HAL open science

Effective signal processing methods for robust respiratory rate estimation from photoplethysmography signal

Safa Cherif

To cite this version:

Safa Cherif. Effective signal processing methods for robust respiratory rate estimation from photoplethysmography signal. Signal and Image processing. Ecole nationale supérieure Mines-Télécom Atlantique, 2018. English. NNT : 2018IMTA0094 . tel-02160161

HAL Id: tel-02160161

https://theses.hal.science/tel-02160161v1

Submitted on 19 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPERIEURE MINES-TELECOM ATLANTIQUE
BRETAGNE PAYS DE LA LOIRE - IMT ATLANTIQUE
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies
de l'Information et de la Communication
Spécialité : *Signal, Image, Vision*

Par
Safa CHERIF

Effective signal processing methods for robust respiratory rate estimation from photoplethysmography signal

Thèse présentée et soutenue à IMT Atlantique- Brest, le 12 octobre 2018
Unité de recherche : LabSTICC UMR CNRS 6285
Thèse N° : 2018IMTA0094

Rapporteurs avant soutenance :

Valérie Louis-Dorr	Professeur à l'Université de Lorraine, Nancy
Laurent Navarro	Maître de Conférences HDR à ENSM, Saint Etienne

Composition du Jury :

Président :	Lotfi Senhadji	Professeur à l'université de Rennes 1
Examinateurs :	Julien De Jonckheer	CHRU de Lille, Lille
	Laurent Navarro	Maître de Conférences HDR à ENSM, Saint Etienne
	Valérie Louis-Dorr	Professeur à l'Université de Lorraine, Nancy
Dir. de thèse :	Dominique Pastor	Professeur à l'IMT Atlantique, Brest
Co-dir. de thèse:	Erwan L'Her	Professeur à l'UBO, Brest

Remerciements

En tout premier lieu, je tiens à remercier mon directeur de thèse M. Dominique Pastor, professeur à IMT Atlantique Bretagne - Pays de la Loire pour avoir cru en moi et m'a accueilli au sein de son équipe. Grâce à ses compétences scientifiques et pédagogiques, M. Pastor a su m'orienter dans mes recherches par son écoute et ses conseils avisés. Je lui suis aussi reconnaissante pour le temps et la patience qu'il m'a accordés. Je lui adresse ma gratitude pour ses qualités humaines et ses encouragements, surtout durant les deux dernières années, qui m'ont permis d'aller jusqu'au bout de ma thèse.

Mes plus sincères remerciements s'adressent également à mon codirecteur de thèse M. Erwan L'Her, Professeur de réanimation médicale et de médecine d'urgences au CHRU de Brest, pour ses précieux conseils et accompagnement durant mes travaux de thèse.

Je remercie également tous les membres du département Signal et Communications de l'IMT Atlantique pour m'avoir permis de travailler dans d'aussi bonnes conditions.

Je remercie M. Lotfi Senhadji, Professeur à l'Université de Rennes 1 et directeur du LTSI (Laboratoire de Traitement du Signal et de l'Image), qui m'a fait l'honneur de présider le jury de ma thèse. Je remercie également les rapporteurs, Mme. Valérie Louis-Dorr, Professeur à l'Université de Lorraine et M. Laurant Navarro, Maître de Conférence à l'ENSM à Saint Etienne, pour leurs appréciations et remarques pertinentes sur les travaux de cette thèse. Je remercie également M. Julien de Jonckheere, Chargée de recherche au CHRU de Lille, d'avoir bien voulu examiner ce travail.

Je remercie toutes les personnes formidables que j'ai eu la chance de les rencontrer au cours de ces années de thèse et qui ont été pour moi une deuxième famille à Brest.

Enfin, j'adresse toute ma gratitude à mes parents et mon frère Wael pour tout leur accompagnement et sacrifices tout au long de mon parcours et sans lesquels je n'en serais pas là aujourd'hui. Je remercie également mon mari pour son soutien et sa patience durant mes années de thèse. Merci à toutes les personnes des familles Chérif, BelHadj et Zhioua qui ont été indispensables à la réussite de ce travail par leur bienveillance et encouragement.

Dédicace

A la prunelle de mes yeux, mes enfants Mohamed Aziz et Zeyneb,

A mes parents et mon frère, témoins de mes joies, de mes hauts et bas.

A mon mari, ma famille et mes amis.

A tous mes enseignants qui m'ont inculquer l'amour du savoir.

Abstract

One promising area of research in clinical routine involves using photoplethysmography (PPG) for monitoring respiratory activities. PPG is an optical signal acquired from oximeters, whose principal use consists in measuring oxygen saturation. Despite its simplicity of use, the deployment of this technique is still limited because of the signal sensitivity to distortions and the non-reproducibility between subjects, but also for the same subject, due to age and health conditions. The main aim of this work is to develop robust and universal methods for estimating accurate respiratory rate regardless of the intra- and inter-individual variability that affects PPG features. For this purpose, firstly, an adaptive artefact detection method based on template matching and decision by Random Distortion Testing is introduced for detecting PPG pulses with artefacts. Secondly, an analysis of several spectral methods for Respiratory Rate (RR) estimation on two different databases, with different age ranges and different respiratory modes, is proposed. Thirdly, a Spectral Respiratory Quality Index (SRQI) is attributed to respiratory rate estimates, in order that the clinician may select only RR values with a large confidence scale. Promising results are found for two different databases. The RMSE found when comparing the estimated RR values to RR references is about $0.66 breath/min$ for the Capnobase data with an acceptance rate of 67%. For Reastoc data, the RMSE is about $0.31 breath/min$ with an acceptance rate of 75.6%.

