Uncovering audio patterns in music with Nonnegative Tucker Decomposition for structural segmentation

A. Marmoret, J.E. Cohen, N. Bertin, F. Bimbot

Univ Rennes, Inria, CNRS, IRISA, France - axel.marmoret@irisa.fr

Poster summary

This poster presents a tensor factorization technique called Nonnegative Tucker Decomposition (NTD), and its application to MIR.

We find that NTD can extract new representations of music signals and uncovers audio patterns in music pieces.

We evaluate this representations in the task of structural segmentation of music in audio form.

NTD was first applied to music modeling by [1].

Time-Frequency-Bar tensor

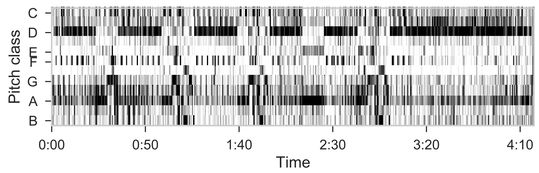


Figure 1. Chromagram of "Come Together", by The Beatles.

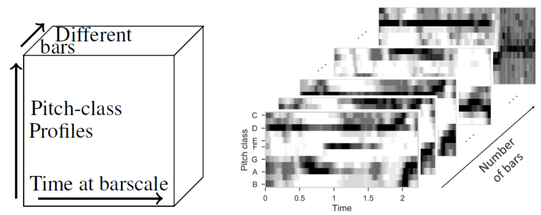
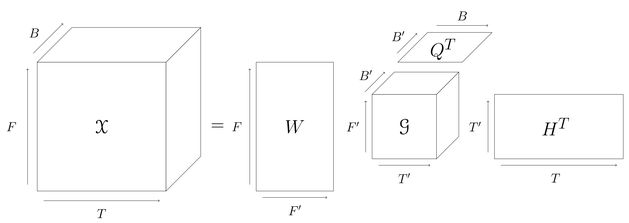


Figure 2. TFB tensor model. It can be seen as a concatenation of chromagrams for each bar along a third dimension.

We estimate bars using the toolbox madmom [2]. Then, we cut the chromagram on each downbeat, and fold them in a tensor with two temporal dimensions: inner-bar time and bar indexes.

Nonnegative Tucker Decomposition



The Tucker Decomposition can be written, element-wise:

$$\mathcal{X}(f,t,b) \approx \sum_{f',t',b'=1}^{F',T',B'} \mathcal{G}(f',t',b') W(f,f') H(t,t') Q(b,b')$$

Musical pattern

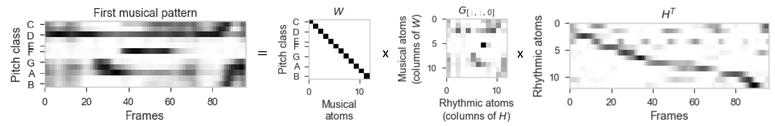


Figure 3. Musical pattern, representing the first bar of "Come Together" (far-left): this bar is decomposed as a linear combination of columns of W (center-left, chroma information) and of H (far-right, rhythmic information). The combination is defined by the first slice of $\mathcal{G}$ (center-right).

Each slice of the core defines a musical pattern, and is associated with a column of the Q matrix.

Q matrix: musical patterns as features for the bars.

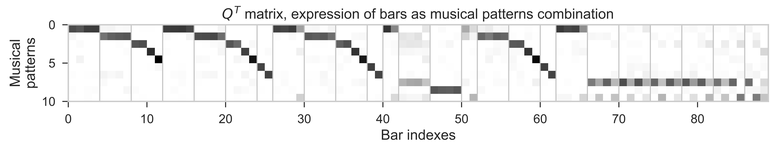


Figure 4. Q^T matrix of "Come Together", with $T' = 12$ and $B' = 10$. Grey lines: segmentation annotation.

Hence, Q^T is a barwise representation of the song, with musical patterns as features.

