

# Un enseignement expérimental de traitement du signal audio

Gilles CHARDON<sup>1</sup> José PICHERAL<sup>1</sup> Sarah ROUAL<sup>1,2</sup>

<sup>1</sup>Université Paris-Saclay, CNRS, CentraleSupélec  
Laboratoire des signaux et systèmes  
91190, Gif-sur-Yvette, France.

<sup>2</sup>Safran Tech, Rue des jeunes bois, 78114 Magny-les-Hameaux, France

**Résumé** – Nous présentons ici un cours de traitement de signal audio qui est proposé en deuxième année du cursus ingénieur de CentraleSupélec. Il s’agit d’un module optionnel d’enseignement expérimental se déroulant sur une semaine à plein temps. Cet enseignement s’articule autour de trois thèmes principaux : estimation de réponses impulsionnelles de salle, séparation de sources et localisation de sources. Chaque thème est introduit sur une journée par un cours et un TD le matin, et un TP l’après-midi. Les deux journées restantes sont consacrées à l’approfondissement en binôme d’un de ces thèmes au choix. Pendant la semaine, les étudiants obtiennent des résultats concrets, à partir de codes et données qu’ils ont eux-mêmes produits à l’aide du matériel mis à disposition (microphones de mesure, couples ORTF, réseaux de microphones MEMS, etc.). L’un des objectifs pédagogiques consiste à faire le lien entre les aspects théoriques et expérimentaux suivant la démarche suivante : 1) s’appuyer sur des fondements théoriques, 2) mettre en œuvre des méthodes de traitement, 3) valider sur des données synthétiques, 4) réaliser des acquisitions expérimentales, 5) approfondir les méthodes et améliorer les traitements.

**Abstract** – Here, we present an audio signal processing course offered in the second year of the engineering curriculum at CentraleSupélec. This is an optional, full-time, one-week experimental teaching module. The course is structured around three main topics: room impulse response estimation, source separation, and source localization. Each topic is introduced over a day with a lecture and tutorial in the morning, followed by a practical session in the afternoon. The last two days are dedicated to in-depth study, in pairs, of one chosen topic. Throughout the week, students obtain concrete results using codes and data they generate themselves with the provided equipment (measurement microphones, ORTF pairs, MEMS microphone arrays, etc.). One of the educational objectives is to bridge theoretical and experimental aspects through the following approach: 1) rely on theoretical foundations, 2) implement processing methods, 3) validate on synthetic data, 4) conduct experimental acquisitions, 5) refine methods and improve processing.

## 1 Introduction

Le cursus ingénieur de CentraleSupélec comprend un cours de traitement du signal en promotion complète en première année. Cependant, le temps disponible et l’effectif ne permettent pas d’aborder des applications concrètes des méthodologies vues en cours et travaux dirigés. Nous avons profité de l’opportunité de proposer des enseignements expérimentaux, en deuxième année du cursus, pour concevoir un enseignement de traitement du signal laissant la place à l’exploration, l’expérimentation, l’acquisition et la manipulation de données réelles, dans le domaine du traitement du signal audio. Ce domaine particulier permet d’aborder les notions importantes du traitement du signal avec des applications dont l’intérêt est immédiatement perceptible, et nécessitant un matériel peu onéreux.

L’objectif est de permettre aux étudiants de comprendre et d’établir des liens entre les concepts théoriques du traitement du signal et leurs applications pratiques. Cela implique tout d’abord de leur fournir les notions théoriques en lien avec les applications envisagées sous forme de rappels et d’approfondissements avant de passer à la mise en œuvre pour laquelle on demande aux étudiants la démarche suivante : 1) implémenter les méthodes de traitement, 2) valider les méthodes sur des données synthétiques 3) réaliser des mesures pour valider expérimentalement les données. Cette partie s’apparente à un TP nécessitant un taux d’encadrement important pour avancer rapidement. Enfin une dernière partie d’approfondis-

sement s’apparentant à un mini projet laisse plus de liberté aux étudiants pour travailler sur un thème particulier.

Cet enseignement s’inscrit dans la deuxième année du cursus ingénieur de CentraleSupélec sous la forme d’un module optionnel se déroulant sur une semaine entière à plein temps et s’adressant à des effectifs réduits (10 à 20 étudiants). Deux occurrences sont proposées, en novembre et juin, en concurrence avec des enseignements de sciences humaines et sociales, voire même des vacances. Une troisième occurrence servant de rattrapage pour d’autres enseignements est parfois proposée en juin.