Keywords: photoplethysmography, respiratory rate, artifacts detection, quality index.

Résumé

1. Introduction

La surveillance de l'évolution des signaux physiologiques, en routine clinique et pour les applications de télémedicine, requiert des procédés automatisés de suivi et de détection d'anomalie. Il y a un besoin évident de méthodes non invasives, fiables et simples pour le suivi en temps réel des activités cardio-respiratoires des patients.

La photopléthysmographie (PPG) est une technologie prometteuse qui peut assurer, en théorie, d'estimer les fréquences cardiaques et respiratoires. Il s'agit d'un signal optique acquis à partir de l'oxymètre de pouls, dont l'usage principal consiste à mesurer la saturation en oxygène. Le signal PPG est largement utilisé par les cliniciens vu sa simplicité d'acquisition mais à ce jour il n'existe pas de méthodes automatisées pour le traitement de ce signal.

Dans la littérature, plusieurs méthodes ont été proposées pour l'étude du signal PPG dans des conditions parfaites loin d'être similaires aux conditions réelles en milieux hospitalier. Nous avons constaté beaucoup de différences de performances d'une base de données à une autre et entre les différents patients de la même base. En dépit de sa simplicité, le signal du PPG est sensible aux distorsions et à la non-reproductibilité entre les sujets, mais aussi pour les mêmes sujets, en raison de l'âge et des conditions de santé.

L'originalité des travaux de cette thèse réside essentiellement dans la recherche de méthodes plutôt universelles qui s'appliquent à tous les patients quelques soient les conditions d'acquisition et par la suite réussir à automatiser le traitement de PPG dans les conditions réelles. Dans ce contexte, les travaux de cette thèse ont pour objectifs le développement de méthodes robustes et universelles afin d'avoir une estimation précise de la fréquence respiratoire indépendamment de la variabilité intra et interindividuelle du signal PPG.

Dans une première partie nous présentons les principales caractéristiques et application du signal PPG. Dans une deuxième partie nous exposons la problématique de la détection des artéfacts à partir du signal de PPG en proposant une méthode orig-

inale de comparaison de pulses. Dans une troisième partie, une analyse de plusieurs méthodes spectrales d'estimation de la fréquence respiratoire (FR) est proposée. Finalement, dans une quatrième partie, un indice de qualité respiratoire spectral (SRQI) est conçu dans le but de communiquer au clinicien que les valeurs d'estimation de la fréquence respiratoire ayant un certains indice de confiance.

2. Les caractéristiques du signal PPG

2.1. La forme d'onde du PPG

Le PPG est un signal complexe constitué de différents composants. La composante principale du PPG est la forme d'onde pulsatile qui est synchronisée à chaque battement cardiaque (AC). Cette composante est superposée et modulée par un composant (DC) qui varie lentement et qui est relié à la respiration et aux activités vasomotrices. Par conséquence, la fréquence cardiaque (FC) correspond à l'inverse des distances entre les pulses appelé PtoP et la FR corresponds la fréquence des modulations du PPG.

La modélisation mathématique de la forme d'onde du PPG permet aux chercheurs d'évaluer les performances de leurs algorithmes. Certains chercheurs ont proposés quelques modèles gaussien et logarithmiques. Mais ces modèles ne reflètent pas nécessairement les signaux de PPG acquis dans des conditions réelles. En effet, la représentation du caractère impulsionnel du signal ainsi que les différentes modulations respiratoires n'est pas si évident. La non reproductibilité du signal et la variabilité selon les conditions du sujet limite la fiabilité d'une telle modélisation.

2.2. Application clinique du signal PPG

La principale application directe du PPG est la mesure de la saturation en oxygène dans le sang par oxymètre de pouls. D'autres applications subordonnées ont été étudiées durant ces dernières années dont on cite le suivi des activités cardiaques et respiratoires. En effet, la composante pulsatile de PPG est synchronisée avec le cycle cardiaque. Par conséquence, la fréquence cardiaque peut être estimée en calculant la distance entre les maxima consécutifs des pulses. Cette application de PPG est intéressante dans le cadre de suivi clinique où les électrodes ECG peuvent être inconfortables à porter, comme pour les sujets sous traitement d'hémodialyse. L'irrégularité des battements cardiaques peut alors être détectée en utilisant des informations sur la morphologie du pouls et de la fréquence cardiaque. Ainsi, différents types d'arythmies peuvent être détectées à

partir du PPG. Quant à la fréquence respiratoire, les méthodes existantes de mesure et de suivi des FR sont encore limitées par leur difficulté d'utilisation et leur inexactitude. En fait, le fait de placer une bande étroite autour du thorax du patient constitue une mesure non reproductible et de plus inconfortable pour le patient. D'autres techniques, comme la capnographie ou les mesures acoustiques, sont limitées aux unités de soins intensifs et aux salles d'opération. Une approche alternative est donc proposée: c'est d'extraire la FR du signal PPG et d'étudier son évolution au cours du temps.

