Audio Optimal Transport: A Generalized Portamento

Constant Bourdrez
Department of Computer Science
PSL University - IASD Master
constant.bourdrez@dauphine.eu

Abstract

This study presents a generalized framework for audio transport, specifically addressing smooth transitions, or portamento, between audio signals such as musical notes. The relevance of this work lies in its potential to enable expressive and musically meaningful transformations, with applications in sound design, music synthesis, and AI-driven music composition. By leveraging the theory of optimal transport (OT), we provide a mathematically principled approach to modeling transitions that respect the intrinsic properties of audio signals. Building upon the foundational method proposed by Henderson [1], this study not only reproduces the original results but also extends the framework to handle unbalanced transport and interpolate between more than two signals.

Our contributions include a theoretical framework for audio transport, novel numerical algorithms for solving transport problems, and practical demonstrations of the method's effectiveness for expressive and stylistic audio transformations. We also explore real-time implementations and highlight the strengths and limitations of this approach, particularly in applications involving melodic audio signals.

1 Introduction

Optimal Transport (OT) theory has gained prominence as a powerful tool for solving problems involving the transformation of one probability distribution into another. With its roots in mathematics, OT has found applications in diverse fields such as economics, computer graphics, and machine learning. This study extends the application of OT to audio processing, focusing on the creation of smooth transitions, or portamento, between audio signals like musical notes.

The foundational work by Peyré and Cuturi [2] offers a comprehensive overview of computational OT, laying the groundwork for its adoption in handling complex data distributions. Building on this foundation, we investigate how OT can be employed to achieve musically meaningful and expressive transformations in audio, pushing the boundaries of what is possible in sound design and synthesis.

1.1 Prior Work

Prior studies have demonstrated the versatility of OT in signal processing. Elvander [3] employed OT to estimate inharmonic pitch signals, effectively clustering frequencies even under significant inharmonicity. Flamary [4] applied OT to music transcription, showcasing its ability to handle the intricate spectral characteristics of musical signals. Montesuma [5] demonstrated the potential of OT for domain adaptation in music-related tasks.

Expanding its scope, Rolet et al. [6] combined OT with Non-Negative Matrix Factorization for blind source separation, while Torres et al. [7] used it for unsupervised harmonic parameter estimation. Chazelles et al. [8] applied OT to define distances between power spectral densities of time series, broadening its application to a more generalized signal processing context.

However, transition-based applications of OT in audio remain underexplored. Henderson [1] was among the first to address transitions between signals as an OT problem. Related approaches include using Generative Adversarial Networks for smooth transitions in DJ sets [9], diffusion models for conditioned music generation [10], and style transfer with Variational Autoencoders [11].

1.2 Motivations and Challenges

A portamento represents the continuous transition between pitches, a feature available on only a few musical instruments. Electronic devices, on the other hand, are often limited to specific scenarios such as monophonic glide or offline processing. Henderson's work demonstrated that such an effect could be modeled as an OT problem, offering a novel perspective on audio transitions. Building on this foundation, we propose enhancements to refine this effect by addressing key limitations.

Puche et al. [12] introduced a framework for timbre and pitch interpolation using Variational Autoencoders (VAEs), interpolating signals in the latent space to generate the final data. While this approach achieves reasonable results, its reliance on latent space interpolation often neglects the geometric structure of spectrogram distributions, leading to artifacts or spectral inconsistencies. This highlights the need for a transport approach that ensures both smoothness and optimality.

Our proposed audio transport effect relies on solving a one-dimensional OT problem, mapping pitches in one signal to those in another. The approach is computationally efficient, enabling real-time implementation in a low-resource programming language. It performs best with audio signals that are predominantly melodic and minimally affected by additional effects. However, rhythmic components often lose relevance in this framework, suggesting areas for further refinement.

2 Optimal Transport Overview

2.1 Background

Optimal transport (OT, [2]) compares two probability distributions paying special attention to the geometry of their domain. This comparison is achieved by finding the lowest cost to transfer the mass from one probability distribution onto the other. Specifically, the optimal transport between two measures μ and ν defined on $\mathbb{R}^d$ is given by:

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) \mathrm{d}\pi(x,y), \tag{1}$$

where π is a joint distribution for x and y, referred to as the transport plan, belonging to the space $\Pi(\mu,\nu)$ of the product measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν; and $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a cost function. In this manuscript, we seek the optimal plan π_* obtained by setting $c(x,y) = \|x-y\|^p$, $p \geq 1$, in eq. (1).

$$\pi^* = \arg\min_{\pi} \iint_{\mathbb{R}^2} \|x-y\|^p \, d\pi(x,y), \tag{2}$$

This quantity is subject to nonnegativity as well as conservation of mass for source and target distributions ρ_v and ρ_w:

$$\int_{\mathbb{R}} \pi(x,y)\, dy = \rho_v(x) \quad \text{and} \quad \int_{\mathbb{R}} \pi(x,y)\, dx = \rho_w(y). \tag{3}$$

The p-th root of the optimal value provides an intuitive way to measure the similarity between two distributions known as the p-Wasserstein distance or also Earth Mover's distance. In the rest of this paper, we will use $p = 2$. The corresponding least squares Wasserstein distance satisfies all metric axioms among other attractive properties [13, 14].

