

Separation de sources d'électrocardiogrammes

LEWY Nathan, SAINTAVIT Pierre

April 9, 2025

Abstract

Le signal électrocardiographique (ECG) reflète l'activité électrique du cœur et est couramment utilisé pour diagnostiquer et surveiller divers troubles cardiaques. Dans certaines situations cliniques, notamment lors du suivi prénatal, le signal enregistré contient une superposition de deux sources : l'ECG de la mère (MECG) et celui du fœtus (FECG).

L'objectif principal de ce projet est de séparer ces deux composantes à partir de mesures mixtes, issues de plusieurs électrodes placées sur le corps de la mère, dans le but de détecter avec précision les **complexes QRS** qui constituent les phases les plus caractéristiques du cycle cardiaque. Ces complexes, en particulier les pics R, permettent d'estimer la fréquence cardiaque et de diagnostiquer d'éventuelles anomalies.

Pour ce faire, nous avons étudié trois approches différentes:

- le **Matching Pursuit**, une méthode de décomposition sur dictionnaire adaptée aux signaux impulsionnels ;
- l'**Analyse en Composantes Indépendantes** (ICA), permettant de séparer des sources supposées statistiquement indépendantes ;
- une approche combinant **Analyse en Composantes Principales** (PCA) et étude de l'**autocorrélation**, pour extraire la périodicité des QRS.

Contents

1	Première approche: Matching Pursuit	2
1.1	Principe	2
1.2	Implémentation	2
1.3	Problèmes rencontrés	2
1.4	Conclusion	2
2	Deuxième approche : Analyse en composantes indépendantes	3
2.1	Principe de l'ICA	3
2.2	Principe de la FastICA	3
2.3	Implémentation sur l'ECG	3
2.4	Résultats	4
2.5	Conclusion	5
3	Troisième approche : Analyse en composantes principales et autocorrélative	6
3.1	Décomposition en composantes principales	6
3.2	Analyse de l'autocorrélation	6
3.3	Tableau des résultats	7
4	Conclusion	8

1 Première approche: Matching Pursuit

1.1 Principe

Le **Matching Pursuit** est une méthode gloutonne de décomposition de signaux. L'idée est de projeter successivement le signal sur des éléments (appelés *atomes*) d'un dictionnaire de fonctions. Dans notre cas, on choisit des *peignes de Dirac* parfaits, simulant les impulsions cardiaques (pics R) dans un ECG.

On cherche à approximer le signal $s(t)$ comme une somme :

$$s(t) \approx \sum_{n=1}^N a_n g_{\gamma_n}^*(t)$$

où :

- $g_{\gamma_n}^*(t)$ est un atome de norme 1 du dictionnaire paramétré par γ_n (fréquence, largeur, amplitude, position... suivant le dictionnaire),
- a_n est le coefficient de projection du signal sur cet atome.

1.2 Implémentation

Le dictionnaire utilisé est composé de **peignes** paramétrisés (que l'on divise ensuite par leur norme pour être de norme 1):

$$g_{\gamma}(t) = \sum_k \delta(t - (kT + \phi))$$

où T est la période du peigne, choisie en fonction de la fréquence cardiaque et ϕ le déphasage. Dans le cas idéal, les pics des complexes QRS du fœtus et de la mère coïncident parfaitement avec les deux premières projections sur le dictionnaire.

1.3 Problèmes rencontrés

- Des **pics parasites** apparaissent dans le résidu, ce qui rend l'extraction de composantes difficile.
- L'idée de ne garder que les deux premières composantes extraites (mère et fœtus) ne fonctionne pas bien.
- Le recours à l'**Orthogonal Matching Pursuit**, censé améliorer la qualité en rendant les projections orthogonales, ne donne pas de meilleurs résultats.
- En changeant les peignes pour des sommes de triangles de largeur variable, le problème ne disparaît pas.
- L'échec vient du fait que les rythmes cardiaques mère/fœtus (MHR/FHR) suivent une **distribution continue de fréquence**, ce qui rend l'hypothèse de peigne fixe trop simpliste.

