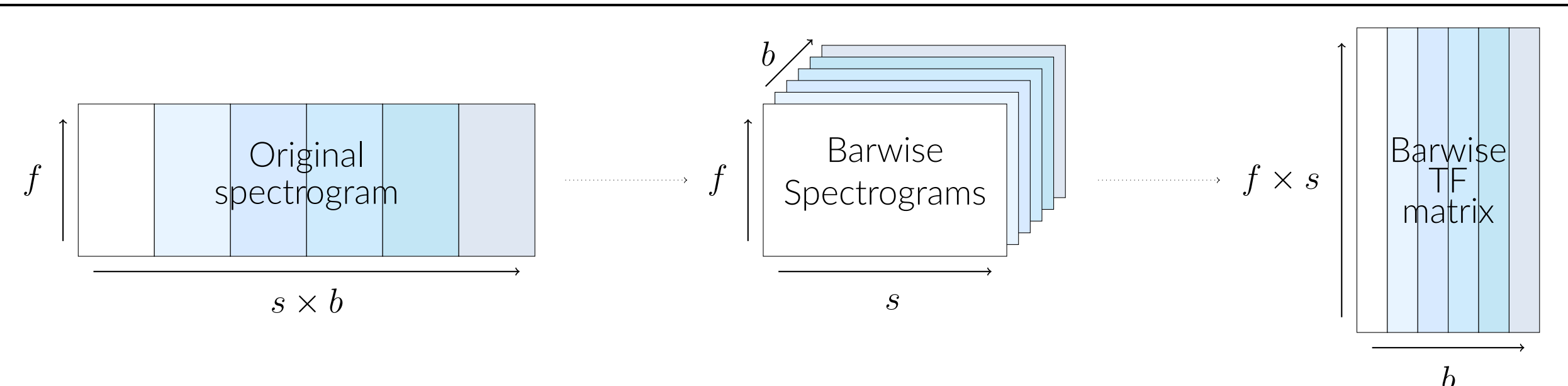


Barwise Compression Schemes for Audio-Based Music Structure Analysis

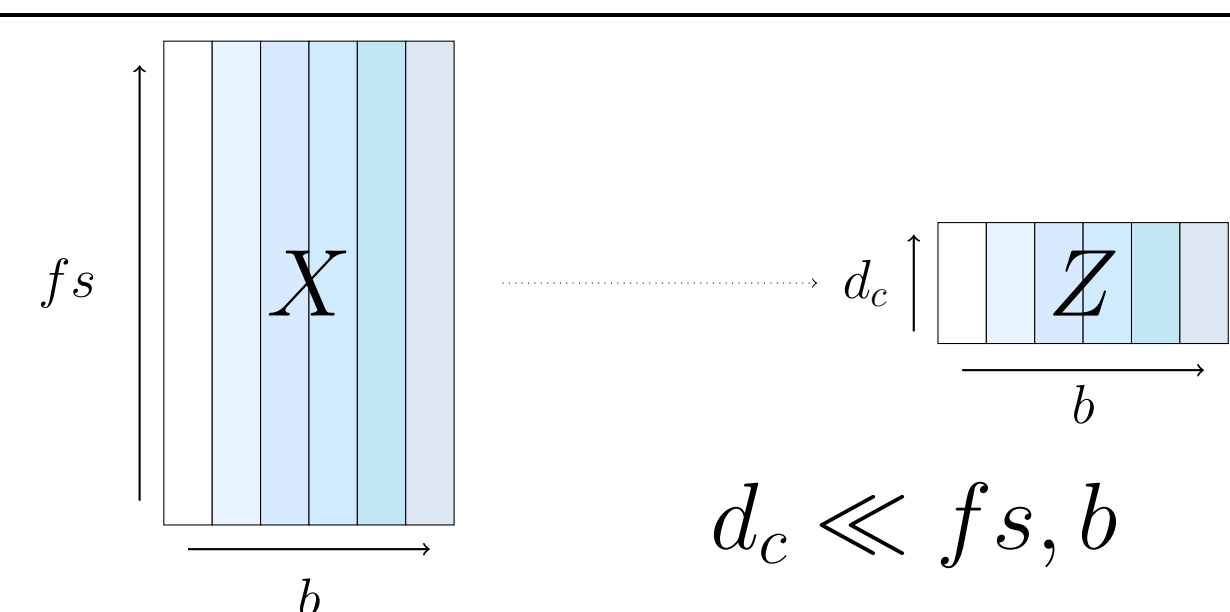
Overall process



Barwise Processing



Barwise Compression



- Low-rank approximations: Nonnegative Matrix Factorization (NMF), Principal Component Analysis (PCA),
- Single-Song Autoencoders (SSAE): Autoencoders optimized on a unique song.

NMF

Computes two nonnegative matrices W and Z such that:

