Real-Time Musical Score Following

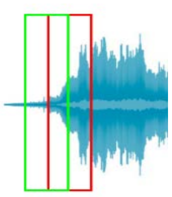
Pierre-Alexandre Hébert, Emilie Poisson Caillault, Ayat Amokrane
Université du Littoral Côte d'Opale, Laboratoire d'Informatique Signal Image Côte d'Opale

Introduction

Score following is the real-time synchronization of a live musician playing a score with the score itself. The area of research stands at the intersection of artificial intelligence, pattern recognition, signal processing, and musicology.

Features Extraction

First thing, we **extract suitable features** that capture relevant key aspects while suppressing irrelevant details.
We apply **The Short-time Fourier transform (STFT)** to our signals.
The procedure to divide a longer time signal into **shorter segments of equal length** and fix a **window function (**which is nonzero for only a short period of time) then slid the window along the time axis and compute the Fourier transform for each of the resulting windowed signals. [1]



$$S(m,k) = \sum_{n=0}^{N-1} x(n+mH) \cdot w(n) \cdot e^{-i2\pi n \frac{k}{N}}$$

Fig 1 : Application of STFT

This reveals **the Fourier spectrum** on each shorter segment whose linear frequency axis (Hertz) is converted into a logarithmic axis (heights).

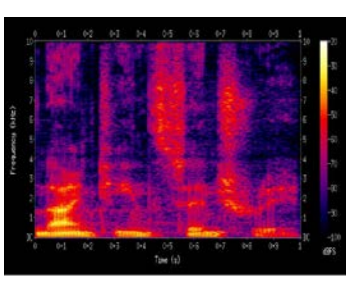


Fig 2 : Spectrogram

We then derive a temporal chromatic representation by appropriately combining the pitch bands that correspond to the same pitch or chroma class. [1]