Les étudiants inscrits à ce cours ont généralement reçu des cours sur les fondements du traitement du signal (par ex. le cours de première année à CentraleSupélec) et de statistiques, mais ils ont généralement peu d’expérience pratique dans l’implémentation de traitements numériques de signaux car il n’y a pas de travaux pratiques associés aux cours de fondements. Ils sont cependant familiarisés à l’usage de python, utilisé notamment durant les *coding weeks* de première année. Certains, mais pas tous, ont un intérêt particulier pour l’audio et les applications du traitement du signal.

## 2 Liste de matériel

Le matériel nécessaire à la mise en place de ce module est, pour la plus grande part, disponible chez les fournisseurs de

TABLE 1 : Liste de matériel

- Microphone de mesure omnidirectionnel (Superlux ECM999) (Fig. 1)
- Couple de microphones ORTF (Superlux S502) (Fig. 1)
- Interface audio USB avec entrées XLR (M-Audio M-Track DUO). On prendra garde à activer l'alimentation fantôme.
- Hauts-parleurs (VISATON BF-32), avec enceintes fabriquées par la Fabrique (Fig. 2), fablab de CentraleSupélec, avec fixation sur pied de micro. Tout haut-parleur de qualité raisonnable est adapté.
- Réseau de 32 microphones MEMS  $M\mu$  (Fig. 2) [6] fourni par l'Institut Jean le Rond d'Alembert, qui sera remplacé par des réseaux de  $4 \times 4$  microphones moins performants mais plus maniables (autant mécaniquement qu'informatiquement) et disponibles dans le commerce (miniDSP UMA-16)
- Absorbant acoustique (mousse PU)
- Câble et adaptateurs divers (XLR, jack, etc.)
- Pieds de micros



FIGURE 1 : Couple ORTF et microphone de mesure.

matériel de production audio. Les références données dans la table 1 sont bien entendues indicatives. Ce matériel est également utilisé pour de la médiation scientifique, lors de visites de lycées à CentraleSupélec, ou de présentations des activités de recherches aux étudiants de l'école.

### 3 Contenu

Le module, dans sa forme actuelle, est structuré en trois journées, chacune consacrée à un thème, suivies par deux journées d'approfondissement conclues par une restitution du travail effectué.



FIGURE 2 : Réseau de microphones MEMS et deux haut-parleurs

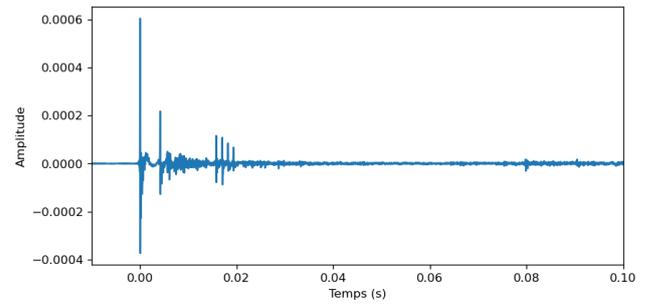


FIGURE 3 : Réponse impulsionnelle estimée de l'atrium du bâtiment Eiffel de CentraleSupélec

Les trois premières journées ont pour thème l'estimation de réponse impulsionnelle, la localisation de sources, et la séparation de sources. Le matin est consacré à un cours théorique et une séance de travaux dirigés. Les concepts introduits sont alors appliqués l'après-midi sur des données réelles produites par les étudiants avec le matériel fourni.

Le choix de ces thèmes a été guidé par les connaissances théoriques des étudiants, les possibilités d'application expérimentale en temps limité, et le matériel disponible.

Le temps disponible le permettant, quasiment aucun code n'est fourni aux étudiants, ce qui leur permet de coder les méthodes intégralement.

#### 3.1 Estimation de réponse impulsionnelle

Le premier thème est l'estimation de réponse impulsionnelle de salle. Sous conditions que l'environnement de propagation ne varie pas, et l'amplitude acoustique reste dans le domaine de linéarité, la propagation des ondes entre deux points de l'espace peut s'écrire comme une convolution. Ainsi, négligeant les effets du matériel, le signal mesuré par un microphone  $y$  pour un signal émis  $x$  s'écrit  $y = h \star x$  avec  $h$  la réponse impulsionnelle entre le microphone et le haut-parleurs.