The optimal transport plan can be utilized to perform displacement interpolation between two distributions. This process visualizes the movement of mass by interpolating along the optimal transport paths, effectively sliding each particle of mass from its source to its target. This method is widely employed in applications such as computer graphics and related fields. Figure 2 illustrates two distinct approaches to interpolating between distributions. The top panel showcases linear interpolation, which, if viewed as audio spectra, corresponds to fading one set of pitches out while another fades in. In contrast, the bottom panel demonstrates displacement interpolation, where the mass shifts smoothly from one location to another. For audio spectra, this movement would resemble the effect of a portamento, creating a sliding transition between pitches.

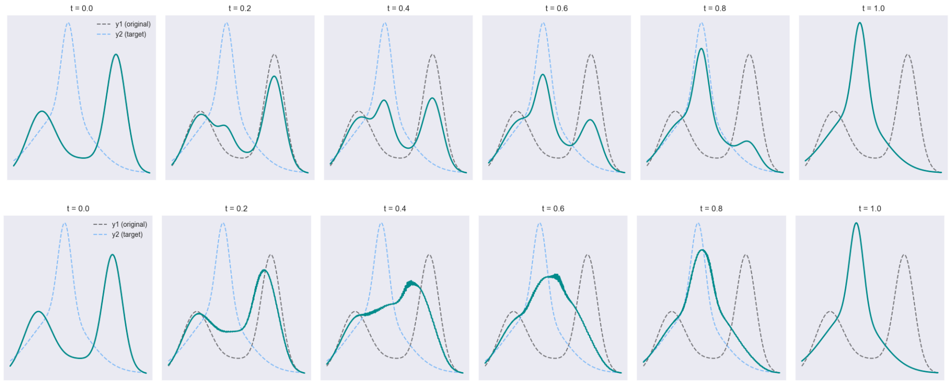

Figure 1: Top: Regular interpolation between two mixtures of Gaussians. Bottom: Optimal transport interpolation between two mixtures of Gaussians.

A notable limitation of these techniques is the computational complexity of solving the optimal transport problem for dimensions $d \geq 2$. However, for $d = 1$, efficient algorithms exist that operate in linear time.

2.2 Audio Transport

The audio transport effect operates by applying displacement interpolation to input audio spectra, enabling a continuous transition of pitches from one signal to another as the interpolation parameter varies. To adapt the spectra dynamically over time, the algorithm follows the phase vocoder paradigm.

2.2.1 Phase Vocoder Paradigm

The phase vocoder [15] is a time-frequency processing method widely used in audio signal manipulation. It analyzes the signal's spectrogram to extract both amplitude and phase information. This information is then modified and re-synthesized to achieve time-stretching, pitch-shifting, or other spectral transformations. Extensions of this paradigm [16] enhance its ability to handle complex modifications.

In the context of audio transport, the phase vocoder's principles ensure that spectral changes remain perceptually smooth, particularly when transitioning between distinct spectral configurations. By leveraging this paradigm, the algorithm achieves a natural and continuous interpolation effect.

This method produces two well-known artifacts: vertical incoherence and horizontal incoherence. Solutions to these two phenomena are described in [1], but our work didn't focus on resolving this.

2.2.2 Transport Between Two Spectrograms

Consider discrete spectra represented as complex vectors X and Y, with their corresponding frequency vectors denoted as ω^X and ω^Y, respectively. To compute the optimal transport plan between these discrete spectra, we formulate a discrete analog of the continuous optimal transport problem described in equation (2).

The goal is to determine the optimal transport plan $\pi^* \in \mathbb{R}^{|X|\times|Y|}$ that minimizes the following cost function:

$$\pi^* = \underset{\pi \geq 0}{\arg\min} \sum_{i,j} \left|\omega_i^X - \omega_j^Y\right|^2 \pi_{ij}, \tag{4}$$

where π_{ij} represents the amount of mass transported from ω_i^X to ω_j^Y, and $\left|\omega_i^X - \omega_j^Y\right|^2$ is the squared cost of transporting this mass.

This optimization is subject to the conservation of mass constraints:

$$\sum_j \pi_{ij} = |X_i|, \quad \sum_i \pi_{ij} = |Y_j|, \tag{5}$$

which ensure that the total mass leaving each frequency bin ω_i^X equals the magnitude $|X_i|$, and the total mass arriving at each frequency bin ω_j^Y equals $|Y_j|$. Additionally, the problem assumes that the total mass in X and Y is equal, i.e., $\sum_i |X_i| = \sum_j |Y_j|$. For spectra with different total magnitudes, the optimal transport plan can be computed on normalized spectra, and the scaling can then be linearly interpolated over the interpolation process.