$$\operatorname{argmin}_{W \geq 0, Z \geq 0} \|X - WZ\|_F^2 \quad (1)$$

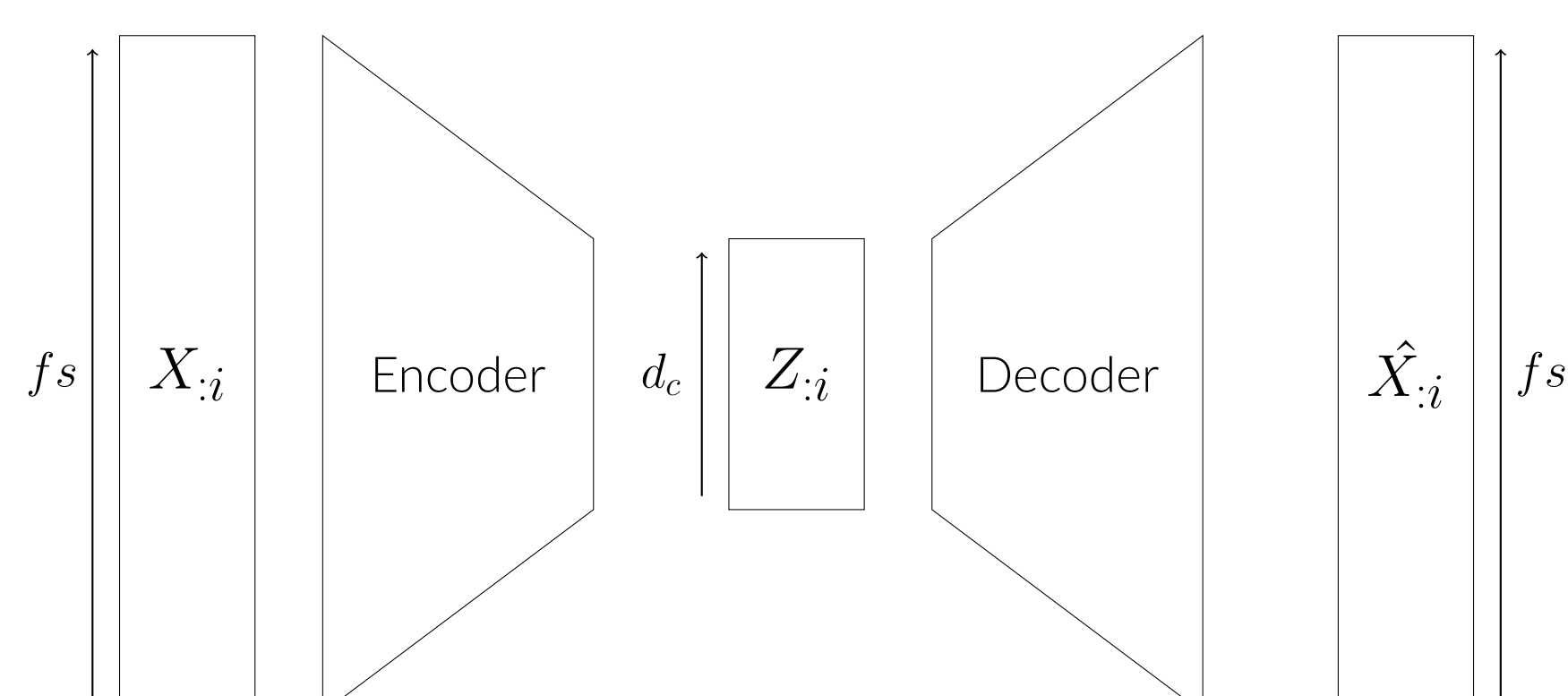
PCA

Starting from the Singular Value Decomposition of the centered matrix $X - \mu = U\Sigma V^T$, PCA projects the data on the d_c largest left singular vectors, i.e.:

$$Z = U_{[1, d_c]}^T (X - \mu) \quad (2)$$

SSAE

Optimizing an autoencoder on the barwise representation of the song (each $X_{:,i}$ is a column of X , i.e. a barwise spectrogram):



The Autoencoder solves:

$$\operatorname{argmin}_{\hat{X}} \|X - \hat{X}\|_F^2 \quad (3)$$

References

[Foo] Foote, J. Automatic audio segmentation using a measure of audio novelty. ICME2000.

[McF+] McFee, B., & Ellis, D. Analyzing Song Structure with Spectral Clustering. ISMIR2014.

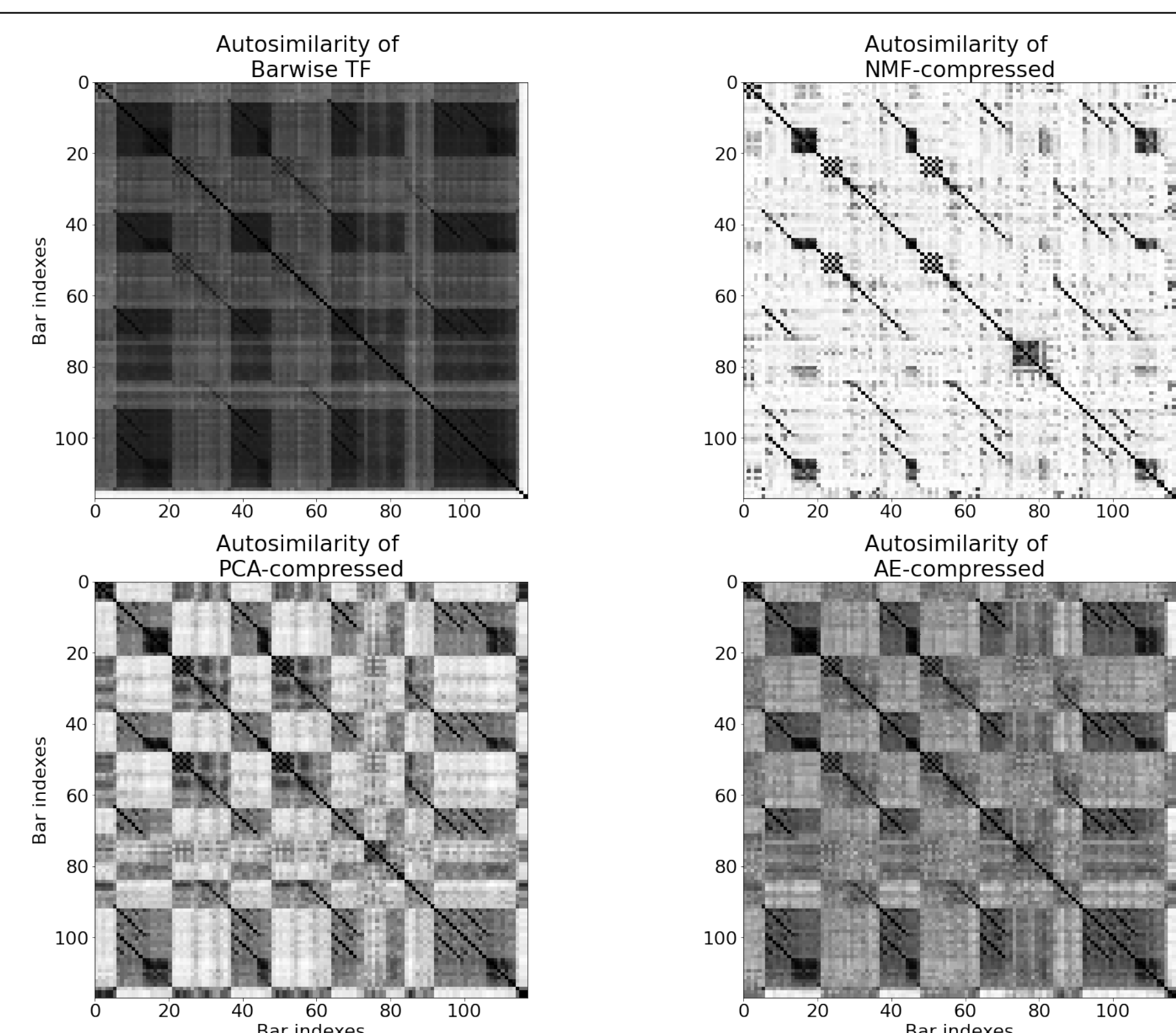
[Mar+] Marmoret, A., Cohen, J. E., Bertin, N., & Bimbot, F. Uncovering audio patterns in music with Nonnegative Tucker Decomposition for structural segmentation. ISMIR2020.