La réponse impulsionnelle peut s'estimer en émettant un bruit fort et court (clap, fouet), mais une mesure plus précise est obtenue en utilisant des signaux particuliers, par exemple un chirp linéaire. La corrélation entre le signal émis et le signal mesuré est  $\gamma = \gamma_c \star h$  avec  $\gamma_c$  l'autocorrélation du chirp, proche d'une impulsion. Une alternative utilisant des signaux déterministes est les codes de Golay, paires de signaux dont la somme des autocorrélations est une impulsion [2]. L'utilisation de signaux aléatoires est également possible.

La diversité des espaces rencontrés sur le campus (amphithéâtres, halls, escaliers, parking souterrain, etc.) permet de mesurer des réponses très variées et de bien démontrer l'intérêt de la mesure des réponses impulsionnelles en les mettant en lien avec les différences de perception acoustique dans ces environnements. Dans des espaces de grandes dimensions, les premières réflexions sont aisément identifiables (voir figure 3), permettant de faire le lien avec la géométrie de l'environnement de propagation.

Cette journée permet de réviser les concepts de base de traitement du signal tels que : convolution, intercorrélations aux sens déterministe et aléatoire, processus aléatoires, bruit blanc, formule des interférences.

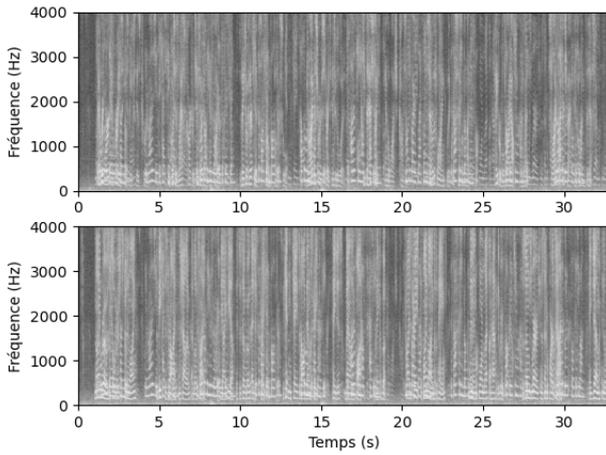


FIGURE 4 : Spectrogrammes des sources mélangées.

### 3.2 Séparation de sources

Le deuxième thème est la séparation de sources. L'analyse par composantes indépendantes [3] est d'abord introduite, sous sa forme simple permettant de traiter des mélanges instantanés. Le but est de retrouver un vecteur aléatoire  $\mathbf{X}$  de composantes indépendantes à partir d'observations  $\mathbf{Y} = \mathbf{MX}$  avec  $\mathbf{M}$  une matrice de mélange inconnue. L'algorithme consiste à d'abord blanchir les observations, donnant un vecteur  $\mathbf{Z}$  de composantes décorrélées, puis à appliquer une matrice de rotation les rendant le plus indépendantes possible.

L'indépendance étant difficile à quantifier, elle est généralement remplacée par une mesure de non-gaussianité, par exemple le kurtosis. La rotation éloignant le plus le kurtosis de celui d'un vecteur gaussien est alors choisie. Cette méthode simple est limitée au cas de mélange instantané. Les variantes plus générales sont trop complexes pour être abordées durant le temps imparti.

La méthode DUET [5] est ensuite introduite, permettant de traiter le cas indéterminé, avec des mélanges non nécessairement instantanés. L'utilisation de cette méthode est également l'occasion d'utiliser la transformée de Fourier à court terme. Sous l'hypothèse que le support des différentes sources soient disjoints dans le plan temps-fréquence (un exemple de représentation temps-fréquence est donné figure 4), le tracé des coefficients dans le plan voie 1 / voie 2 permet de mettre en évidence des droites pour chacune des sources, dont les pentes sont liées aux coefficients de mélange des sources dans les deux canaux, cf. figure 5. Un simple clustering sur la base d'un histogramme des angles (figure 6) des points dans le plan voie 1 / voie 2 permet d'identifier le nombre de sources et de les séparer.