After computing the optimal transport plan π^*, the spectra can be interpolated using a parameter $k \in [0, 1]$. For each transported mass π_{ij}^*, the corresponding frequency is placed at a linearly interpolated location:

$$(1-k)\omega_i^X + k\omega_j^Y \tag{5}$$

If multiple masses are displaced to the same frequency bin, their magnitudes are summed. This approach ensures a smooth transition of spectral components between the two spectra, with the interpolation parameter k controlling the progression from X ($k = 0$) to Y ($k = 1$). The phase attribution is detailled in 2.3.

Our work presents a novel algorithm for computing the optimal solution in linear time using CVXPY and compares it with Henderson's algorithm, which has a complexity of $O(|X| + |Y|)$.

Algorithm 1 Computing The Optimal Transport Matrix according to [1], π^*

```
$\pi^*_{i,j} \leftarrow 0$
$\rho_X, \rho_Y \leftarrow |X_0|, |Y_0|$                    ▷ $\rho$ is the mass left in a bin

loop
    if $\rho_X < \rho_Y$ then
        $\pi^*_{ij} \leftarrow \rho_X$                       ▷ Assign as much mass as possible

        $i \leftarrow i + 1$                                 ▷ Refill the emptied bin
        if $i \geq |X|$ then break
        $\rho_X \leftarrow |X_i|$

        $\rho_Y \leftarrow \rho_Y - \rho_X$                  ▷ Decrease the capacity of the other

    else
        Symmetric to the case above

return $\pi^*$
```

In one dimension, the optimal transport plan is monotonic, meaning that no mass crosses over any other mass. This property allows Equations 4 and 5 to be solved using the greedy strategy outlined in Algorithm 1.

The algorithm starts with the initial bins of the two spectra. Given the monotonicity, all of the mass in the smaller bin must be assigned to the larger bin. Once this assignment is made, we can conceptually remove the smaller bin and reduce the mass of the larger bin by the amount assigned. The algorithm then proceeds inductively, solving the smaller subproblem.

This algorithm is tested and compared with another approach that utilizes linear programming and the OSQP Solver, which is based on a quadratic programming formulation to efficiently solve optimal transport problems in linear time for one-dimensional distributions. However, it should be noted that for 1D problems, such a solver may not be required, as Algorithm 1 remains efficient in this context.

Algorithm 2 Optimal Transport using CVXPY (with OSQP Solver)

```
Input: Two spectras $X$ and $Y$ with dimensions $n$ and $m$ and their support $x$ and $y$, respectively
Output: Optimal transport matrix $P$
Step 1: Compute the distance matrix $C_{i,j} := \|x_i - y_j\|_2$
Step 2: Define the optimization variable $P \in \mathbb{R}^{n \times m}$
Step 3: Define the auxiliary variables $u = \mathbf{1}_{m \times 1}$ and $v = \mathbf{1}_{n \times 1}$
Step 4: Set up the constraints:
    • $0 \leq P$
    • $P \cdot u = X$ (Mass conservation for $X$)
    • $P^T \cdot v = Y$ (Mass conservation for $Y$)
Step 5: Set up the optimization problem:
    $$\text{minimize} \sum_{i,j} P_{i,j} C_{i,j}$$
    with constraints defined in Step 4
Step 7: Solve the optimization problem using OSQP solver
Step 8: Return the optimal transport matrix $\pi_*$
```

The algorithm was implemented in Python using the CVXPY package for convex optimization [17]. Our code is open source and can be found at `https://github.com/constantbourdrez/Optimal_Audio_Transport`.

Once the map is computed, it can be displayed, allowing the visualization of the shortest path between distributions or spectra.

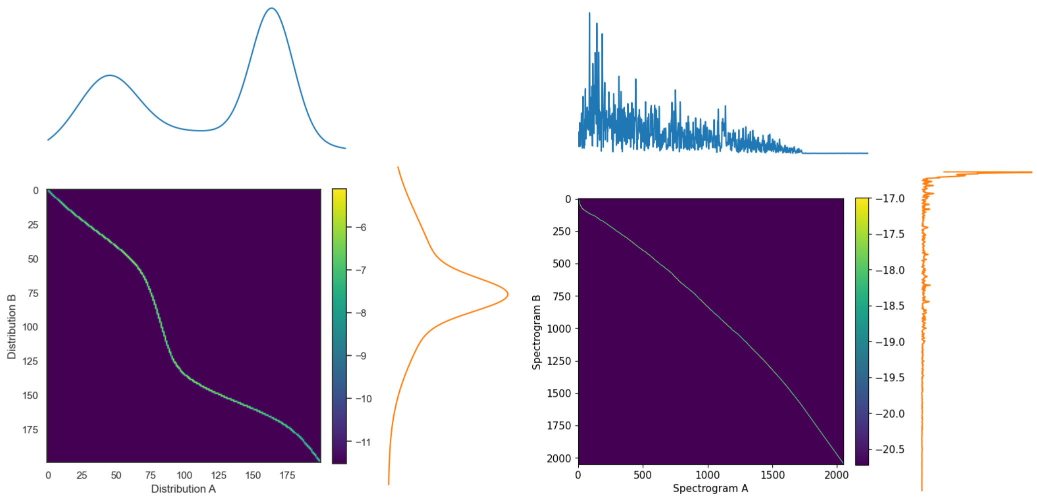

Figure 2: Visualization of the optimal transport map between two mixtures of Gaussians (left) and two spectra (right). Parameters of the two mixtures of Gaussians: $\boldsymbol{\mu}_1 = (0.2, 0.5, 0.75)\,, \boldsymbol{\mu}_2 = (0.4, 0.36)\,, \boldsymbol{\sigma}_1 = (0.1, 0.3, 0.07)\,, \boldsymbol{\sigma}_2 = (0.05, 0.2)\,, \boldsymbol{A_1} = (2, 0.7, 4)\,, \boldsymbol{A}_2 = (1.5, 1)$.