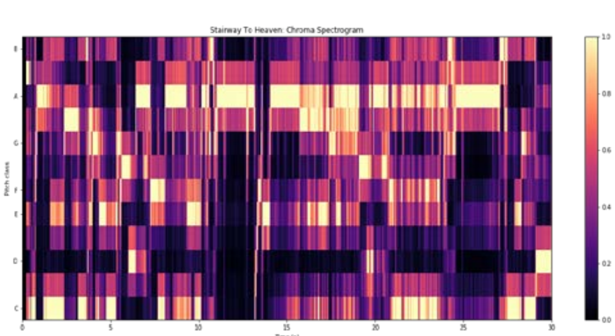


Fig 3 : Chromatic representation

System Overview

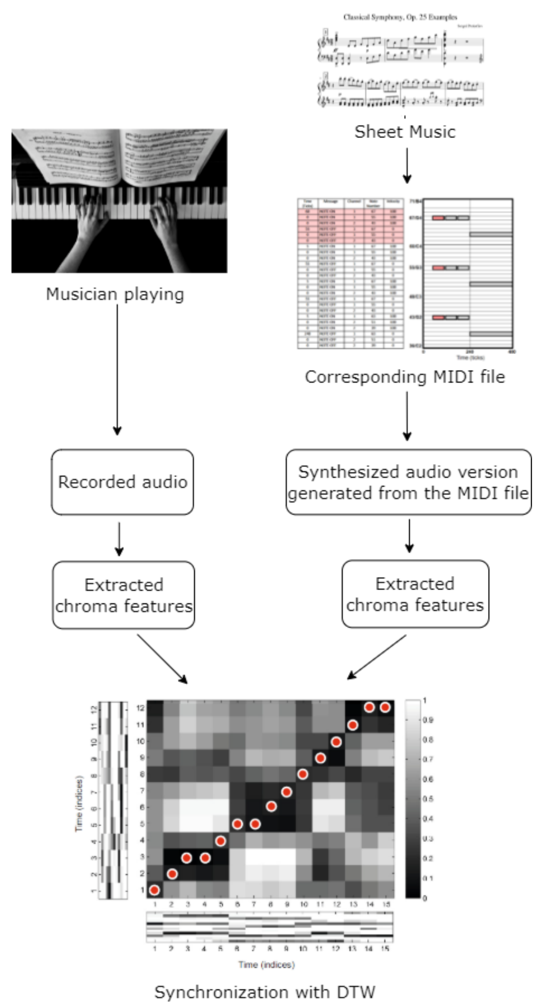


Fig 4 : Music synchronization

Method

- **Musical synchronization** is a procedure which, for a given position in one representation of a piece of music, determines the corresponding position in another representation. [1]
- **Chromatic representation** of the signal has proven to be a powerful tool for analyzing music as it shows a high degree of robustness to variations in timbre and dynamics. [1]
- After extracting chroma features, we apply **Dynamic Time Warping** (DTW) to find optimal temporal correspondences between the elements of two given chroma sequences. [1]

Dynamic Time Warping

DTW aligns time series U = $u_1, \ldots, u_m$ and V = $v_1, \ldots, v_n$ by finding a minimum cost path W = $W_1, \ldots, W_l$, where each W_k is an ordered pair (i_k, j_k), such that $(i, j) \in W$ means that the points u_i and v_j are aligned. [1]
Several constraints are placed on W :

Bounds: $W_1 = (1,1)$
$W_l = (m,n)$

Monotonicity: $i_{k+1} \geq i_k$ for all $k \in [1, m-1]$
$j_{k+1} \geq j_k$ for all $k \in [1, n-1]$

Continuity: $i_{k+1} \leq i_k + 1$ for all $k \in [1, m-1]$
$j_{k+1} \leq j_k + 1$ for all $k \in [1, n-1]$

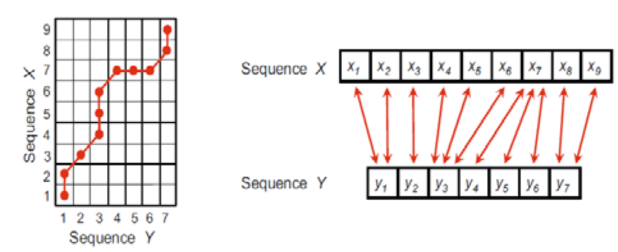


Fig 5 : Sequences alignment

Conclusion

The resulting correspondences establish a musically meaningful linking structure between the given music representations.

Using this method, different approaches [2] can be employed to achieve **online synchronization** where data streams have to be processed in **real time**. The level of acceptance of these techniques lies in their **efficiency** and **accuracy**.

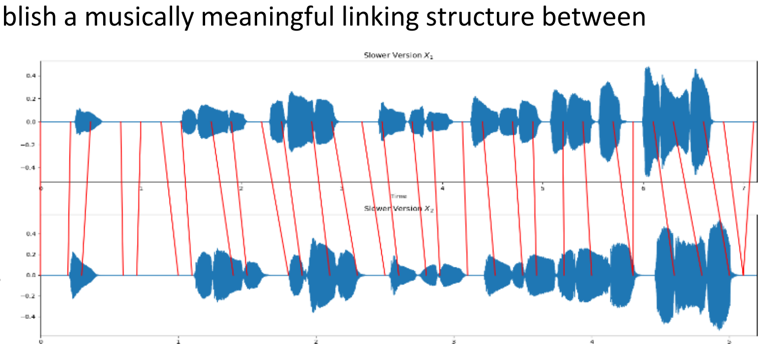


Fig 6 : Visualizing the warping path directly on time domain signal

References

[1] Müller, Meinard - Fundamentals of Music Processing_ Audio, Analysis, Algorithms, Applications (2015, Springer International Publishing) - libgen.lc
[2] Simon Dixon - Live tracking of musical performances using on-line time warping (Austrian Research Institute for Artificial Intelligence)