

Ph.D. thesis opening

Human-in-the-Loop Audio Source Separation for Aircraft Cockpit Recordings

Keywords:

Cockpit Voice Recorder; Sounds; Speech; Blind Source Separation; Human-in-the-Loop Processing; Nonnegative Matrix or Tensor Factorization; Informed Source Separation

Description:

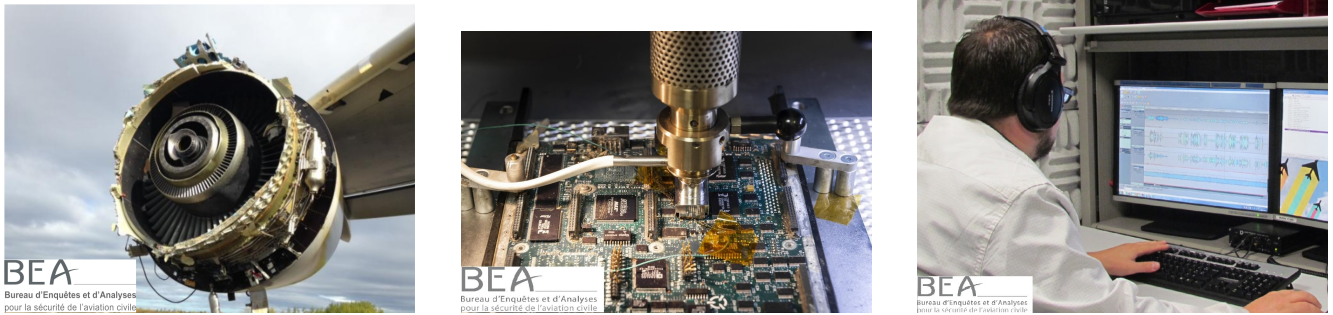


Figure 1: After a civil aircraft incident / accident (left), the CVR is opened (middle), and an audio analyst transcribes its content (right). Source: [1].

Public and state transportation aircraft are fitted with two crash-survival flight recorders—also known as “black boxes”—i.e., the Cockpit Voice Recorder (CVR) and the Flight Data Recorder. Both need to be retrieved and analyzed by air accident authorities in case of incident or accident. 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 CVR investigations, for civil and State aircrafts, respectively. CVR contents are “manually” transcribed by specialized investigators (a.k.a. audio analysts) for the benefits of the safety investigation (Fig. 1).

In a CVR recording, the causes of speech intelligibility degradation are numerous. In particular, the CVR design itself generates a significant amount of superimposed—a.k.a. mixed—speech signals over the audio channels which are simultaneously recorded. Moreover, in case of an aircraft accident or incident, superimposed speech signals are more likely to occur—since voice and cockpit sound activities become denser—which may yield to the loss of crucial information for the safety investigators. In our recent work [1], we reverse-engineered the CVR audio mixing model (see Fig. 2) and we found that state-of-the-art blind source separation (BSS) algorithms could be applied. BSS is a generic problem which aims to estimate unknown source signals from observed ones while the propagation channels from the sources to the sensors are also unknown [2]. We noticed that *classical* BSS algorithms¹ could help the audio analyst to transcribe a CVR recording. In particular, allowing the audio analyst to listen the outputs of different methods significantly helped him in his tasks. However, there remained some cases where these classical techniques were not helpful.

The objective of this Ph.D. thesis is two-fold.

1. First, we aim to develop BSS methods which are providing a sufficient performance while not requiring too much energy to that end [5]. For that purpose, we will propose Human-in-the-Loop BSS methods which will be based on the audio-analyst—BSS interactions. In particular, the goal is to first let the analyst use simple

¹While they were found to provide a significant BSS enhancement for other applications [3,4], deep learning BSS techniques cannot be used in the context of safety investigations.

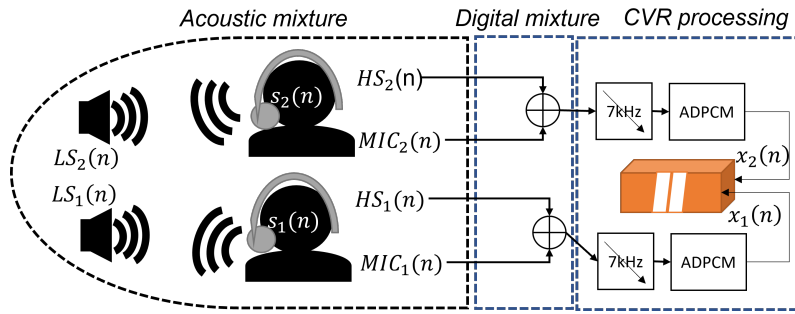


Figure 2: Cockpit sound mixing process in the CVR. Source: [1].

yet efficient BSS algorithms, and then to complexify the BSS method (and allow it more computational time) if the obtained BSS output is unsatisfactory. The latter will be measured by both objective and subjective criteria. Adding information in BSS will be the first way to improve the BSS method, as it was found to be useful for other applications [6–8].