1.4 Conclusion

Cette méthode, bien qu'intuitive et directement liée à la nature impulsionnelle des signaux ECG, manque de robustesse dans un contexte de séparation de sources. Les composantes se mélangent, et les choix de dictionnaire contraignent trop la modélisation. Impossible d'aboutir à des résultats concluants.

2 Deuxième approche : Analyse en composantes indépendantes

2.1 Principe de l'ICA

L'ICA (Independent Component Analysis) vise à retrouver des sources $s_1(t), s_2(t), \dots, s_n(t)$ indépendantes à partir de signaux mixtes observés $x(t)$.

On suppose que le mélange est linéaire :

$$X(t) = As(t)$$

où :

- $X(t)$ est la matrice des signaux mesurés
- A est une matrice de mélange inconnue,
- $s(t)$ est le vecteur des sources **indépendantes** à retrouver.

L'objectif de l'ICA est d'estimer une matrice W telle que :

$$s(t) = Wx(t)$$

2.2 Principe de la FastICA

Pour ce faire, la FastICA met en place l'algorithme suivant :

Algorithm 1 Algorithme d'extraction de caractéristiques

```
1: for  $p = 1$  to  $C$  do
2:    $\mathbf{w}_p \leftarrow$  Random vector of length  $N$ 
3:   while  $\mathbf{w}_p$  changes do
4:      $\mathbf{w}_p \leftarrow \frac{1}{M} \mathbf{X} g(\mathbf{w}_p^T \mathbf{X})^T - \frac{1}{M} g'(\mathbf{w}_p^T \mathbf{X}) \mathbf{1}_M \mathbf{w}_p$ 
5:      $\mathbf{w}_p \leftarrow \mathbf{w}_p - \sum_{j=1}^{p-1} (\mathbf{w}_p^T \mathbf{w}_j) \mathbf{w}_j$ 
6:      $\mathbf{w}_p \leftarrow \frac{\mathbf{w}_p}{\|\mathbf{w}_p\|}$ 
7:   end while
8: end for
9:  $\mathbf{W} \leftarrow [\mathbf{w}_1, \dots, \mathbf{w}_C]$ 
10:  $\mathbf{S} \leftarrow \mathbf{W}^T \mathbf{X}$ 
```

Où X a été centrée et normalisée au préalable et :

$$\begin{aligned} f(u) &= \log \cosh(u), \\ g(u) &= \tanh(u), \\ g'(u) &= 1 - \tanh^2(u). \end{aligned}$$

2.3 Implémentation sur l'ECG

On prend :

- $x(t) \in \mathbb{R}^{4 \times T}$ est des observations. Chaque $x_i(t) \in \mathbb{R}^T$ est formé en prenant la moyenne sur 6 à 8 électrodes consécutives. Ainsi, les observations sont lissées.
- $s(t) \in \mathbb{R}^{2 \times T}$ où T est la durée entière du signal. Il y a bien deux composantes : une pour le fœtus, une pour la mère.

On obtient des résultats sur un des signaux de la base de données :

On a obtenu les abscisses des maximums avec la fonction "find peaks" de scipy. On a choisi des paramètres dans cette fonction pour sélectionner des pics qui sont assez éloignés les uns des autres et qui valent au moins la moitié du plus grand pic du signal.

