

MSc Research Internship Opening (6 months)

Unrolled nonnegative matrix factorization techniques for audio source separation in “aircraft black box” recordings

Keywords:

Cockpit Voice Recorder; sounds; speech; blind source separation; unrolled techniques; nonnegative matrix factorization; clipped mixtures; audio coding.

Description:

Civil and state aircraft are equipped with two flight recorders (also known as “black boxes”), *i.e.*, the Cockpit Voice Recorder (CVR) and the Flight Data Recorder. In the event of an incident or accident, both must be found and analyzed by the competent authorities. The audio service of BEA (*Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile*) and RESEDA are the French authorities in charge of analyzing the CVRs of civil and state aircrafts, respectively. CVRs are reinforced equipment protecting important information for determining the contributing factors to the occurrence of an accident or an air incident. For this reason, the audio quality of the recordings, the audibility of the cockpit alarms, and the intelligibility of the voice exchanges are crucial for BEA and RESEDA investigators. The contents of the CVRs are then transcribed by specialized investigators (audio analysts) for the benefit of the security investigation (Fig. 1).

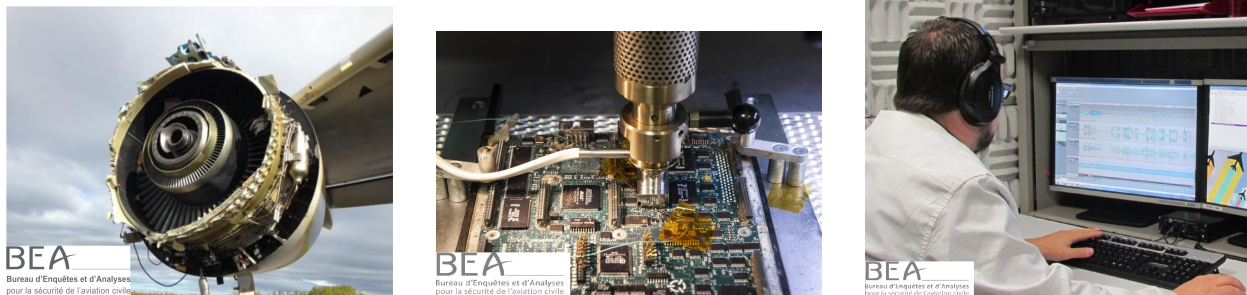


Figure 1: After a civil aircraft incident or accident (left image), the CVR is opened (middle image) and an audio analyst transcribes its contents (right image). Source: [1].

The CVR design limitations force aircraft manufacturers to mix the sound sources which are audible in the cockpit and in the pilots’ headsets (radio transmissions and receptions, intercom exchanges, announcements to passengers, audible alarms, etc.) into a reduced number of audio channels sent to the CVR [1], which makes a digital acquisition and protects this data in the event of an accident. Thus, the simultaneous activation of several sound sources can lead to a significant reduction in the intelligibility of voice exchanges, or even a complete masking of one or more audio sources.

Extracting information from CVR data relies on the experience of BEA/RESEDA audio analysts and their ability to detect sound sources buried in difficult audio mixes [2]. In our recent work [1], we proposed a model of audio mixing in CVRs by reverse engineering (see fig. 2) and we showed that state-of-the-art Blind Source Separation (BSS) methods could be applied. BSS aims to estimate a set of unknown signals from mixtures of them, the mixture itself being unknown [3].

Within the framework of the ANR BLeRIOT project—which funds this internship—we do not wish to develop deep learning methods requiring a large training set. On the contrary, the constraints of safety investigations lead us to develop unsupervised or weakly supervised approaches that are interpretable.