Autosimilarity QQ^T: barwise similarity coefficients

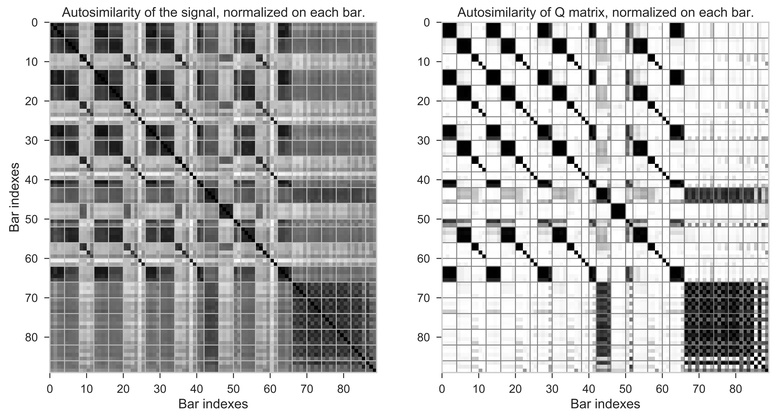


Figure 5. Barwise autosimilarities of the chromagram (left) and of the Q^T matrix (right). Grey lines: segmentation annotation.

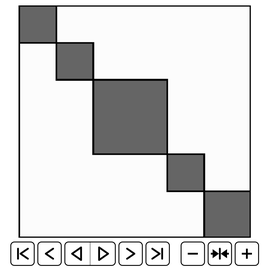
The autosimilarity of Q^T, *i.e.* QQ^T, seems to polarize the autosimilarity distribution. This is due to the sparsity of Q. Meanwhile, high similarity blocks seem to be preserved.

Segmentation algorithm

Our goal is to frame dark blocks (zones of high similarity) as segments, as they share the same (or a very similar) musical patterns representation, *i.e.* musical content.

To this end, we developed a dynamic programming algorithm, which aims at optimally fitting these dark zones.

This is a maximization algorithm, where the cost is a convolution with a predetermined kernel, weighted by the size of the segment.



Note: this gif only animates in Adobe Acrobat Reader©. Otherwise, you can see it a this link: https://gph.is/g/a9YJjN5

Segmentation results on RWC Pop

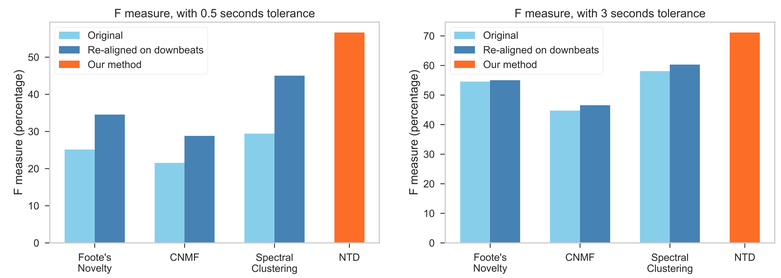


Figure 6. Segmentation scores (F measure only), with baselines, respectively [3], [4] and [5]. The condition "Re-aligned on downbeats" means that we manually aligned the original frontiers on our estimated downbeats.

Going further: article, code and notebooks

See our article: "Uncovering audio patterns in music with Nonnegative Tucker Decomposition for structural segmentation".

The code of our entire process is open source:

https://gitlab.inria.fr/amarmore/musicntd/-/tree/0.1.0

This folder also contains *Notebooks*, which present detailed experimental results and more technical information about our technique. Feel free to dig it for more details!

References

[1] J. B. Smith and M. Goto, "Nonnegative tensor factorization for source separation of loops in audio," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 171–175, IEEE, 2018.

[2] S. Böck, F. Korzeniowski, J. Schlüter, F. Krebs, and G. Widmer, "madmom: a new Python Audio and Music Signal Processing Library," in *Proc. of the 24th ACM International Conference on Multimedia*, (Amsterdam, The Netherlands), pp. 1174–1178, 10 2016.

[3] J. Foote, "Automatic audio segmentation using a measure of audio novelty," in *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proc. Latest Advances in the Fast Changing World of Multimedia (Cat. No. 00TH8532)*, vol. 1, pp. 452–455, IEEE, 2000.

[4] O. Nieto and T. Jehan, "Convex non-negative matrix factorization for automatic music structure identification," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 236–240, IEEE, 2013.

[5] B. McFee and D. Ellis, "Analyzing song structure with spectral clustering.," in *ISMIR*, pp. 405–410, 2014.