2.3 Return To The Time Domain

After transporting mass across frequencies, the goal is to reintroduce phase information into the signal to be able to go back into the time domain.

To retrieve the phase in the transport process, we use a phase-vocoder approach to ensure smooth transitions between frequency bins. We begin by computing the frequencies corresponding to each bin based on the signal's sampling rate and the FFT size. The phase is then updated by adding a term that reflects the progression of each frequency over time, ensuring a continuous phase evolution. This term is computed by multiplying the frequency with a factor that accounts for time, producing a phase change that is consistent with the signal's progression. The updated phase is then used to compute the interpolated phase, which is combined with the magnitude of the signal to maintain its spectral characteristics. This phase retrieval method ensures that the phase information is preserved and smoothly interpolated during the transport, allowing the signal to be accurately transformed back into the time domain.

$$\begin{aligned} \phi_i(t) &= 2\pi f_i \cdot t + \phi_{i-1}(t), \\ \phi_{i,\text{interp}}(t) &= A_i(t) \cdot \sin(\phi_i(t)), \end{aligned} \tag{6}$$

where $A_i(t)$ is the amplitude of the spectrogram at w_i and interpolation time t. The original signal is reconstructed by applying the inverse Short-Time Fourier Transform (iSTFT) to the processed spectrogram.

2.3.1 Vertical and Horizontal Incoherence

A limitation of the Short-Time Fourier Transform (STFT) is the inherent trade-off between time and frequency resolution. As time resolution increases, the frequency representation becomes "smeared." This smearing affects the relationship between a peak frequency and its components, which is critical for maintaining perceptual quality. Treating these components independently often introduces phasing artifacts, known as vertical incoherence, within a window. To address this, phase vocoder techniques commonly "lock" regions around a peak frequency, ensuring that the relative phase between bins within these regions remains unchanged.

Another issue arises when transposing a specific spectral region, as its phase evolves at a different rate. This leads to interference when transferring the phases of consecutive windows from the original signal to the corresponding windows in the transposed signal, a phenomenon known as horizontal incoherence.

Both of these limitations were discussed in [1], and incorporating solutions for them into our framework is a direction for future work.

2.4 Unbalanced Audio Optimal Transport

The main idea of unbalanced optimal transport is to lift the mass conservation restriction by replacing the hard constraint encoded in equation (5) by a soft penalization as described by Sejourné et al. [18] The previous constraint in optimal transport requires the transport of all samples from the input distributions, which can be problematic in the presence of outliers, or irrelevant samples. In contrast, relaxing this constraint, as done in unbalanced optimal transport (defined below), allows for the exclusion of such outliers. In the context of audio signals, this technique is particularly useful because the transport problem is defined in the Fourier space, and numerical artifacts introduced by the Short-Time Fourier Transform (STFT) can lead to unwanted elements that should not be considered in the transport process.

The unbalanced optimal transport problem can be formulated using a Cizarr f-divergence as follows:

$$W_\rho(a,b) = \min_{P \in \mathbb{R}_+^{n\times m}} \langle P, C\rangle + \rho D_\varphi(Pu \mid X) + \rho D_\varphi(P^\top v \mid Y), \tag{7}$$

where D_φ represents a Cizarr f-divergence, defined as:

$$D_\varphi(h \mid b) = \sum_i \varphi\left(\frac{h_i}{b_i}\right) b_i.$$

A well-known example of an f-divergence is the Kullback-Leibler (KL) divergence, which arises when $\varphi(s) = s\log(s) - s + 1$. For the remainder of the article, we will use this specific form of f-divergence. The parameter ρ controls the amount of mass conservation relaxation. As $\rho \to +\infty$, one recovers the usual (balanced) OT. When $\rho \to 0$, no transport is performed.

In this setup, the problem is referred to as either "Kantorovitch-Hellinger" or "Wasserstein-Fisher-Rao".

