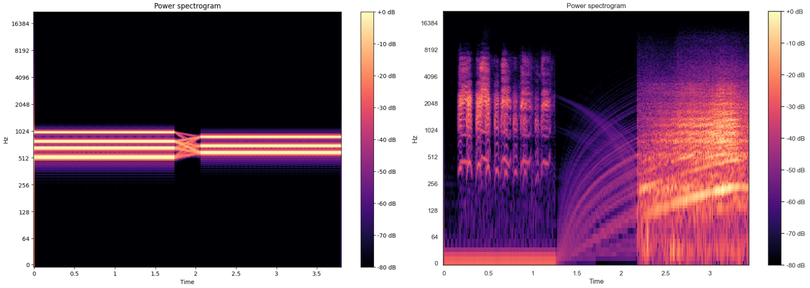


Figure: Left: Unbalanced transport on **polyphonic** sine waves (523, 659, 784, 988 Hz to 587, 698, 880 Hz , Right: Unbalanced transport between a **duck** and a car sound.

Conclusion

Conclusion

- **Audio Transport Effect**: Facilitates portamento-like transitions between any two audio signals.
- **Single Control Parameter**: Easy for musicians to use in both live performances and studio settings.
- **Flexibility**: Implemented in Python, with potential for hardware integration in low-level languages.
- **Future work**:
  - Glide effect creation using a single audio source with dual inputs.
  - Addressing incoherence artifacts through unbalanced transport to solve mass destruction issues.
  - Simultaneous interpolation of multiple signals via Wasserstein barycenters.
- **Creative Potential**: Designed for artistic exploration, offering exciting opportunities for musicians and producers.

Thank you !

constant.bourdrez@dauphine.eu

References

[1] Trevor Henderson and Justin Solomon. *Audio Transport: A Generalized Portamento via Optimal Transport*. 2019. arXiv: 1906.06763 [eess.AS].

[2] Gabriel Peyré and Marco Cuturi. "Computational Optimal Transport: With Applications to Data Science". In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607. ISSN: 1935-8237. DOI: 10.1561/2200000073.

[3] Thibault Séjourné, Gabriel Peyré and François-Xavier Vialard. "Chapter 12 - Unbalanced Optimal Transport, from theory to numerics". In: *Numerical Control: Part B*. Ed. by Emmanuel Trélat and Enrique Zuazua. Vol. 24. Handbook of Numerical Analysis. Elsevier, 2023, pp. 407–471. DOI: https://doi.org/10.1016/bs.hna.2022.11.003.

[4] J. L. Flanagan and R. M. Golden. "Phase Vocoder". In: *Bell System Technical Journal* 45.9 (1966), pp. 1493–1509.