[Gri+] Grill, T., & Schlüter, J. Music Boundary Detection Using Neural Networks on Combined Features and Two-Level Annotations. ISMIR2015.

Structural Segmentation

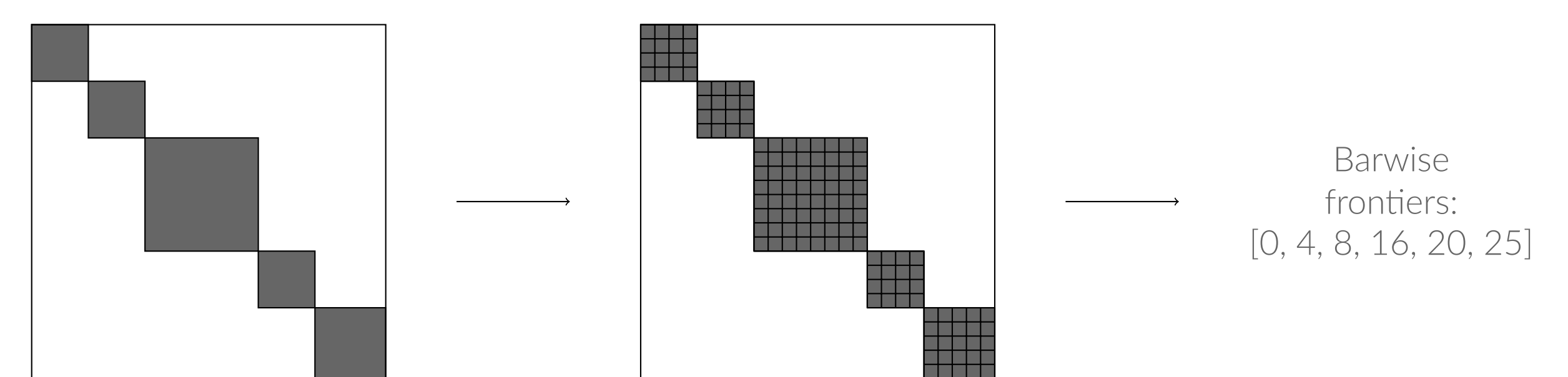
Structural segmentation is obtained by segmenting the **autosimilarity matrix** $Z^T Z$ of Z , via a convolution based **dynamic programming algorithm**.