Nonnegative matrix factorization (NMF) methods constituted the state of the art before the emergence of deep learning in the early 2010s and fulfill both criteria mentioned above. However, they do not always guarantee a

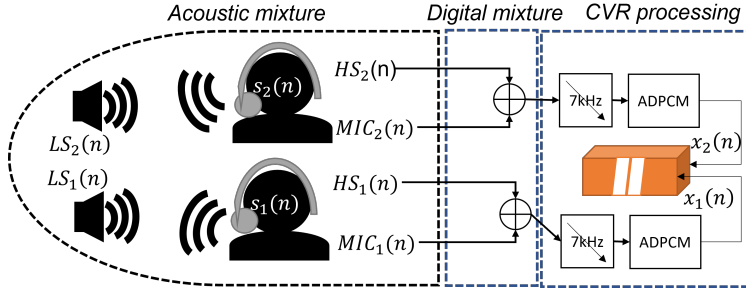


Figure 2: Mixing process in the CVR. Source: [1].

good separation performance when applied to CVR signals. The principle of unrolled or unfolded techniques allows us to transform classical iterative approaches into deep learning methods and enable a good compromise between performance, model interpretability, and training set size [4]. These techniques have been combined with NMF for a number of applications [5–8], each with specific characteristics of its proposed method.

In this internship, we aim to propose unrolled NMF methods that we will apply to CVR data. In particular, we have identified situations where the model proposed in [1] is no longer valid, either due to audio channel saturation effects caused by the avionics audio system or lossy coding effects due to a particular CVR model. These different effects have been studied separately in the literature—e.g., by [9, 10] for audio declipping and by [11] for audio coding effects on separation quality—but, to the best of our knowledge, they have never been considered jointly or specifically for CVR signals. An important aspect of the internship will be to compare the performance of classical or unfolded NMF methods for such scenarios.

Internship context:

This internship is funded by the ANR ASTRID “BLerIOT” and will take place in the newly-created LoRAL team (Low-Rank Approximation and Learning) at LISIC. This team has currently 8 permanent professors, 2 emeritus professors, 3 postdoctoral researchers, and 11 Ph.D. students. A large part of the team works on BSS-type problems and the recruited intern will be able to interact with the entire team. The laboratory is located in the heart of the Regional Natural Park “Caps and Marais d’Opale”, close to Lille, England, Belgium, and Northern Europe. Its premises in Longuenesse are next to the public university residence and close to all amenities (everything can be done on foot or by bus). The teaching team can also put the recruited intern in contact with private landlords.

To apply:

Student in data science (signal and image processing, computer science with a focus on artificial intelligence / machine learning, applied mathematics), you are curious and very comfortable in programming (Matlab and / or Python). You read and speak fluent English without difficulty. You are able to clearly explain your work, even to non-expert audiences. Although not required, a first experience in matrix factorization (for example, in matrix or tensor decomposition, in source separation, in dictionary learning) will be appreciated. **Candidates must be of French nationality or citizens of a member state of the European Union, a state that is part of the European Economic Area or the Swiss Confederation.**

To apply, please send an email to {gilles.delmaire, matthieu.puigt} [at] univ-littoral.fr while attaching the documents that can support your application:

- CV;
- Cover letter;
- Transcripts since the last year of Bachelor;
- two letters of recommendation or the names and contact details of two academic referees.

References

- [1] Matthieu Puigt, Benjamin Bigot, and Hélène Devulder. Introducing the “cockpit party problem”: Blind source separation enhances aircraft cockpit speech transcription. *Journal of the Audio Engineering Society*, 73(1/2), 2025.
- [2] BEA. [Guide sur la vérification des enregistrements CVR](#).
- [3] Pierre Comon and Christian Jutten, editors. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Elsevier, 2010.
- [4] Vishal Monga, Yuelong Li, and Yonina C Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Process. Mag.*, 38(2):18–44, 2021.