Given the fact that $n = m$ and that the sampling points are equal, $x_i = y_i$, it holds that

$$\frac{W_\rho(X,Y)}{\rho} \to \sum_i \left(\sqrt{X_i} - \sqrt{Y_i}\right)^2 \quad \text{as} \quad \rho \to +\infty.$$

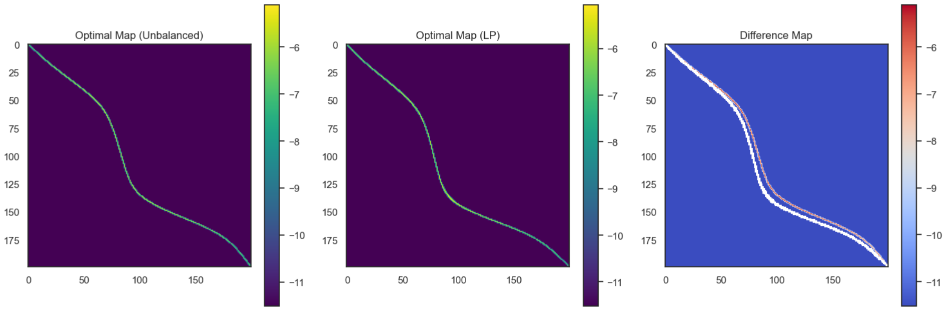

Figure 3: Comparison of the maps obtained with unbalanced (right), and balanced (left) transport on the two mixture of gaussians from fig. 2 with $\rho = 0.7$ (logscale)

Figure (3) shows that the transport plans computed are different from one another, highlighting the impact of the unbalanced and balanced optimal transport formulations. In the balanced case, the mass is strictly conserved between the two distributions, leading to a more rigid transport plan. In contrast, the unbalanced optimal transport formulation allows for the relaxation of mass conservation, enabling a more flexible transport plan that can handle outliers or irrelevant samples more effectively.

We also used CVXPY for the implementation and the algorithm is described just below:

Algorithm 3 Unbalanced Optimal Transport using CVXPY (with OSQP Solver)

```
Input: Two spectras $X$ and $Y$ with dimensions $n$ and $m$ and their support $x$ and $y$, respectively
Output: Unbalanced optimal transport matrix $P$
Step 1: Compute the distance matrix $C_{i,j} := \|x_i - y_j\|_2$
Step 2: Define the optimization variable $P \in \mathbb{R}^{n \times m}$
Step 3: Define the auxiliary variables $u = \mathbf{1}_{m\times 1}$ and $v = \mathbf{1}_{n\times 1}$
Step 4: Define the penalization terms:
    • $q = \sum_{i,j} \text{KL}\left(P_{ij} v_j \mid X_i\right)$
    • $r = \sum_{i,j} \text{KL}\left(P_{ji}^T u_i \mid Y_j\right)$
Step 5: Set up the constraint:
    • $0 \leq P$
Step 5: Set up the optimization problem:
    $$\text{minimize} \sum_{i,j} P_{i,j} C_{i,j} + \rho \times q + \rho \times r$$
    with constraints defined in Step 5
Step 7: Solve the optimization problem using OSQP solver
Step 8: Return the optimal transport matrix $\pi_*$
```

3 Experiments

Our analysis of the methods discussed above was initially focused on simple signals commonly found in music. We began by testing our approach on monophonic and polyphonic signals, as well as well-known simple waveforms such as sawtooth waves. Subsequently, we extended our investigation to evaluate the effectiveness of our method on natural sounds and music within a broader context.

3.1 Monophonic Signals

Figure 4 illustrates the process of smoothly interpolating between two notes. The right figure also highlights that the transport is not linear, a phenomenon that introduces additional smoothness and artistic variation, rather than simply fading the pitch.

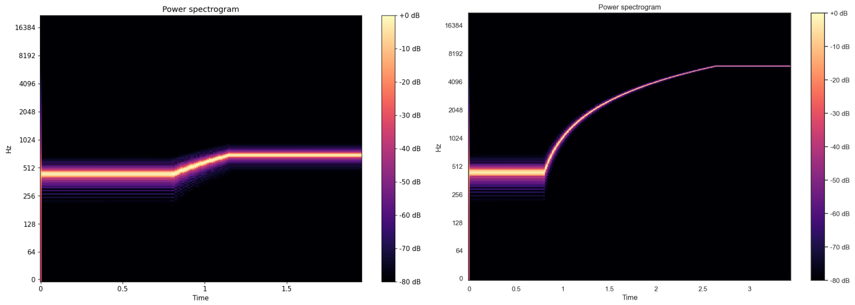

Figure 4: Left: Transition between La (440 Hz) and Ré# (660 Hz), Right: Transition between La (440 Hz) and B6 (6000 Hz)

The temporal aspect is crucial when using such techniques. If the transition time between two signals is too short, it can result in an abrupt transition, leading to audio artifacts, as demonstrated in Figure 5.