Autosimilarities in practice

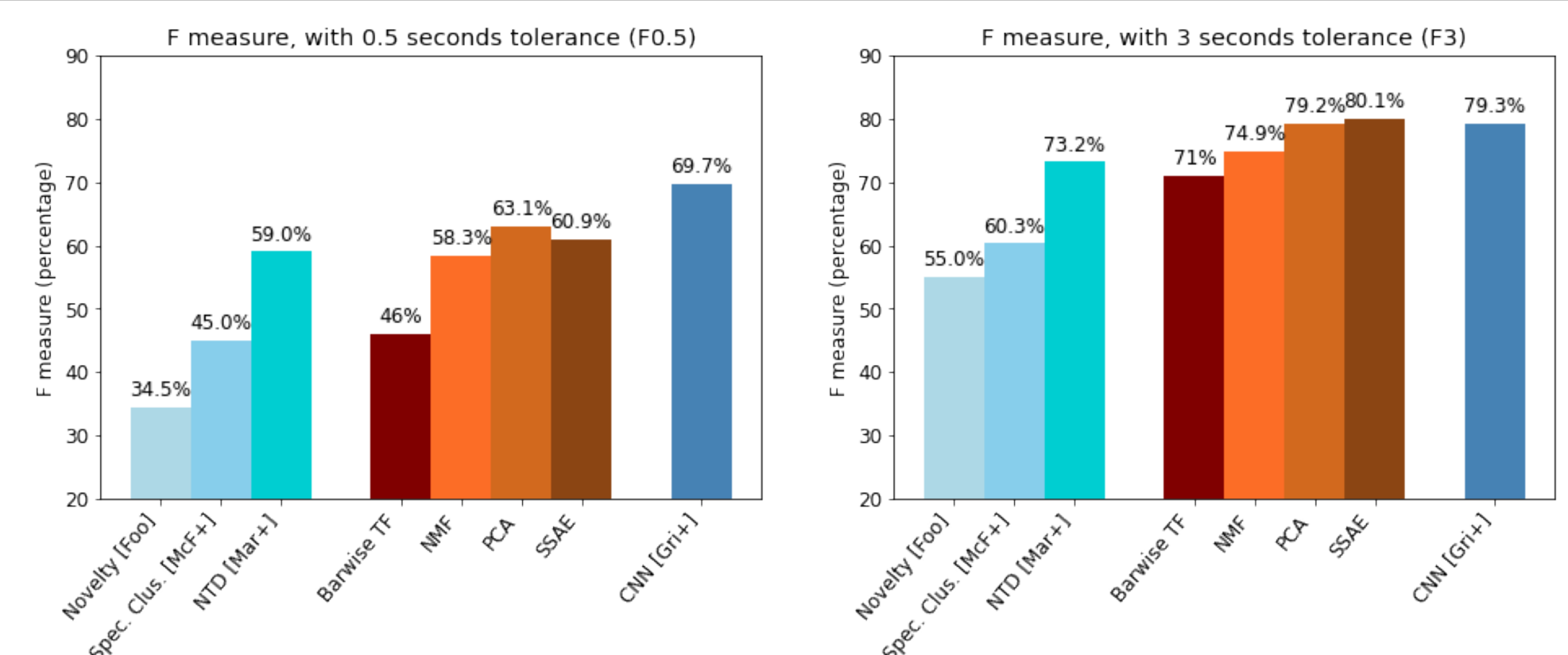


Idea of the segmentation algorithm

The goal is to frame squares of high similarity, by computing a local cost as the average similarity on the diagonal:



Performance



Conclusions

- ✓ Comparable performance with fully trained CNN
- ✓ PCA and NMF are fast and easy-to-use
- ✓ SSAE can be largely improved
- ✓ SSAE is time consuming
- ✓ Depends on barwise division of music