2.3. Description des données

Capnobase

Capnobase est une base de données en ligne qui contient des signaux physiologiques recueillis de patients, adultes et enfants, sous anesthésie à l'hôpital d'enfant de Britsh Colombia et à l'hôpital St Paul à Vancouver. Cette base de données contient 42 enregistrements de longueur de 8 minutes. Pour chaque patient, il y a trois signaux enregistrés simultanément: l'ECG, la capnométrie et le PPG. Nous allons utiliser l'ECG et la capnométrie comme vérité terrain avec lesquelles on compare la fréquence cardiaque et respiratoire estimées à partir du PPG.

Réastoc

Il s'agit de données recueillies au service de réanimation à l'hôpital de la Cavale Blanche à Brest. Les enregistrements (ECG, PPG, ABP) ont été réalisés grâce au logiciel Synapse développé par l'unité INSERM Rennes au LTSI de Rennes. Pour notre étude, nous avons sélectionné des séquences de 22 patients de longueur variable selon le patient. La fréquence cardiaque et respiratoire de référence ont été notées par le personnel soignant. Cette base de données est en cours d'évolution et d'autres informations y seront intégrées.

3. La détection des artéfacts du signal PPG par la RDT

Une estimation précise de FC et FR repose sur la qualité des signaux PPG. En effet, le signal de PPG est sensible aux artéfacts associés aux conditions de mesure et aux mouvements du patient. En observant la morphologie de la forme d'onde PPG, nous

remarquons qu'elle est composée d'impulsions de formes similaires et de légères différences d'amplitude et de forme. En cas d'artefacts, la forme des impulsions change soudainement, ce qui entraîne une distorsion du signal. Autrement, lorsque l'activité physiologique varie, par exemple en FC ou FR, les changements de forme du pouls sont lents et faibles.

3.1. Aperçu de la littérature

Les artefacts ont une distribution inconnue et peuvent avoir plusieurs formes. Comme la plupart des signaux physiologiques, le PPG n'est pas stationnaire et ses paramètres changent considérablement au fil du temps. Ainsi, la comparaison avec un modèle ou des signaux de référence n'est pas évidente.

La première approche proposée est basée sur le filtrage. Les méthodes de filtrage ordinaires ne permettent pas d'éliminer les artefacts, car il existe un chevauchement entre la bande de fréquences des artefacts et celle des signaux PPG utiles. D'autres méthodes de filtrage sont proposées comme le filtrage adaptatif, les ondelettes ou la Décomposition en Modes Empiriques. En dépit de la cohérence méthodologique de ces méthodes de filtrage, ils ont des performances limitées pour les applications en temps réel. En fait, il n'y a pas de modèle statistique pour le PPG, les artefacts et leurs variations dans le temps. Pire encore, des informations utiles sur les activités respiratoires ou cardiaques peuvent être omises lors de l'application du filtrage sur des portions propres de PPG.

La seconde approche proposée est basée sur la classification. Il s'agit de décomposer le signal en un ensemble de pulses et de les classer en pulses propres ou pulses altérés par des artéfacts. Mais cette approche a également ses limites. En effet, les méthodes de classification nécessitent de larges bases de données représentatives ; ce qui semble compliqué à mettre en œuvre. Jusqu'à présent, il n'y a pas de normalisation du protocole de l'acquisition du signal du PPG. Il existe donc une grande variabilité entre les enregistrements. De plus, la nature dynamique du comportement physiologique engendre un signal PPG non-stationnaire avec une distribution inconnue. Il existe donc un besoin d'un système autonome avec peu de paramètres prédéfinis, qui pourrait être adapté à chaque enregistrement et pourrait mettre à jour ses paramètres lorsque des modifications substantielles se produisent.

3.2. Méthodes proposées et résultats

A partir de la littérature existante, les méthodes de comparaison de modèles sont les mieux adaptées pour détecter les signaux PPG avec des artefacts sans endommager les informations utiles. Mais le problème est toujours de savoir comment choisir le seuil et la métrique de comparaison les plus appropriés. Comme dans Sukor & al., l'idée générale consiste à calculer un modèle de pulse et à utiliser une métrique pour comparer ce modèle avec les autres pulses. Pour se faire, nous introduisons une approche universelle afin d'avoir un système autonome qui pourrait être adapté à tout enregistrement, sans avoir de paramètres à ajuster. Nous avons proposé deux approches: la première avec une détection simple et la deuxième avec une détection adaptative.

Détection simple

Tout d'abord un prétraitement est effectué afin de détecter les maximas et d'extraire les pulses. La deuxième étape consiste à sélectionner un template. Le signal est découpé en plusieurs segments. Sur chaque segment, la corrélation entre le pulse moyen et les autres pulses du même segment est calculée. Le segment le plus stable est celui où la moyenne des corrélations est maximale. Ainsi sur ce segment, le template, la moyenne des corrélations μ et la variance des corrélations σ sont initialisées.