L'ICA permet d'aborder la modélisation aléatoire de signaux sous un autre angle, la loi des modèles étant d'importance capitale, en comparaison avec la modélisation stationnaire au sens large utilisée pour l'estimation de réponse impulsionnelle. Comme évoqué plus haut, la méthode DUET est l'occasion d'introduire la transformée de Fourier à court terme.

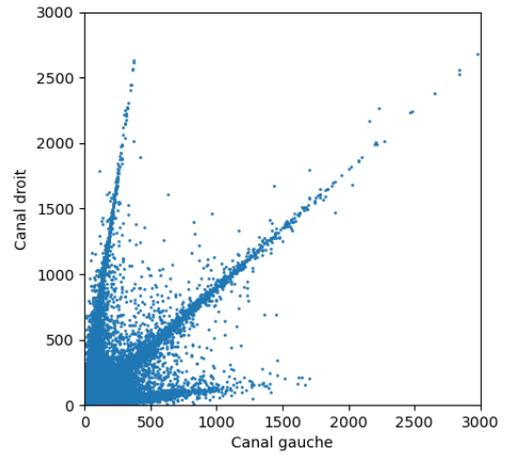


FIGURE 5 : Valeurs absolues des coefficients dans le plan voie 1 / voie 2.

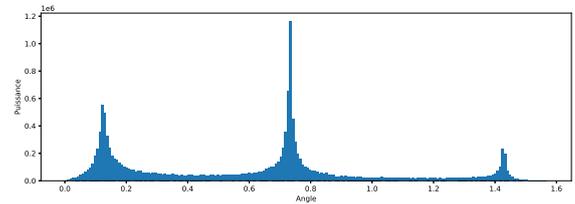


FIGURE 6 : Histogrammes des angles pondérées par les puissances.

### 3.3 Localisation de sources

La localisation de sources acoustiques est abordée dans la troisième journée. Il s'agit de l'application la plus délicate car les étudiants ne sont pas familiers avec les modèles spatiotemporels des signaux. De plus, les étudiants doivent être accompagnés d'assez près pour l'implémentation des traitements qui est un peu moins évidente que pour les applications précédentes.

La méthode de la formation de voies (*beamforming*) [4] est introduite comme cas particulier du maximum de vraisemblance [1], vu en cours de statistique de première année.

Dans sa formulation pour signaux déterministes, étant donné un vecteur de coefficients de Fourier sur les microphones du réseau  $\mathbf{m}$  et le modèle de acoustique de source  $\mathbf{g}(\mathbf{x})$  (fonction de Green), le critère de la formation de voie s'écrit

$$b(\mathbf{x}) = \left| \left( \frac{\mathbf{g}(\mathbf{x})}{\|\mathbf{g}(\mathbf{x})\|} \right)^H \mathbf{m} \right|^2 \quad (1)$$

On notera que dans des conditions habituelles de propagation, l'utilisation d'un modèle de champ libre n'empêche pas d'obtenir des résultats satisfaisants, au moins pour un TP, comme sur la figure 7.

Les notions d'estimation paramétrique, de maximum de vraisemblance, de vecteurs aléatoires et de propagation des ondes sont revues pendant cette journée.

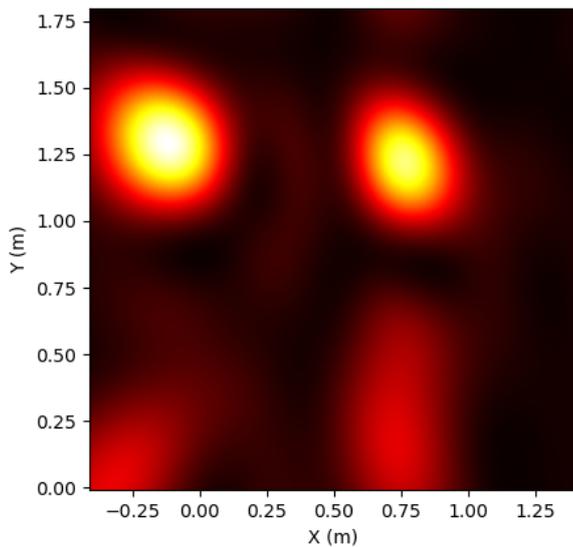


FIGURE 7 : Critère de la formation de voie  $b(\mathbf{x})$  à 2000 Hz pour la localisation de deux sources à 1.5m du réseau.