- [5] Scott Wisdom, Thomas Powers, James Pitton, and Les Atlas. Deep recurrent nmf for speech separation by unfolding iterative thresholding. In *Proc. WASPAA'17*, pages 254–258. IEEE, 2017.
- [6] Rami Nasser, Yonina C Eldar, and Roded Sharan. Deep unfolding for non-negative matrix factorization with application to mutational signature analysis. *Journal of computational biology*, 29(1):45–55, 2022.
- [7] Christophe Kervazo, Abdelkhalak Chetoui, and Jérémy E Cohen. Deep unrolling of the multiplicative updates algorithm for blind source separation, with application to hyperspectral unmixing. In *Proc. EUSIPCO'24*, pages 656–660. IEEE, 2024.
- [8] Sheng Shu, Ting-Zhu Huang, Jie Huang, and Gemine Vivone. Nesterov-accelerated non-negative matrix factorization unrolling network for hyperspectral unmixing. *Neurocomputing*, page 131202, 2025.
- [9] Amir Adler, Valentin Emiya, Maria G Jafari, Michael Elad, Rémi Gribonval, and Mark D Plumbley. Audio inpainting. *IEEE Audio, Speech, Language Process.*, 20(3):922–932, 2011.
- [10] Çağdaş Bilen, Alexey Ozerov, and Patrick Pérez. Joint audio inpainting and source separation. In *Proc. ICA/LVA'15*, pages 251–258. Springer, 2015.
- [11] Matthieu Puigt, Emmanuel Vincent, Yannick Deville, Anthony Griffin, and Athanasios Mouchtaris. Effects of audio coding on ICA: an experimental study. In *Proc. IEEE ECMSM'13*, Toulouse, France, June 24–26, 2013.

Offre de stage de Master 2 (6 mois)

Techniques dépliées de factorisation matricielle non-négative pour la séparation de sources audiophoniques dans les enregistrements de « boîtes noires aéronautiques »

Mots clés :

Cockpit Voice Recorder ; sons ; parole ; séparation aveugle de sources ; approches dépliées ; factorisation matricielle non-négative ; mélanges saturés ; codage audio

Description :

Les avions civils et étatiques sont équipés de deux enregistreurs de vol (aussi connus sous le nom de « boîtes noires »), c.-à-d. le *Cockpit Voice Recorder* (CVR) et le *Flight Data Recorder*. En cas d'incident ou d'accident, tous deux doivent être retrouvés et analysés par les autorités compétentes. Le service audio du BEA (Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile) et RESEDA sont les autorités françaises en charge des analyses des CVR des avions respectivement civils et étatiques. Les CVR sont des équipements renforcés protégeant une information importante pour la détermination des facteurs contributifs à la survenue d'un accident ou d'un incident aérien. Pour cette raison la qualité audio des enregistrements, l'audibilité des alarmes du cockpit et l'intelligibilité des échanges vocaux sont cruciales pour les enquêteurs du BEA et RESEDA. Les contenus des CVR sont alors transcrits par des enquêteurs spécialisés (analystes audio) pour les bénéfices de l'enquête de sécurité (Fig. 1).

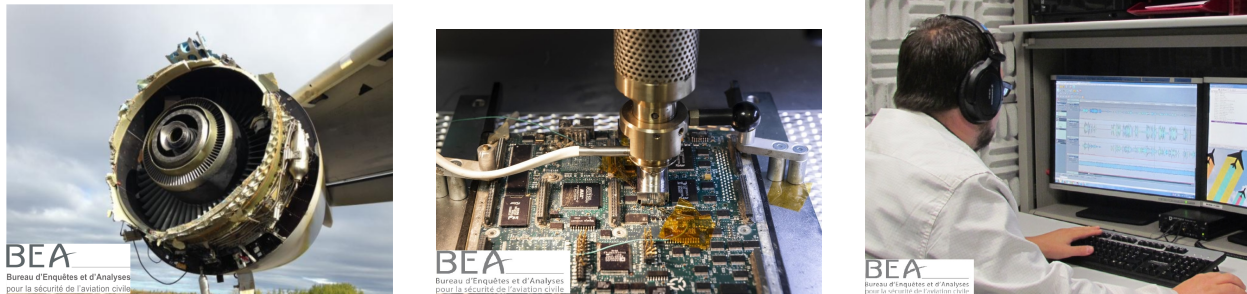


FIGURE 1 – Après un incident ou accident d'avion civil (image à gauche), le CVR est ouvert (image au milieu) et un analyste audio transcrit son contenu (image à droite). Source : [1].

