Discriminant Approach within Non-negative Matrix Factorization for Musical Components Recognition

Zafar Rafii, Raphaël Blouet & Antoine Liutkus

Mist Technologies R&D - Paris Cybervillage - 204 rue de Crimée, 75019 Paris, France
firstname.lastname@mist-technologies.com

Application context

- **Mist Technologies** develops system for single sensor source separation : Upmix for professionals (DVD reissues, Movies:*La Vie en Rose*) and general public applications (www.songcooker.com).
- Prior information on mixture components allows to obtain good performance but requires musical instruments identification. **This task is handmade and very costly.**
- As instrument detection in polyphonic recordings appears to be a very difficult task, we first focus on close set instruments identification : *Given a close set of musical instruments, is it possible to automatically identify some components in polyphonic music ?*
- This system could also be used as a general front-end to any high level information retrieval systems (genre detection, transcription, ...)

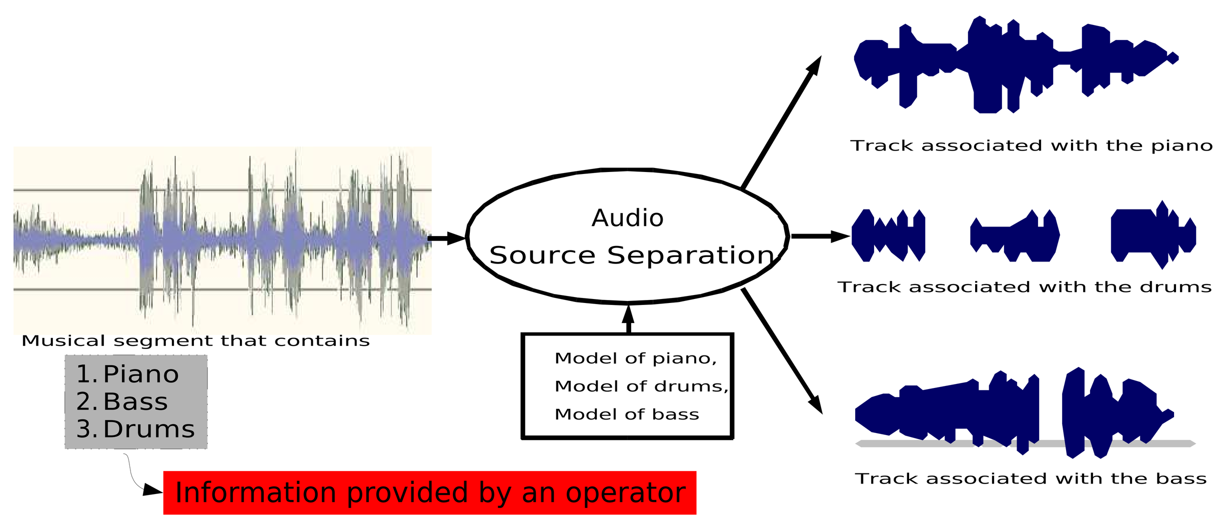

FIGURE 1: General Single Sensor Source Separation System Overview

Scientific context

- Investigation of the use of Non-negative Matrix Factorization (NMF) to model audio signals.
  . Virtanen takes advantage of NMF for sound source separation [Vir06]
  . Smaragdis proposes a modified version of the NMF algorithm which is able to identify components with temporal structure [Sma04].
  . Cont [CD07] applies NMF on modulation spectrum, but also on spectrogram for pitch estimation and instrument recognition.
  . Benetos, Kotti and Kotropoulos use NMF on matrix of nonnegative feature vectors extracted from audio files for instrument classification [BKK06].