2. The second objective of the Ph.D. thesis is to be able to jointly process all the CVR channels. Indeed, one microphone named Cockpit Area Microphone (CAM) was not investigated in [1], mainly because it is sampled at 12 kHz while the other CVR signals are sampled at 7 kHz. However, the CAM channel provides additional information (e.g., mechanical noise)—mixed with the other sounds in the cockpit—which is usually not recorded in the other channels while being crucial to analyze. While jointly processing data with different resolutions is quite classical for other applications—e.g., hyperspectral imaging [9]—it has been much less investigated for audio signals.

Context of the Ph.D. thesis:

This Ph.D. thesis is funded within the “BLerIOT” ANR ASTRID project (Jan. 2025 – Dec. 2027). The BLerIOT consortium is a balanced group of research laboratories—located in Toulouse (IRIT) and Longuenesse (LISIC)—and of French authorities in charge of aircraft accident or incident (BEA, RESEDA, both being located near Paris). The selected Ph.D. student is expected to start his/her Ph.D. thesis at the same time as the project, in early 2025.

The Ph.D. thesis will take place in the new antenna of LISIC in Longuenesse. This antenna currently has 7 permanent researchers, 1 postdoc researcher, and 5 Ph.D. students (many of them working on BSS). LISIC is located in the heart of the Regional Natural Park of “Caps et Marais d’Opale”, close to Lille, England, Belgium, and Northern Europe. Its premises are next to a student residence and near all the facilities. The recruited Ph.D. student will work in close collaboration with all the BLerIOT partners, in particular with BEA and RESEDA audio analysts. The monthly gross salary is 2.2 k€. The job also brings health insurance and retirement contribution.

Application:

Recently or nearly graduated in the field of data sciences (signal and image processing, computer science with a focus in artificial intelligence / machine learning, applied mathematics), you are curious and are very comfortable in programming (Matlab, Python). You read and speak fluent English with ease. You also own communication skills so that you can explain your work to non-experts of your field, e.g., during project meetings. Although not compulsory, speaking French as well as a first experience in low-rank approximation—e.g., matrix or tensor decomposition, blind source separation, dictionary learning—will be appreciated. **Applicants must be French or citizens of Member State of the European Union, or of a State forming part of the European Economic Area, or of the Swiss Confederation.**

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

- your resume;
- a cover letter;
- your transcripts from the last year of B.Sc to the last year of M.Sc. (if the latter is already available);
- two reference letters or the names and means of contact of two academic advisers.

Applications will be reviewed on a rolling basis until the position is filled.

References

- [1] Matthieu Puigt, Benjamin Bigot, and H el ene Devulder. Introducing the “cockpit party problem”: Blind source separation enhances aircraft cockpit speech transcription. *Journal of the Audio Engineering Society*, to appear.

- [2] Pierre Comon and Christian Jutten, editors. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Elsevier, 2010.
- [3] DeLiang Wang and Jitong Chen. Supervised speech separation based on deep learning: An overview. *IEEE/ACM Trans. Audio, Speech, Language Process.*, 26(10):1702–1726, Oct. 2018.
- [4] Hendrik Purwins, Bo Li, Tuomas Virtanen, Jan Schlüter, Shuo-Yiin Chang, and Tara Sainath. Deep learning for audio signal processing. *IEEE J. Sel. Topics Signal Process.*, 13(2):206–219, May 2019.
- [5] Romain Couillet, Denis Trystram, and Thierry Ménéssier. The submerged part of the AI-ceberg. *IEEE Signal Process. Mag.*, 39(5):10–17, 2022.
- [6] Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel. Informed nonnegative matrix factorization methods for mobile sensor network calibration. *IEEE Trans. Signal Inf. Process. Netw.*, 4(4):667–682, 2018.
- [7] Gilles Delmaire, Mahmoud Omidvar, Matthieu Puigt, Frédéric Ledoux, Abdelhakim Limem, Gilles Roussel, and Dominique Courcot. Informed weighted non-negative matrix factorization using $\alpha\beta$ -divergence applied to source apportionment. *Entropy*, 21(3):253, 2019.
- [8] Sarah Roual, Claude Sensiau, and Gilles Chardon. Informed source separation for turbofan broadband noise using non-negative matrix factorization. In *Forum Acousticum 2023*, 2023.
- [9] Laetitia Loncan, Luis B De Almeida, José M Bioucas-Dias, Xavier Briottet, Jocelyn Chanussot, Nicolas Dobigeon, Sophie Fabre, Wenzhi Liao, Giorgio A Licciardi, Miguel Simoes, et al. Hyperspectral pansharpening: A review. *IEEE Geosci. Remote Sens. Mag.*, 3(3):27–46, 2015.