En troisième étape, le calcul de tous les coefficients de corrélation entre le template générée et les autres pulse est réalisé. Enfin, un test d'hypothèse est proposé pour prendre une décision sur la présence ou pas des artéfacts. Le test utilisé est le RDT (Random Distortion Testing) qui est une technique qui nous permet de tester si la distance entre la corrélation correspondante à chaque pulse et le moyenne des corrélation μ dépasse une certaine tolérance τ. Ainsi, si cette distance est supérieur au seuil RDT, on décide que le pulse est pollué par des artéfacts

Pour évaluer les performances de cette méthode, la base de données de Sukor *& .al* est utilisé. Cette base contient 104 enregistrement de PPG contenants des annotations manuelles des artéfacts. Chaque enregistrement est de longueur d'une minute. Le 1/3 des pulses sont des artéfacts. Des bonnes performances ont été retrouvé surtout par rapport à la spécificité et la précision qui sont respectivement $85\% \pm 12$ et $83\% \pm 8$. L'avantage de notre méthode basé sur le test RDT est qu'elle utilise un seul seuil spécifique à chaque enregistrement alors que Sukor utilise 6 seuils empiriques. De plus, vu que nous utilisons peu de paramètre, il n'y a pas de besoin de faire un nouveau apprentissage à chaque changement de base de donnée.

Détection adaptative des artéfacts

Les variations de la FR et FC peuvent causer des légères fluctuations de la forme des pulses PPG surtout pour les longs enregistrements. Il faut donc différentier ces effets physiologiques des vrais artéfacts. D'où la nécessité d'avoir une détection adaptative pour les longs enregistrements.

Comme pour la détection simple, une étape primordiale est d'initialiser les paramètres de la détection. La première minute de l'enregistrement est choisie pour identifier un segment stable sur lequel le template, μ et σ sont initialisés. Après initialisation, nous continuons notre comparaison avec les autres pulses. Si il n'y a pas de détection d'artéfact, le template, μ et σ sont actualisés. Sinon, en cas des artéfacts, on passe au pulse suivant sans actualisation. Evidement ici notre seuil de décision qui dépend de μ et σ devient adaptative vu l'actualisation des paramètres.

Cette méthode adaptative a été testée sur une portion des deux bases de données Réactoc et Capnobase. Une nette amélioration de la précision est observée par rapport à la détection simple. Pour Capnobase, la précision est passée de 74% en utilisant une détection simple à 91% en utilisant la détection adaptative. De même pour Réastoc, la précision est passée de 80% en utilisant une détection simple à 92% en utilisant la détection adaptative.

3.3. Discussion

La méthode de détection adaptative des artéfacts basée sur l'utilisation d'un seuil RDT adaptative a permis une nette diminution du taux de fausses alarmes et par la suite une amélioration de la précision. En effet, par cette technique un ajustement du template au fur et à mesure de l'enregistrement est réalisé sans être biaisé par les artéfacts.

Les performances trouvées sont semblables pour les deux bases de données sans pour autant avoir recours à une nouvelle étape d'apprentissage à chaque changement de base de données.

L'approche proposée peut être appliquée à d'autres signaux biomédicaux qui ont les mêmes caractéristiques que le PPG. En d'autres termes, ces signaux sont quasi-périodiques ayant une forme similaire qui se répète au cours du temps e que se déforme brusquement suite à des artéfacts.

4. L'estimation de la fréquence respiratoire à partir du PPG

La fréquence respiratoire (FR) est un indicateur physiologique largement utilisé en soins cliniques. C'est un marqueur vital de la détérioration clinique lorsque des changements suspects sont détectés. Le signal de PPG constitue une bonne alternative aux méthodes traditionnelles de mesure de la FR, grâce à sa simplicité d'utilisation et son caractère non invasif.

4.1. Aperçu de la littérature

Les effets physiologiques liés à l'activité respiratoire et cardiaque induisent plusieurs modulations des signaux PPG. De ce fait, la plupart des algorithmes, dans la littérature, proposent une étape préliminaire pour extraire les modulations du signal PPG. La deuxième étape consiste à estimer les valeurs de FR à partir de chaque modulation. Enfin, la troisième étape est d'analyser les FRs extraites des modulations et puis calculer la FR finale. Nous nous intéressons plus particulièrement à la méthode de Karlen qui est une méthode de référence dans la littérature.

Le processus usuel pour estimer la FR, dans la littérature, est comme suie: La première étape est l'estimation de chaque modulation après extraction des minimas et des maximas:

- Modulation des amplitudes (AM): C'est la variation des amplitudes des pulses. Il s'agit de calculer la hauteur entre le maximum et le minimum de chaque pulse.
- Modulation de la bande de base (BW): C'est la variation d'intensité induite par les voies respiratoires. Il s'agit du milieu de la distance entre le maximum et le minimum de chaque pulse
- Modulation des fréquences (FM): C'est la variation des fréquences cardiaques. Il s'agit de la distance entre les maxima successifs des pulses.

La deuxième étape consiste à estimer FR de chaque modulation. La FR correspond à la fréquence du pic maximal de chaque spectre de modulation sur la bande fréquentielle correspondante aux fréquences respiratoires usuelles [0.013,1 Hz].