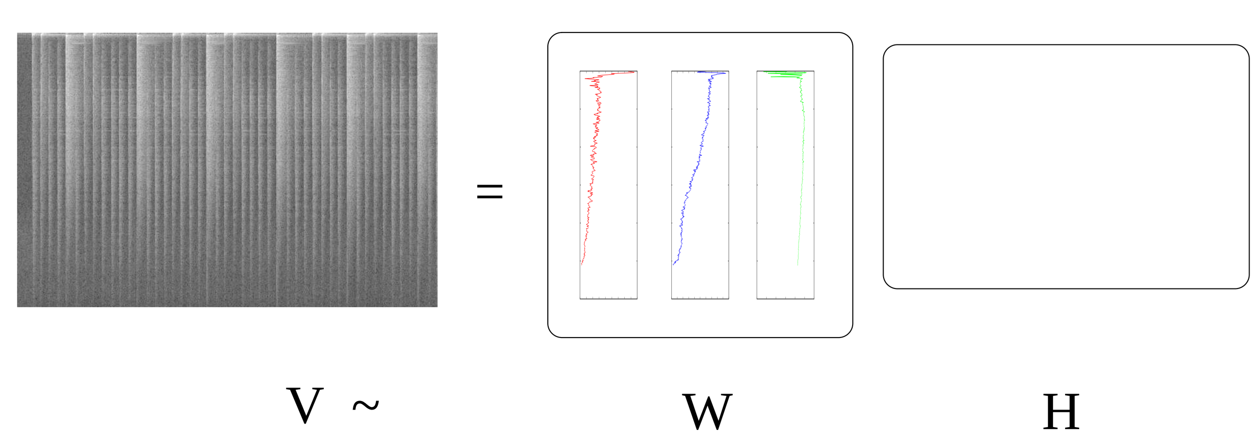

FIGURE 2: NMF on a spectrogram computed from a polyphonic musical excerpt with $r = 3$ components

- NMF factorizes a nonnegative matrix V into two nonnegative matrices W & H seeking to minimize a specific cost function $\mathcal{F}$.
- NMF being not unique, appropriate additional constraints can lead to different solutions, with different properties of the representation (*e.g* sparsness, smoothness).
- NMF is an unsupervised learning algorithm, therefore we propose a discriminant approach to enforce an enhanced part-based decomposition more adapted to instrument detection.

Enhancements

NMF is not unique: the factorization depends not only on the update rules but also on the starting point.

Initialization

Initializations for W & H :
- Random positive matrices
- Spherical K-means clustering
- *Nonnegative Double Singular Value Decomposition (NNDSVD)* [BG07]
  Experimental results show that NNDSVD :
  . leads to rapid reduction of the approximation error
  . leads to better results than other methods: less redundancy, better sparseness and more localized parts within the extracted components

Sparseness/Smoothness Constraints

Additional constraints can also be included in the update rules to enforce a convergence. Favoring components whose gains are sparse and slowly varying [Vir06].

Number of Components

There is no reliable method for the automatic estimation of the number of components r. We determine r by applying a first coarse NMF, using a large number of components. To that end, Virtanen's NMF, initialized with a NNDSVD, with a strong sparseness criteria helps to keep only the components with relevant energy. Those components are then filtered from their null values and used as an initialization for a second standard NMF.