Les limitations de conception des CVR contraignent les constructeurs d'avion à mélanger les sources sonores audibles dans le cockpit et dans les casques des pilotes (émissions et réceptions radio, échanges sur l'intercom, annonces aux passagers, alarmes sonores, etc) en un nombre réduit de canaux audio envoyés vers le CVR [1], qui en fait une acquisition numérique et protège cette donnée en cas d'accident. Ainsi l'activation simultanée de plusieurs sources sonores peut conduire à une réduction significative de l'intelligibilité des échanges vocaux, voire un masquage complet d'une ou de plusieurs sources audio.

L'extraction d'informations dans les données CVR s'appuie sur l'expérience des analystes audio du BEA/RESEDA et sur leur capacité à détecter des sources sonores enfouies dans des mélanges audio difficiles [2]. Dans nos récents travaux [1], nous avons proposé un modèle de mélange audio dans les CVR par rétro-ingénierie (voir fig. 2) et nous avons montré que les méthodes de la littérature de séparation de sources (SAS) pouvaient être appliquées. La SAS cherche à estimer un ensemble de signaux inconnus à partir de mélanges de ceux-ci, le mélange étant lui-aussi inconnu [3].

Dans le cadre du projet ANR BLeRIOT qui finance ce sujet de stage, nous ne souhaitons pas développer des méthodes d'apprentissage profond nécessitant une grande base d'apprentissage. Au contraire, les contraintes des

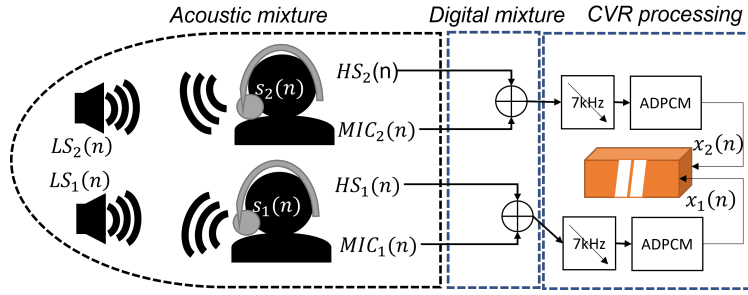


FIGURE 2 – Processus de mélange dans le CVR. Source : [1].

enquêtes de sécurité nous poussent à développer des approches qui soient non ou faiblement supervisées et qui soient interprétables.

Les méthodes de factorisation matricielle non-négative (NMF) ont constitué l'état de l'art avant l'apparition du deep learning au début des années 2010, et remplissent les deux critères ci-dessus. Cependant, elles ne fournissent pas toujours de bonnes performances de séparation lorsqu'on les applique à des signaux CVR. Le principe des techniques déroulées ou dépliées permet de transformer des approches classiques itératives en méthodes d'apprentissage profond et permettent un bon compromis entre performance, interprétabilité du modèle et taille de la base d'apprentissage [4]. Ces techniques ont été combinées à la NMF pour un certain nombre d'applications [5–8], avec des spécificités propres aux méthodes proposées.

Dans le cadre de ce stage, nous proposerons des méthodes de NMF dépliées que nous appliquerons aux données issues de CVR. En particulier, nous avons identifié des cas où le modèle proposé dans [1] n'est plus valide, soit à cause d'effets de saturation des canaux audio dus au système audio avionique, soit à cause d'effets de codage avec perte dus à un modèle de CVR particulier. Ces différents effets ont pu être étudiés séparément dans la littérature – par exemple [9, 10] pour le déclippage audio ou [11] pour les effets de codage sur la qualité de séparation – mais jamais à notre connaissance conjointement ni pour les signaux CVR. Un point important du stage consistera à comparer les performances de méthodes classiques ou dépliées de NMF pour de tels scénarios.