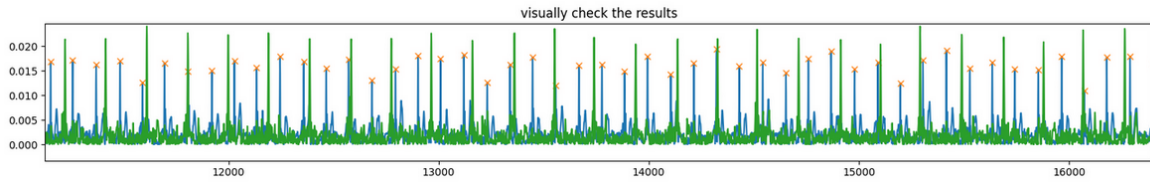
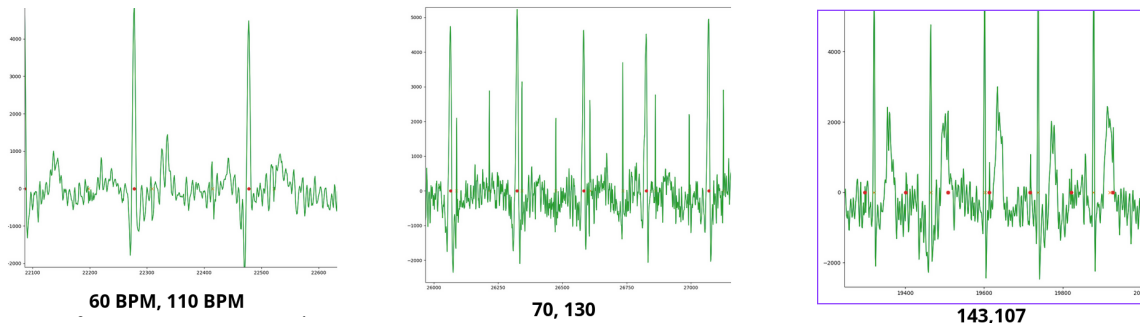


Figure 1: Résultat de la séparation des signaux.

2.4 Résultats

On essaye l'algorithme sur des cas stationnaires et on fait varier les BPM :

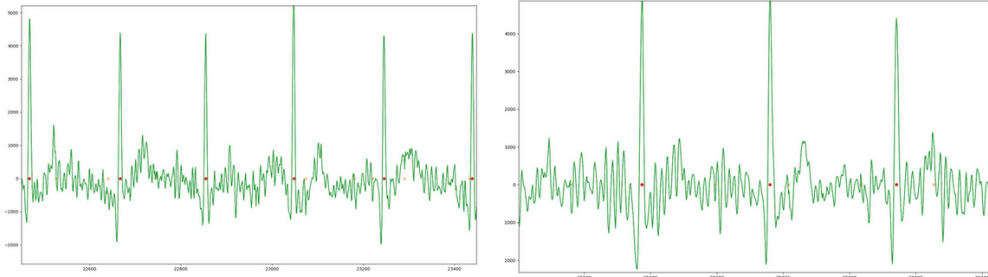


On remarque que plus les fréquences sont proches, moins la détection est précise.

On essaye aussi avec différents niveaux de bruit :

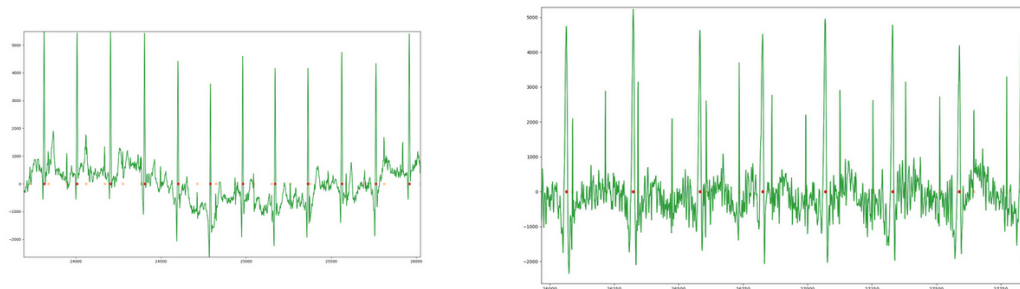
06

03



12

09



On observe une sensibilité au bruit.

2.5 Conclusion

L'ICA est facile à implémenter et donne de bons résultats sur des signaux simples. Mais pour des signaux très bruités ou des fréquences similaires entre la mère et le fœtus, l'algorithme fonctionne moins bien.

Il y a plusieurs pistes d'amélioration :

- chercher $s(t) \in \mathbb{R}^{3 \times T}$ voire $s(t) \in \mathbb{R}^{4 \times T}$
- fenêtrer le signal afin de pouvoir traiter les signaux non stationnaires
- faire un choix d'électrodes plus pertinent pour avoir de meilleures observations

Cependant, nous allons présenter une approche qui fonctionnait mieux que l'ICA et nous avons donc décidé de ne pas poursuivre cette piste.