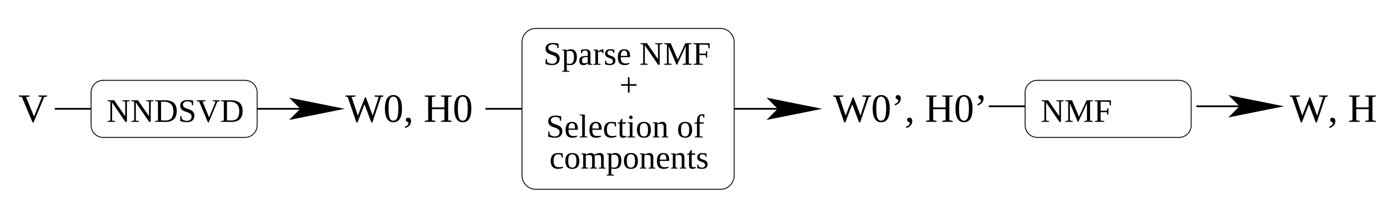

FIGURE 3: Enhanced NMF System Overview for Musical Component Separation

Forced Discriminant NMF

We use our enhanced NMF algorithm described above to extract the fixed magnitude spectrum components from a training database formed of K=5 instrument classes $\mathcal{C}_k$ (Bass, Drums, Piano, Saxophone,Trumpet), and stored them in dictionaries.
However, NMF is an unsupervised algorithm, namely singly applied on solo instrument excerpts, it would not take into account the class-specific information. Therefore, some NMF methods including discriminant constraints have been proposed, basically based on the Fisher Linear Discriminant Analysis (FLDA).
Likewise, we propose our own discriminant method, *the Forced Discriminant NMF (FD-NMF)* extending our enhanced NMF algorithm to the whole database in the same time, including a discriminant approach into the process to enhance the separability between classes.