La troisième étape est d'estimer la FR finale. Dans la méthode Fusion de Karlen, il s'agit de la moyenne des trois FR estimés des modulations AM, BW et FM. Karlen

a ensuite proposé un processus de décision, appelé Smart Fusion, afin de limiter les erreurs dues aux valeurs aberrantes. En effet, si l'écart type des trois FR est supérieur à 4 respirations / minute, la valeur de la FR finale est rejetée. Ce processus de décision a permis d'améliorer nettement le taux d'erreur mais le taux de rejet reste très important. Ce qui peut limiter l'application de cette méthode dans les conditions réelles de surveillance de FR.

4.2. Méthodes proposées

Dans ce travail de thèse, nous proposons de tester un ensemble de méthodes spectrales afin de trouver une méthode générique d'estimation de la FR quel que soit l'état du patient. La première approche consiste à appliquer le consensus spectrum sur un signal PPG brut sans passer par l'extraction des modulations. Il s'agit de multiplier des spectres successifs de FFT. Cette méthode, que nous appelons CS-PPG, permet d'accentuer les pics correspondants à la fréquence respiratoire et de diminuer ceux résultants des artéfacts. Le pic maximal sur la bande de fréquence respiratoire correspond à la fréquence respiratoire du sujet. La deuxième approche consiste à utiliser les signaux de modulation respiratoire du PPG qui sont AM, BW et FM. Quatre méthodes sont proposées:

- *FFT-Med: Il s'agit d'appliquer la FFT sur une fenêtre de chaque modulation. La FR_{Mod} est le pic maximum de chaque spectre. La FR_{finale} est donnée par la médiane des trois FR_{Mod}.*

- *CS-Med: Il s'agit d'appliquer le consensus spectrum sur n fenêtres consécutives de chaque modulation. La FR_{Mod} est le pic maximum de chaque spectre résultant du consensus spectrum. La FR_{finale} est donnée par la médiane des trois FR_{Mod}.*

- *SF-Med: Il s'agit d'appliquer la FFT sur une fenêtre de chaque modulation puis fusionner les spectres des trois modulation par la médiane. La FR_{finale} est donnée par le pic maximum du spectre résultant.*

- *SF-CS: Il s'agit d'appliquer la FFT sur une fenêtre de chaque modulation puis fusionner les spectres des trois modulation par le consensus spectrum. La FR_{finale} est donnée par le pic maximum du spectre résultant.*

4.3. Résultats et discussion

Les performances des algorithmes proposés sont testées sur deux bases de données Capnobase et Réastoc pour différentes tranches d'âge et pour deux modes de respiration: spontanée et sous ventilation mécanique. Les meilleures performances pour Capnobase sont données par la méthode CS-Med avec un RMSE de $1.75(0.59 - 3.63)$ respiration/min. Alors que pour Réastoc, les meilleures performances sont données par la méthode CS-PPG avec un RMSE de $0.35(0.18 - 2.11)$ respiration/min

Cette divergence des performances selon les bases de données peut être expliquée par le fait que la qualité des signaux de modulations dépend de différents facteurs comme le type de capteur, la fréquence d'échantillonnage, l'environnement d'acquisition...

Pour une même base de donnée, les performances divergent selon l'âge et le mode de respiration. En effet, l'estimation de FR est meilleure chez les patients sous ventilation mécanique. Ceci peut être expliqué par le fait qu'ils ont une respiration plus stable. Les performances des algorithmes dépendent aussi du l'âge du patient. Ceci peut être expliqué par le fait que l'effort respiratoire dépend du l'âge du patient.

Les performances des algorithmes ont été comparées à certaines méthodes de la littérature, principalement la méthode Smart Fusion proposée par Karlen *& al.*

Pour Capnobase, l'algorithme que nous proposons CS-Med se comporte bien par rapport à l'algorithme Fusion. Cependant, l'algorithme Smart Fusion de Karlen présente les meilleures performances aux dépens d'une diminution du taux d'estimation. En fait, pour Capnobase, seules 55 % des fenêtres sont conservées. Néanmoins, contrairement aux méthodes de Karlen, avec notre méthode CS-Med, aucun prétraitement pour l'élimination des artefacts n'a été effectué et 100% du signal a été analysé.

5. Optimisation de l'estimation de FR

L'estimation du FR à partir du signal PPG constitue une alternative prometteuse. Mais jusqu'à présent, son utilisation est encore limitée dans les conditions réelles en raison de la non-robustesse des méthodes utilisées. En fait, comme indiqué dans la section précédente, les performances des algorithmes varient considérablement en fonction des bases de données et des caractéristiques des sujets. Dans ce chapitre, nous présentons une analyse des facteurs qui affectent les signaux respiratoires et nous proposons quelques outils pour afficher au clinicien uniquement les valeurs de fréquences respiratoires pertinentes.

5.1. Optimisation par élimination des artéfacts

Des tests ont été conduits afin de vérifier la pertinence de l'élimination des pulses pollués par des artéfacts avant de réaliser une estimation de FR. Il s'agit de voir l'évolution du RMSE en fonction des pulses éliminés. Nous avons constaté une légère amélioration des résultats de RMSE après l'élimination des artéfacts. Mais cette amélioration n'est pas suffisante pour garantir une estimation robuste de la FR.