3 Troisième approche : Analyse en composantes principales et autocorrélative

3.1 Décomposition en composantes principales

La PCA (Principal Component Analysis) est utilisée [1] pour transformer les données vers un nouvel espace où les axes (composantes principales) sont non corrélés et ordonnés par variance décroissante.

Soit $X = \begin{pmatrix} x_1 \\ \cdots \\ x_N \end{pmatrix} \in \mathcal{M}_{N,T}(\mathbb{R})$ la matrice des $N = 34$ signaux x_i issus des différentes électrodes durant chacun T échantillons. On a centré chaque x_i et l'on a choisi de ne **pas** les normer à 1 (en prenant $\tilde{x}_i = \frac{x_i}{\text{std}(x_i)}$) puisque l'on a constaté que cela n'affecte pas les résultats expérimentaux. On prend en effet tous 34 les signaux issus des électrodes car la littérature [2] constate que les résultats par BSSpca s'améliorent avec le nombre de signaux traités. On calcule la matrice de covariance :

$$\Sigma = \frac{1}{n} X X^\top$$

Puis on diagonalise Σ :

$$\Sigma = P D P^\top$$

où :

- P contient les vecteurs propres (axes principaux),
- D est une matrice diagonale contenant les valeurs propres (variances le long de chaque axe).

On projette les données dans la base des composantes principales:

$$X' = P^\top X$$

On a remarqué expérimentalement que notre analyse était plus fine avec 3 composantes au lieu de 2. En effet en prenant les composantes associées au 2 valeurs propres les plus grandes on se retrouve parfois avec les complexes QRS de la mère comme première composante et les relaxations des ventricules de la mère en deuxième composante. Dans ces cas, les complexes QRS du fœtus étaient systématiquement en troisième composante.

3.2 Analyse de l'autocorrélation

On a choisi d'analyser l'autocorrélation. En effet, au cours d'une fenêtre d'analyse, les pics des ECG sont proches de peignes de Dirac idéaux et dans le meilleur cas, l'autocorrélation est non nulle seulement pour les multiples de la fréquence cardiaque. L'estimation de Bartlett (choisie pour sa variance plus faible que l'estimateur de Blackman-Tukey) de l'autocorrélation du signal $x_i(t)$ avec T échantillons est :

$$R(\tau) = \frac{1}{T} \sum_{k=0}^{T-\tau} x(k)x(k+\tau) dt$$

Dans la fenetre de recherche des frequences cardiaques (on peut convertir un lag τ en fréquence connaissant le fréquence d'échantillonnage du signal), c'est-à-dire entre 50 bpm et 180 bpm, on cherche le pic d'autocorrélation de fréquence f_0 la plus haute pour la composante principale en cours de traitement. On force ensuite la recherche des pics sur la composante tels que ces pics soient d'une fréquence autour de f_0 avec une marge d'erreur de 80%.