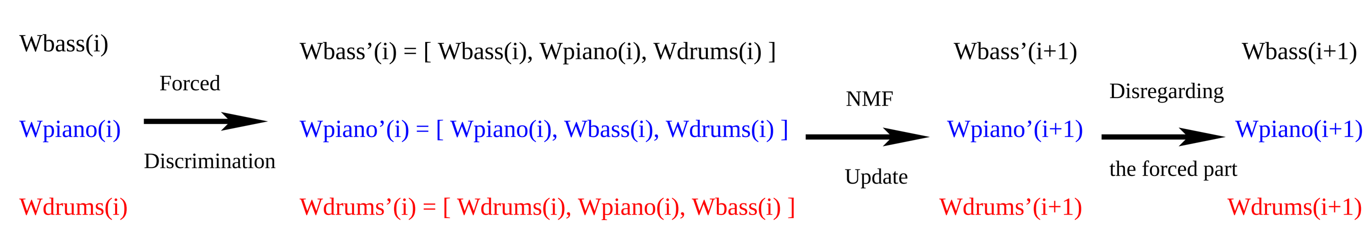

FIGURE 4: one Forced Discriminant NMF iteration

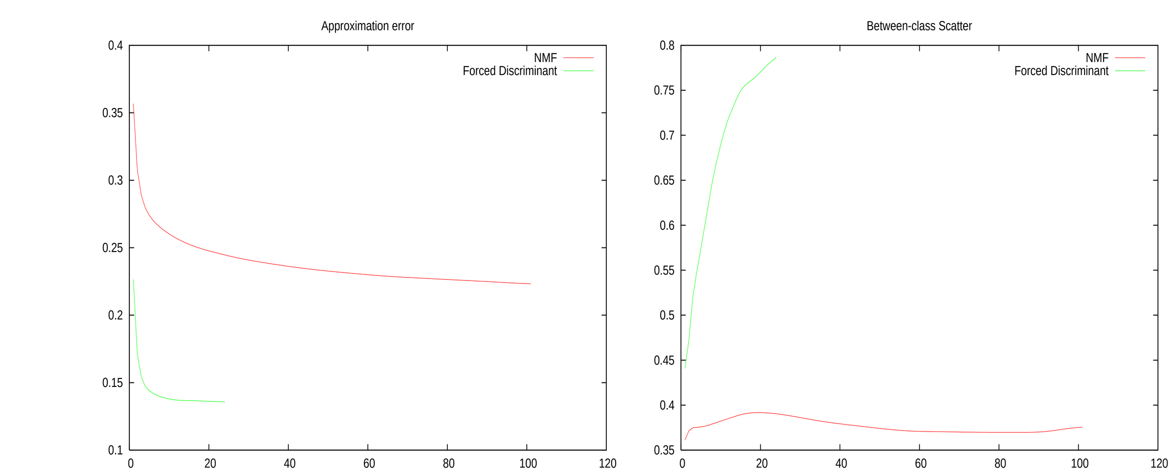

FIGURE 5: Error approximation & Between-class Scatter for NMF & FDNMF

Evaluation & Conclusions

System overview

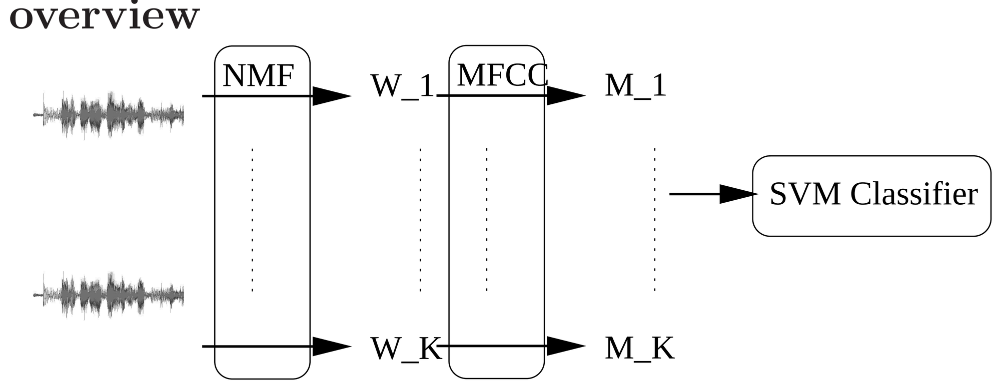

FIGURE 6: Musical Components Recognition Training phase (NMF & FDNMF)

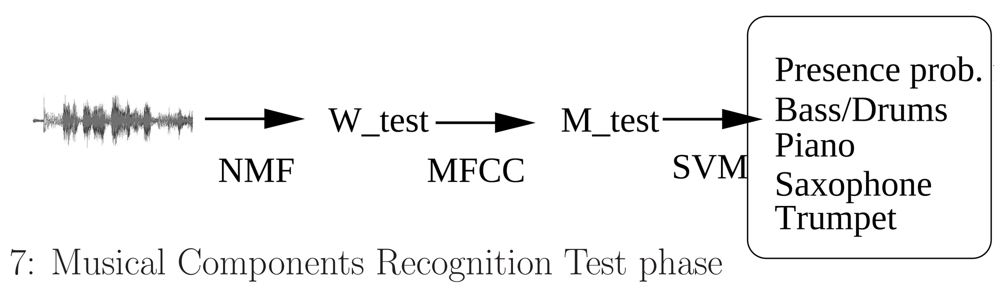

FIGURE 7: Musical Components Recognition Test phase

Evaluation

The test set consists of 150 musical excerpts from the RWC database. Evaluation is done using 5 target instruments : Bass, Drums, Piano, Saxophone & Trumpet:

	Enhanced NMF		FDNMF	
	P_{FA}	P_{Miss}	P_{FA}	P_{Miss}
All instruments	39%	43%	33%	42%

Conclusion

- A NMF based system for Musical Component Recognition has been proposed
- A discriminant approach has been added to the standard NMF system which provides a slight performance enhancement
- Moreover, test database contains extra unknown components which bias the results (a reject option should be included)
- Future works should also include the use of gain coefficients for better analysis

References

[BG07] Christos Boutsidis and Efstratios Gallopoulos. Svd based initialization: a head start for nonnegative matrix factorization. Technical report, 2007.

[BKK06] Emmanouil Benetos, Margarita Kotti, and Constantine Kotropoulos. Application of non-negative matrix factorization to musical instrument classication, 2006.

[CD07] Arshia Cont and Shlomo Dubnov. Realtime multiple-pitch and multiple-instrument recognition for music signals using sparse non-negative constraints. In *Proceedings of Digital Audio Effects Conference (DAFx)*, 2007.

[Sma04] Paris Smaragdis. Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs. In *International Symposium on Independent Component Analysis and Blind Signal Separation*, pages 494–499, September 2004.

[Vir06] Tuomas Virtanen. Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006.