L'analyse menée dans ce chapitre confirme que les pulses avec des artefacts ne sont pas la seule cause d'une mauvaise estimation de la FR. En effet, les méthodes de détection d'artefacts, notamment celles basées sur la comparaison de la morphologie des impulsions, ne concernent que les caractéristiques temporelles et pulsatiles du signal PPG sans se préoccuper des modulations du signal et de la qualité de la composante spectrale.

Une analyse plus approfondie du spectre de certains segments PPG pour lesquels CS-Med et CS-PPG n'ont pas permis d'obtenir un FR précis révèle que les pics correspondant à FR ne sont pas toujours dominants, même en l'absence d'artefacts. Dans certains cas, des pics d'harmoniques apparaissent dans les spectres avec une amplitude supérieure à celle du pic de la FR réel. Dans d'autres cas, des pics parasites non identifiables sont observés. Ces pics ont des amplitudes dans la même plage que le pic FR et leurs fréquences correspondantes sont inférieures à la fréquence de la FR de référence.

Une nouvelle métrique est alors nécessaire pour évaluer les propriétés spectrales des signaux respiratoires. La conception de cette métrique est le sujet de la section suivante.

5.2. Optimisation par indice de qualité spectrale de l'activité respiratoire

L'analyse spectrale du signal respiratoire est un outil prometteur pour évaluer la pertinence de l'estimation de la FR. En effet, un indice est nécessaire pour évaluer si le pic maximum dans la bande de fréquence respiratoire $(0,15 - 1Hz)$ est suffisamment dominant pour être considéré comme une fréquence respiratoire pertinente, par rapport aux autres pics du même intervalle. À cet égard, l'indice de qualité respiratoire spectrale (SRQI) est défini comme le rapport entre la puissance autour du pic dominant et la puissance totale du signal dans la bande de fréquences respiratoires. Tel qu'il est définit , le SRQI a tendance à décroitre en fonction de l'énergie des pics fluctuants.

Par exemple si on a deux pics dominants le SRQI tend vers 0.5. Si on a un seul pic, le SRQI tend vers 1 .

Nous avons étudié les variations de RMSE pour différentes valeurs de seuil Q de SRQI entre 0 et 1. Les meilleures performances ont été trouvées pour la méthode CS-Med combinée au SRQI quel que soit le type de donnée. Pour Capnobase, le RMSE trouvé lors de la comparaison des valeurs de FR estimées aux valeurs de références de la FR est $0,66$ respiration / min avec un taux d'acceptation de 67%. Pour Reastoc, le RMSE est de $0,31$ respiration / min avec un taux d'acceptation de $75,6\%$.

Le critère SRQI associé à la méthode CS-Med présente de nombreux avantages par rapport aux méthodes de la littérature. Le principal avantage est la simplicité de la méthode par son nombre réduit de paramètres. Ceci est un élément clé pour la mise en œuvre future dans un contexte en temps réel. Le deuxième avantage est l'universalité de la méthode. En effet, des performances similaires ont été trouvées pour les bases de données Capnobase et Reastoc, pour toutes les tranches d'âge et pour différents modes de respiration. Le troisième avantage est la robustesse de cette méthode par rapport aux artefacts. En effet, le SRQI peut détecter et rejeter les cas où des segments d'artefacts affectent l'estimation de la FR. Par conséquent, aucun module de traitement supplémentaire dédié à la détections d'artefacts n'est nécessaire. Néanmoins, l'amélioration du RMSE est au prix d'une réduction du taux d'acceptation mais ce taux d'acceptation reste raisonnable par rapport à la méthode Smart Fusion de Karlen.

6. Conclusion

Dans cette thèse, nous avons abordé la problématique de développement de méthodes robustes et universelles afin d'avoir une estimation précise de la fréquence respiratoire indépendamment de la variabilité intra et interindividuelle du signal PPG.

Nous avons développé une méthode de détection des artéfacts basée sur la génération d'un modèle adaptatif de pulse et la comparaison par le test RDT. Cette méthode est performante pour différentes base de données et a plusieurs avantages. En effet, la méthode présentée nécessite peu de paramètres. De plus, contrairement aux méthodes de classification, il n'y a pas de besoin d'apprentissage à chaque changement du type de données.

Dans une deuxième phase, un ensemble d'outils spectraux ont été développés afin d'estimer la FR. La méthode CS-MED associée au critère SRQI donne des estimations

précises de la FR sans avoir recours à la détection d'artefact. Grâce à sa simplicité, à sa fiabilité et à ses minimes besoins en ressources, l'approche proposée peut être intégrée dans des traitements cliniques en temps réel pour différentes applications de surveillance, que ce soit dans les hôpitaux ou les soins à domicile. Elle représente également une alternative aux méthodes de classification, qui nécessitent de larges données d'apprentissage.

En résumé, la surveillance des signes vitaux à partir du signal PPG constitue une bonne alternative aux systèmes de mesures traditionnels. Néanmoins, quelques limites liées à la compréhension des phénomènes physiologiques du signal restent à étudier.

Comme perspective, des tests cliniques à large échelle avec plus de données hétérogènes (cas des arythmies, des apnées...) sont à réaliser. D'autre part, une amélioration du taux d'acceptation par une analyse plus approfondie des phénomènes physiologiques affectant la respiration est envisagée. De manière plus globale, il est important de réaliser une étude d'un système qui prend en considération les données propres à chaque patient en plus du signal PPG.