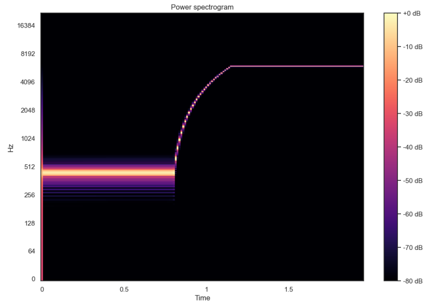

Figure 5: Abrupt transition due to a high transition rate compared to $\frac{1}{\Delta f}$

A limitation of this technique is its reliance on subjective audio evaluation rather than quantitative metrics. Listening to the generated audio reveals that the transitions are indeed smooth and portamento-like. However, artifacts such as vertical and horizontal incoherence, as described in [1], can still be heard. For product-oriented applications, it is essential to address these artifacts by integrating corrective techniques into the processing pipeline. All audios can be found here: `https://www.dropbox.com/scl/fo/awcvwcwlr8620epmsc04u/A083z_ZWKpkGVH65cQC1ybg?rlkey=9cks4r1n4qr4x7jogl85wodya&st=wa4fb30m&dl=0`

3.2 Polyphonic Signals

Evaluating transport on polyphonic signals is crucial, as natural sounds typically feature a fundamental frequency alongside its harmonics. Initially, we aimed to enrich a monophonic signal by interpolating between a sinewave at 587 Hz and a sinewave at 587 Hz with harmonics at 698, 880, and 988 Hz. Figure 6 demonstrates this process, and the resulting audio may be of interest to musicians and producers looking to enhance their existing sounds with additional frequencies.

The previously referenced figure also illustrates a smooth polyphonic transition, resembling a portamento. However, the right part of the figure shows the spectrogram of a transition from a polyphonic sinewave to a polyphonic rectangular wave. The rectangular wave contains significantly more spectral information due to its FFT, which includes all odd harmonics. This can be observed in the figure, where the transport algorithm assigns mass to the nearest frequency, as described in algorithm 1. When listening to the corresponding audio, two key observations can be made: first, a rectangular wave is not as smooth as other waves, and second, the transport process sounds more robotic compared to the previous case. Some artifacts are present, and improving its resolution is left for future work. One clue for this is to use unbalanced transport (Cf 2.4) in order to get rid of artifacts. This will be covered in 3.4.

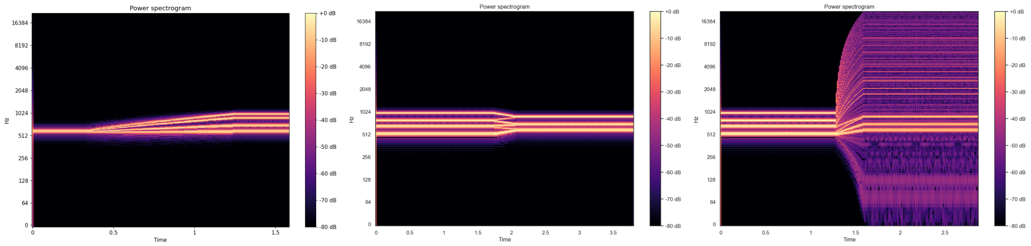

Figure 6: Left: Monophonic to polyphonic sinewave (587 Hz to 587, 698, 880, 988 Hz), Middle: Polyphonic to polyphonic sinewave (523, 659, 784, 988 Hz to 587, 698, 880 Hz), Right: Polyphonic sinewave to polyphonic rectangular wave (523, 659, 784, 988 Hz to 587, 698, 880 Hz).

We also applied the same procedure using a triangle wave instead of a rectangular wave, with the same parameters as shown in Figure 6.

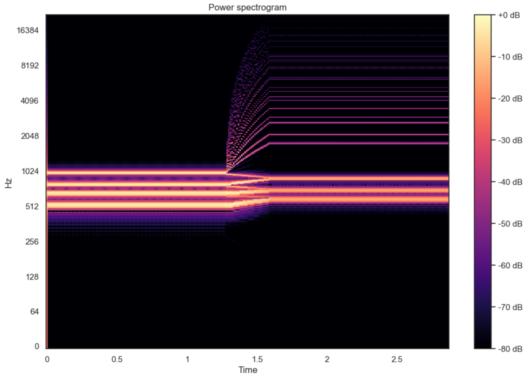

Figure 7: Transition from polyphonic sinewave to polyphonic triangle wave (523, 659, 784, 988 Hz to 587, 698, 880 Hz).

It can be observed that the transition is less smooth, and the low-amplitude frequencies of the triangle wave are influenced by only one frequency of the sine wave. This could result in an audio signal that lacks richness and quality.

3.3 Natural Sounds

Finally, we conducted a test on two natural sounds to evaluate how the algorithm performs with more complex sounds that contain a significant amount of spectral information.We tested transitions between two electronic music tracks and between duck sounds and the acceleration sound of an Aston Martin. Several observations were made: the algorithm cannot be used end-to-end to create smooth transitions; instead, it must be integrated into an artistic framework to fully exploit its potential, particularly for musical applications. It is quite surprising how pleasant the transport sounds for melodic aspects; however, it completely lacks rhythmic integrity and is not executed in an ideal way. For simpler sounds, the effect is quite impressive, although the fading still contains some artifacts; it produces a transition between sounds in a portamento style. The most challenging aspect is blending these effects into a creative context. One possible approach to enriching these transitions is to explore unbalanced transport to address differences in mass between the signals. For example, the mass in the duck sound is much lower compared to the car sound. A key challenge is the ability to create or remove mass during the process, which is why the unbalanced phenomenon could offer promising results.