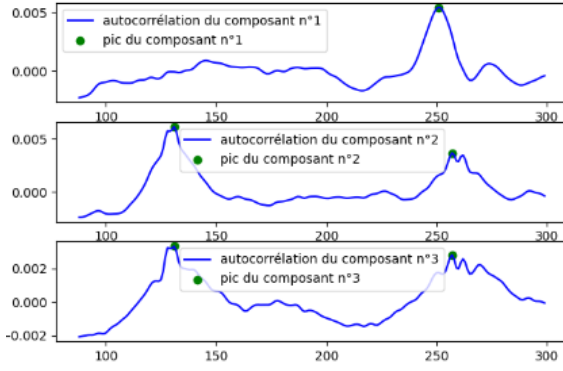


Figure 2: Estimations de l'autocorrélation sur les 3 composantes principales

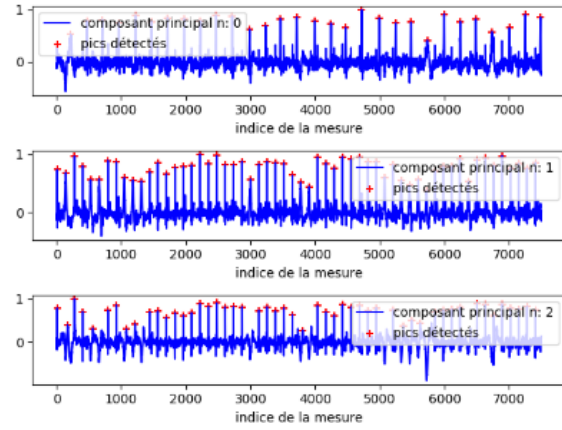


Figure 3: Forçage de la recherche des pics à la fréquence associée

Pour cette simulation au dessus, il se trouve que la deuxième et troisième composantes donnent des pics très similaires. Pour la tâche de classification, il suffit de renvoyer la liste des indices des pics des deux premières composantes. Nous avons donc essayé de définir une métrique calculant la similarité de ces listes pour renvoyer seulement deux listes de complexes QRS, mais nous n'avons pas abouti. Nous avons alors choisi de renvoyer les listes qui avaient le meilleur f1 score (voir plus bas) en faisant l'hypothèse que le médecin en charge du patient se rendra compte si deux des trois composantes renvoyées sont similaires.

3.3 Tableau des résultats

Pour chaque complexe QRS détecté dans cette méthode, on calcule le taux de faux positifs, de vrais positifs et de faux négatifs suivant si le complexe QRS détecté est dans une fenêtre de 150ms (mère) ou 50ms (foetus) du complexe QRS réel.

Toutes les données sont sur des fenêtres de 30s avec un **SNR de 3dB** (qui est la plus élevée non nulle dans cette database), prises aléatoirement parmi la base de données synthétique **Fetal ECG Synthetic Database**. Celle ci contient différentes simulations, avec des scénarios différents.

Toutefois, au vu du temps de calcul requis, nous avons limité le nombre d'itérations (soit de fenêtres aléatoirement choisies) à 500, donc une **variance** des résultats peut encore être légèrement présente.

500 iterations, 30s windows, randomized through synthetic database	c0 Baseline	c1 Foetal movement	c2 MHR/FHR variations	c3 Uterine contraction	c4 Ectopic beats	c5 twin pregnancy
Mother sensitivity	0.99	0.99	0.99	0.91	0.99	0.99
Mother ppv	1	1	1	0.63	1	1
Mother f1-score	0.99	0.99	0.98	0.88	0.99	0.99
Foetus sensitivity	0.88	0.9	0.93	0.61	0.88	0.68
Foetus ppv	1	1	1	0.17	1	0.68
Foetus f1-score	0.89	0.92	0.95	0.62	0.90	0.69

Figure 4: Results of the method across fecgsyn database

On déduit de ces résultats:

- **le cas de base fonctionne :** Le modèle classifie les complexes QRS comme attendu
- **le cas du mouvement foetal fonctionne :** Le modèle est robuste contre les bruits secondaires
- **le cas des variations des fréquences cardiaques et des battements ectopiques fonctionne :** Le modèle est robuste contre les variations des fréquences cardiaques
- **le cas des contractions uterines dyfonctionne:** Le modèle est moins robuste contre les bruits de forte amplitude issus de la mère, qu'il confond avec les complexes QRS du fœtus.
- **le cas des jumeaux dyfonctionne:** Le modèle est moins robuste lorsqu'il y a plusieurs fœtus, puisqu'il n'en détecte qu'un seul, qui se trouve parfois ne pas être celui pris pour référence des complexes QRS.

4 Conclusion

Nous avons pu évaluer différentes méthodes pour la détection des complexes QRS. En particulier, notre méthode dépasse légèrement en F1-score les résultats obtenus par **BSSpca** dans la littérature [2] pour le cas des battement ectopiques bien qu'elle performe moins bien notamment pour les contractions utérines

References

- [1] A.K. Nandi E. Bacharakis and V. Zarzoso. Foetal ecg extraction using blind source separation methods. *Signal Processing*, 2025. Signal Processing Division, Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK.
- [2] Fernando Andreotti et al. An open-source framework for stress-testing non-invasive foetal ecg extraction algorithms. *Physiological Measurement*, 37:627, 2016.