Contents

Remerciements **i**

Dédicace **iii**

Abstract **v**

Résumé **vii**

Abbreviations **xxix**

1 Introduction **1**

1.1 Context 1

1.2 Motivation 2

1.3 Objectives 3

1.4 Outline of the thesis 3

2 The photoplethysmography signal **5**

2.1 Introduction 6

2.2 PPG measuring characteristics 6

2.2.1 Technical facts about PPG 6

2.2.2 Sites and devices for measuring PPG 7

2.2.3 Measurement protocol and reproducibility 8

2.3 PPG waveform characteristics 8

2.3.1 Pulse characteristics 8

2.3.2 PPG waveform modulations 10

2.3.3 Factors affecting PPG waveform 11

2.3.4 PPG mathematical model 12

2.4 PPG for clinical physiological monitoring 14

2.4.1 Monitoring blood oxygen saturation 14
2.4.2 Monitoring heart activity 15
2.4.3 Monitoring respiratory activity 16
2.4.4 Monitoring hypovolemia 17
2.5 PPG database 17
2.5.1 Sukor Data 18
2.5.2 CapnoBase Data 18
2.5.3 ReaStoc data 19
2.6 Conclusion 20

3 Detection of artifacts in PPG signal 21
3.1 Introduction 21
3.2 Artifact causes and impact 22
3.3 State of the art 22
3.3.1 Filtering method with PPG restoration 23
3.3.2 Morphology analysis and artifact detection method 27
3.4 RDT for artifact detection 29
3.4.1 Simple artifact detection for short records 30
3.4.2 Adaptive RDT for artifact detection 35
3.5 Results and discussion 37
3.5.1 Simple RDT detection performance 37
3.5.2 Adaptive RDT performance 40
3.5.3 Possible improvement and extension to other physiological signals 42
3.6 Conclusion 42

4 Respiratory rate estimation from PPG 43
4.1 Introduction 44
4.2 State of art 44
4.2.1 RR estimation from raw PPG 44
4.2.2 RR estimation from derived PPG signals 45

4.2.3 General limits of the existing methods 49
4.3 Consensus spectrum for RR estimation 49
4.4 RR from PPG modulations 52
4.4.1 Extracting PPG modulations 52
4.4.2 Extraction of respiratory modulations 53
4.4.3 RR estimation 55
4.5 Results 60
4.5.1 Results on Capnobase 60
4.5.2 Results on Reastoc 72
4.6 Discussion 75
4.6.1 Comparison between the proposed algorithms 75
4.6.2 Age impact on algorithms performance 76
4.6.3 Comparison with others methods 77
4.7 Conclusion 79

5 Optimization of respiratory rate monitoring from PPG 81
5.1 Introduction 81
5.2 Artifact detection impact on respiratory rate estimation 82
5.2.1 Motivation 82
5.2.2 Analysis methodology 83
5.2.3 Results on Capnobase 83
5.2.4 Results on Reastoc 85
5.2.5 Limits of artifact detection 86
5.3 SRQI impact on respiratory rate estimation 88
5.3.1 SRQI definition 88
5.3.2 Analysis methodology 89
5.3.3 Results on Capnobase 90
5.3.4 Results on Reastoc 93
5.4 Discussion 96

5.5 Conclusion 99

6 Conclusion and perspectives **101**

6.1 Summary and main contributions 101

6.2 Perspectives and future works 104

Bibliographie **105**

List of Figures

2.1 An example of a raw PPG from the Capnobase database: pulsatile component AC and stationary component DC are easily discerned 6

2.2 Comparison between different shapes of pulses with notch 9

2.3 Example of a PPG pulse and its parameters with area of anacrotic/systolic and catacrotic/diastolic phase. x and y are the amplitude of, respectively, systolic and diastolic peak. 9

2.4 Features of PPG pulse 10

2.5 PPG modulated with respiratory signals AM and BW 11

2.6 An example of a clean PPG signal segment from Sukor Database 18

2.7 An example of PPG signal segment from Capnobase 19

2.8 An example of PPG signal segment from ReaStoc 20

3.1 An example of PPG signal with artifacts from Sukor Database [2] . . . 23

3.2 Flow chart for removing artifacts with adaptive filtering method 23

3.3 Flow chart for removing artifacts with wavelet method 25

3.4 Flow chart for removing artifacts with EMD method 26

3.5 Process flow for extracting IMF from a given signal S(t) (figure from [3]) 27

3.6 Flow chart for removing artifacts with RDT method 29

3.7 Detecting blocks of interest by applying moving average on x then estimating mimima on the resulting signal x_{MA}. Pulse maximum is the maximum value on x in the selected block of interest. 31

3.8 Superposing pulses and calculating pulse template by averaging. 32

3.9 Results of RDT artifact detection on PPG signal 34

3.10 ROC curve of the RDT algorithm for different values of $L = 4 : 30$s, $\gamma = 10^{-1}$ and $\tau = 4.5$ 38

3.11 In this example, RDT detects no artifacts however the label indicates that there are bad pulses (red). If we observe the shape of these questionable pulses, we notice no actual artifact. 39