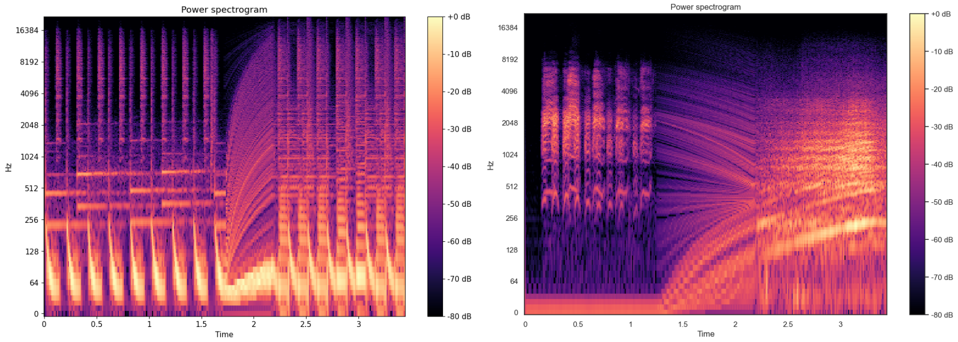

Figure 8: Left: Transition between two electronic tracks, Right: Transition between duck and car sounds.

3.4 Balanced vs Unbalanced Transport

Our experiments revealed that the unbalanced formulation results in either the destruction or conservation of the spectra, but never the creation of mass. This leads to significantly different effects depending on the direction of the transport. For instance, in the duck-to-car transition, switching from the duck to the car does not generate mass, causing the transition to lack volume and preventing the portamento effect. On the other hand, when the process is able to destroy mass, the transition becomes smoother with fewer artifacts compared to the original algorithm. This issue also arises in polyphonic transitions with regular sine waves, making the unbalanced transport appear less promising than the balanced formulation. One potential explanation is that this phenomenon is due to vertical and horizontal incoherence, as changing the ρ parameter does not significantly impact the transport. This remains an area for future exploration.

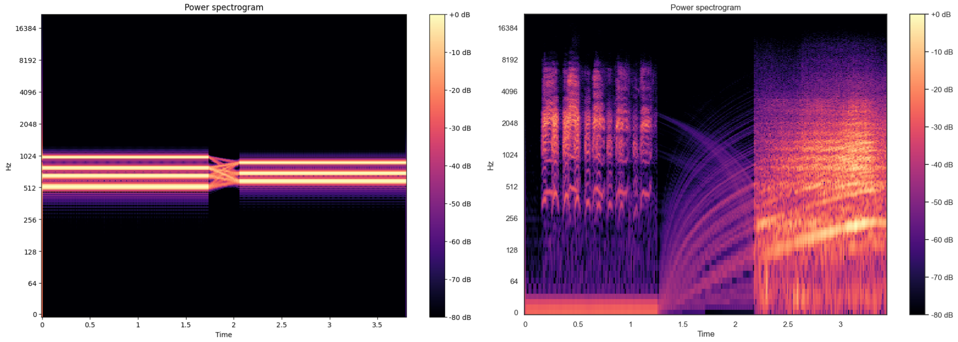

Figure 9: Left: Unbalanced transport on polyphonic sine waves (523, 659, 784, 988 Hz to 587, 698, 880 Hz), Right: Unbalanced transport between a duck and a car sound.

4 Conclusion

This paper explores the audio transport effect, which creates a portamento-like transition between any two audio signals. It is controlled by a single interpolation parameter, making it easy for musicians to incorporate into both live performances and studio recordings. While implemented in Python, it can be replicated in low-level languages for hardware integration. Our approach opens up several other possibilities beyond those explored in [1]. For instance, feeding a single audio source into one input of the effect and routing the output back into the other input can create a glide effect, similar to those used in synthesizers. Another option is to use unbalanced transport and solving incoherence artifacts, which could solve the mass destruction issue. Additionally, mixing more than two signals could be achieved using Wasserstein barycenters [2], enabling simultaneous interpolation of multiple signals. Ultimately, this effect is designed for creative use and should be explored by artists to fully unlock its potential.

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] Filip Elvander et al. “Using optimal transport for estimating inharmonic pitch signals”. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2017, pp. 331–335. DOI: `10.1109/ICASSP.2017.7952172`.

[4] Rémi Flamary et al. *Optimal spectral transportation with application to music transcription*. 2016. arXiv: `1609.09799 [stat.ML]`.

[5] Eduardo F. Montesuma and Fred-Maurice Ngolè Mboula. “Wasserstein Barycenter Transport for Acoustic Adaptation”. In: *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2021, pp. 3405–3409. DOI: `10.1109/ICASSP39728.2021.9414199`.

[6] Antoine Rolet et al. “Blind source separation with optimal transport non-negative matrix factorization”. In: *EURASIP Journal on Advances in Signal Processing* 2018.53 (2018), pp. 1–15. DOI: `10.1186/s13634-018-0576-2`.