### 3.4 Approfondissement

La semaine se termine par des mini-soutenances où les étudiants présentent les méthodes et les résultats obtenus lors de la partie d'approfondissement, ceci permet d'avoir des soutenances variées et originales. Les étudiants participent à l'évaluation.

Voici quelques sujets d'approfondissement ayant été abordés depuis la création du cours :

- pour la réponse impulsionnelle : identification de la réponse du système de mesure (HP et microphone), calcul d'indicateurs sur des réponses impulsionnelles (RT60, clarté, etc.), mise en relation des réflexions avec la géométrie de la salle
- pour la localisation de sources : méthodes MVDR, MUSIC, imagerie 2D et 3D, évaluation des performances (erreur d'estimation, résolution)
- pour la séparation de sources : fastICA, DUET avec couple ORTF, NMF multicanal.

## 4 Retours des étudiants, difficultés rencontrées et évolutions du cours

Les commentaires des étudiants sur le module sont globalement très positifs (voire Table 2), avec comme point fort le temps mis à disposition pour implémenter soi-même et tester les méthodes introduites dans le module.

Des difficultés sont cependant fréquemment rencontrées par les étudiants, qui ont parfois du mal à utiliser efficacement les structures de données de `numpy` et manipuler des données réelles. L'utilisation des réseaux de microphone pose parfois problème, nécessitant (sous Windows) l'installation de pilotes.

Le cours a subi une réduction des heures de cours, avec la perte de 6 heures d'encadrement pendant l'approfondissement. Cette perte s'est ressentie dans la qualité des travaux présentés, en baisse par rapport aux premières itérations, et est source de frustration pour les étudiants.

TABLE 2 : Quelques retours étudiants, positifs ou pas

« J'ai apprécié faire quelque chose de plus technique pour une fois et de pouvoir voir des utilisations plus concrètes de codes python. »  
 « J'ai pas du tout aimé le cours de traitement du signal en 1ère année mais j'ai bien aimé ce cours. »  
 « Le faible effectif permet de mieux comprendre puisqu'on peut poser plus de questions pendant les TD. Le contenu est vraiment intéressant, j'ai vraiment eu l'impression d'acquérir de nouvelles compétences »  
 « Pas beaucoup de temps pour finir la présentation de fin. »  
 « Il est agréable de manipuler un peu et d'être un peu moins guidé que d'habitude »

Ce constat va probablement amener à l'abandon des approfondissements, remplacés par le thème du filtrage adaptatif. L'évaluation sera alors basée sur le travail effectué en travaux pratiques.

## 5 Conclusion

De notre point de vue, l'investissement en matériel et temps de préparation du cours est largement rentabilisé. En plus de la satisfaction des étudiants le suivant, le montage du cours a également été l'occasion de mettre en place des démonstrations de traitement de signal illustrant les cours magistraux du module de première année, et nous a permis de monter en compétence nous-mêmes sur les applications expérimentales des méthodes enseignées.

## Références

- [1] Gilles CHARDON : Theoretical analysis of beamforming steering vector formulations for acoustic source localization. *Journal of Sound and Vibration*, 517:116544, janvier 2022.
- [2] S. FOSTER : Impulse response measurement using Golay codes. In *ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 11, pages 929–932, avril 1986.
- [3] A. HYVÄRINEN et E. OJA : Independent component analysis : algorithms and applications. *Neural Networks*, 13(4):411–430, juin 2000.
- [4] H. KRIM et M. VIBERG : Two decades of array signal processing research : the parametric approach. *IEEE Signal Processing Magazine*, 13(4):67–94, juillet 1996.
- [5] Scott RICKARD : The DUET Blind Source Separation Algorithm. In Shoji MAKINO, Hiroshi SAWADA et Te-Won LEE, éditeurs : *Blind Speech Separation*, pages 217–241. Springer Netherlands, Dordrecht, 2007. Series Title : Signals and Communication Technology.
- [6] C. VANWYNSBERGHE, R. MARCHIANO, F. OLLIVIER, P. CHALLANDE, H. MOINGEON et J. MARCHAL : Design and implementation of a multi-octave-band audio camera for realtime diagnosis. *Applied Acoustics*, 89:281–287, mars 2015.