Contexte du stage :

Ce stage est financé par le projet ANR ASTRID « BLeRIOT » et sera encadré dans la nouvelle équipe LoRAL (Low-Rank Approximation and Learning) du LISIC. Cette équipe compte actuellement 8 enseignant-chercheurs, 2 enseignant-chercheurs émérites, 3 chercheurs post-doctorants et 11 doctorants. Une grosse partie de l'équipe travaille sur des problématiques de type SAS et le ou la stagiaire recruté(e) pourra interagir avec l'ensemble du personnel. Le stage aura lieu dans l'antenne de Longuenesse du LISIC. Le laboratoire est situé en plein cœur du Parc Naturel Régional « Caps et Marais d'Opale », à proximité de Lille, l'Angleterre, la Belgique et l'Europe du Nord. Ses locaux à Longuenesse sont situés à côté de la résidence universitaire du CROUS et proche de toutes les commodités (tout peut être réalisé à pied ou en bus). L'équipe pédagogique peut aussi mettre en contact le ou la stagiaire recruté(e) avec les loueurs privés.

Candidatures :

Etudiant en sciences de données (traitement du signal et des images, informatique avec un focus autour de l'intelligence artificielle / apprentissage automatique, mathématiques appliquées), vous êtes curieux et très à l'aise en programmation (Matlab et/ou Python). Vous lisez et parlez l'anglais courant sans difficulté. Vous êtes capable d'expliquer clairement vos travaux, même à des publics non-experts. Bien que non requis, une première expérience en factorisation matricielle (par exemple, en décomposition matricielle ou tensorielle, en séparation de sources, en apprentissage de dictionnaire) sera appréciée. **Les candidats doivent impérativement être de nationalité française ou citoyens d'un état membre de l'Union Européenne, d'un état faisant parti de l'Espace Economique Européen ou de la Confédération Suisse.**

Pour candidater, merci d'envoyer un mail à {gilles.delmaire, matthieu.puigt} [at] univ-littoral.fr en y annexant les documents pouvant supporter votre candidature :

- CV ;
- Lettre de motivation ;
- Relevés de notes depuis la Licence 3 ;
- deux lettres de recommandation ou les noms et moyens de contact de deux référents académiques.

Références

- [1] Matthieu Puigt, Benjamin Bigot, and Hélène Devulder. Introducing the “cockpit party problem” : Blind source separation enhances aircraft cockpit speech transcription. *Journal of the Audio Engineering Society*, 73(1/2), 2025.

- [2] BEA. [Guide sur la vérification des enregistrements CVR](#).
- [3] Pierre Comon and Christian Jutten, editors. *Handbook of Blind Source Separation : Independent Component Analysis and Applications*. Elsevier, 2010.
- [4] Vishal Monga, Yuelong Li, and Yonina C Eldar. Algorithm unrolling : Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Process. Mag.*, 38(2) :18–44, 2021.
- [5] Scott Wisdom, Thomas Powers, James Pitton, and Les Atlas. Deep recurrent nmf for speech separation by unfolding iterative thresholding. In *Proc. WASPAA'17*, pages 254–258. IEEE, 2017.
- [6] Rami Nasser, Yonina C Eldar, and Roded Sharan. Deep unfolding for non-negative matrix factorization with application to mutational signature analysis. *Journal of computational biology*, 29(1) :45–55, 2022.
- [7] Christophe Kervazo, Abdelkhalak Chetoui, and Jérémy E Cohen. Deep unrolling of the multiplicative updates algorithm for blind source separation, with application to hyperspectral unmixing. In *Proc. EUSIPCO'24*, pages 656–660. IEEE, 2024.
- [8] Sheng Shu, Ting-Zhu Huang, Jie Huang, and Gemine Vivone. Nesterov-accelerated non-negative matrix factorization unrolling network for hyperspectral unmixing. *Neurocomputing*, page 131202, 2025.
- [9] Amir Adler, Valentin Emiya, Maria G Jafari, Michael Elad, Rémi Gribonval, and Mark D Plumbley. Audio inpainting. *IEEE Audio, Speech, Language Process.*, 20(3) :922–932, 2011.
- [10] Çağdaş Bilen, Alexey Ozerov, and Patrick Pérez. Joint audio inpainting and source separation. In *Proc. ICA/LVA'15*, pages 251–258. Springer, 2015.
- [11] Matthieu Puigt, Emmanuel Vincent, Yannick Deville, Anthony Griffin, and Athanasios Mouchtaris. Effects of audio coding on ICA : an experimental study. In *Proc. IEEE ECMSM'13*, Toulouse, France, June 24–26, 2013.