[7] Bernardo Torres, Geoffroy Peeters, and Gaël Richard. “Unsupervised Harmonic Parameter Estimation Using Differentiable DSP and Spectral Optimal Transport”. In: *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2024, pp. 1176–1180. DOI: `10.1109/ICASSP48485.2024.10447011`.

[8] Elsa Cazelles, Arnaud Robert, and Felipe Tobar. “The Wasserstein-Fourier Distance for Stationary Time Series”. In: *IEEE Transactions on Signal Processing* 69 (2021), pp. 709–721. DOI: `10.1109/TSP.2020.3046227`.

[9] Bo-Yu Chen et al. “Automatic DJ Transitions with Differentiable Audio Effects and Generative Adversarial Networks”. In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2022, pp. 466–470. DOI: `10.1109/ICASSP43922.2022.9746663`.

[10] Qingqing Huang et al. *Noise2Music: Text-conditioned Music Generation with Diffusion Models*. 2023. arXiv: `2302.03917 [cs.SD]`.

[11] Shih-Lun Wu and Yi-Hsuan Yang. “MuseMorphose: Full-Song and Fine-Grained Piano Music Style Transfer With One Transformer VAE”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 31 (2023), pp. 1953–1967. DOI: `10.1109/TASLP.2023.3270726`.

[12] Aaron Valero Puche and Sukhan Lee. “Caesynth: Real-Time Timbre Interpolation and Pitch Control with Conditional Autoencoders”. In: *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*. 2021, pp. 1–6. DOI: `10.1109/MLSP52302.2021.9596414`.

[13] Justin Solomon et al. “Convolutional wasserstein distances: efficient optimal transportation on geometric domains”. In: *ACM Trans. Graph.* 34.4 (July 2015). ISSN: 0730-0301. DOI: `10.1145/2766963`.

[14] Cédric Villani. *Optimal Transport: Old and New*. Vol. 338. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer Science & Business Media, 2008. ISBN: 978-3-540-71049-3. DOI: `10.1007/978-3-540-71050-9`.

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

[16] J. Laroche and M. Dolson. “New phase-vocoder techniques for pitch-shifting, harmonizing and other exotic effects”. In: *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA’99 (Cat. No.99TH8452)*. 1999, pp. 91–94. DOI: `10.1109/ASPAA.1999.810857`.

[17] Akshay Agrawal et al. “A rewriting system for convex optimization problems”. In: *Journal of Control and Decision* 5.1 (2018), pp. 42–60.

[18] 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`.

Appendix

Connexion with the course

We leveraged the theory of optimal transport (OT) to model smooth transitions between audio signals, such as musical notes. The core idea of OT, as presented in the course notes, is to find the optimal way to transport mass from one distribution to another while minimizing a given cost function. Mathematically, for two probability measures μ and ν defined on $\mathbb{R}^d$, the optimal transport problem is formulated as:

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\, d\pi(x,y)$$

where π is a joint distribution for x and y, belonging to the space $\Pi(\mu,\nu)$ of the product measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν, and $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a cost function. In the project, this theory is applied to audio spectra to achieve musically meaningful transformations.

The project employs the Kantorovich point of view to handle the transport problem between discrete measures. This relaxation allows for mass splitting, making the transport problem solvable using linear programming methods. The Kantorovich problem is formulated as:

$$\mathrm{L}_{\mathbf{C}}(\mathbf{a},\mathbf{b}) = \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{C}, \mathbf{P} \rangle$$

where $\mathbf{C}$ is the cost matrix, $\mathbf{P}$ is the transport plan, and $\mathbf{U}(\mathbf{a},\mathbf{b})$ is the set of couplings that satisfy the marginal constraints:

$$\mathbf{U}(\mathbf{a},\mathbf{b}) = \left\{ \mathbf{P} \in \mathbb{R}_+^{n \times m} : \mathbf{P}\nVdash_m = \mathbf{a} \quad \text{and} \quad \mathbf{P}^{\mathrm{T}}\nVdash_n = \mathbf{b} \right\}$$

We employed linear programming to solve the optimal transport problem, as discussed in both the course material and the numerical tours. Additionally, we explored unbalanced optimal transport as a potential enhancement to our method. This approach allows for differences in the mass of each signal and permits the destruction or creation of mass during interpolation. While this concept was not directly covered in class, it is mentioned in the course notes and numerical tours as a valuable extension.

In hindsight, it would have been beneficial to introduce or experiment with techniques like Entropic Regularization or Sinkhorn's algorithm. These methods could have improved computational efficiency and added robustness to our approach. Unfortunately, time constraints prevented us from incorporating these techniques into the project, leaving them as promising directions for future work.

The original paper [1] does not delve deeply into the theoretical aspects of Optimal Transport but primarily focuses on its practical application. Given that the problem is one-dimensional, the setup is relatively straightforward compared to more complex scenarios. Nevertheless, building a project based on class material was highly beneficial. This project also provided an opportunity to explore numerous papers published in various conferences and journals authored by experts in the field. This exposure not only deepened our understanding of the topic but also highlighted the breadth of ongoing research and innovative applications within the domain of Optimal Transport.