4.1 Typical steps of most algorithms for estimating RR 47

4.2 Comparison between simple PSD vs. Consensus of four spectra: Peaks are well differentiated in (b). 51

4.3 Each signal AM, FM and BW (a,d,g) is interpolated and oversampled at frequency $4Hz$ to get signals over time (b,e,h). Finally, every signal is filtered to eliminate noise resulted from sampling (c,f,i) 54

4.4 Example of a $32s$ window used for RR estimation from the induced respiratory signals resampled and filtered. The power spectrum is calculated for each respiratory modulation signal (right column) and the maximum power (red line) is selected. In this example, the three respiratory frequencies coincide at the same frequency $0.3125Hz$, which corresponds to $18.75breath/min$ 56

4.5 Example of four spectra from four $32s$ sliding windows with an overlap of $16s$ (left column). The resulted consensus spectrum for each signal modulation is presented in the right column and the maximum peak (red line) is selected. In this example, two respiratory frequencies coincide at the same frequency $0.3125Hz$, which corresponds to $18.75breath/min$. However, in the FM signal, the maximum coincides with an harmonic at $0.6Hz$. 58

4.6 Example of a fused spectrum from the windowed FFTs of the AM, FM and BW signals. The respiratory frequency is estimated at 0.3125 Hz, which corresponds to 18.75 breath/min. The harmonic observed in Figure 4.5d is dimmed in this case 59

4.7 Example of RR estimated with CS-Med method using different window sizes compared to the RR values from reference. 62

4.8 Results from RMSE analysis of capnobase data to select the best FFT window size $\{32s, 64s, 128s, 256s\}$. 63

4.9 Results from RMSE analysis of capnobase data to select the best number of FFTs needed for Consensus Spectrum calculation in CS-PPG and CS-Med 64

4.10 Results from RMSE analysis of capnobase data using 64s window. Comparison of all methods according to age range 65

4.11 Results from RMSE analysis of capnobase data using 64s window. Comparison of all methods according to ventilation mode 65

4.12 Bland-Altman plot of the comparison between RR reference and RR estimate from CS-Med. Mean Difference line is in black at -0.24, Confidence interval limits are between -3.77 and 3.28 68

4.13 Scatter plot comparing RR references with RR estimated from PPG by using CS-PPG, CS-Med, FFT-Med, SF-Med and SF-CS methods. The best fit line is represented by a blue broken line. 68

4.14 Results from RMSE analysis when comparing instantaneous RR references to estimates from Karlen methods: Fusion and Smart Fusion and our proposed method CS-Med 70

4.15 Performance comparison of the CS-Med method, Fusion method and Smart Fusion method with respect to instantaneous and global measures. 71

4.16 Results of the RMSE analysis when comparing RR references to RR estimates for the methods: CS-PPG, CS-Med, FFT-Med, SF-Med, SF-CS and Smart Fusion 73

4.17 Scatter plot comparing RR references with RR estimates. The best fit line is represented by a blue broken line 74

4.18 RMSE analysis according to age range 75

4.19 Summary of the RR estimation methods analyzed in this chapter . . . 78

5.1 Flowchart of the process for RR estimation coupled to artifact detection 83

5.2 RMSE of RR estimates by CS-PPG applied to the Capnobase data, versus the percentage of windows accepted as nPAW increases. 84

5.3 RMSE of RR estimates by CS-Med applied to the Capnobase data, versus the percentage of windows accepted as nPAW increases 85

5.4 RMSE of RR estimates by CS-PPG applied to the Reastoc data, versus the percentage of windows accepted as nPAW increases 85

5.5 RMSE of RR estimates by CS-Med applied to the Reastoc data ,versus the percentage of windows accepted as nPAW increases 86

5.6 Example of a PPG segment where there is no artifact but the CS-Med method fails to estimate good RR. 87

5.7 Spectrum of the PPG signal presented in the latter figure 5.6. Spurious peaks in this spectrum have higher amplitude than the peak corresponding to the real RR (0.156 Hz). This causes erroneous estimation of RR. 87

5.8 Examples of two different cases illustrating the variation of the SRQI according to the number of peaks present in the spectrum. In red, the bins corresponding to the dominant peak. The RR reference value in this example is the 9th bin, which corresponds to 0.14 Hz 89

5.9 RMSE of RR estimates from the CS-PPG method applied to the Capnobase data, versus the percentage of windows accepted as $\mathcal{Q}$ increases. 90

5.10 RMSE of RR estimates from the CS-Med method applied to the Capnobase data, versus the percentage of windows accepted as $\mathcal{Q}$ increases. 91

5.11 RMSE of RR estimates from the Fusion method applied to the Capnobase data, versus the percentage of windows accepted as $\mathcal{Q}$ increases. 92

5.12 RMSE of RR estimates from the Smart Fusion method applied to the Capnobase data, versus the percentage of windows accepted as $\mathcal{Q}$ increases. 92

5.13 RMSE of RR estimates from the CS-PPG method applied to the Reastoc data, versus the percentage of windows accepted as $\mathcal{Q}$ increases. 93

5.14 RMSE of RR estimates from the CS-Med method applied to the Reastoc data, versus the percentage of windows accepted as $\mathcal{Q}$ increases. 94

5.15 RMSE of RR estimates from the Fusion method applied to the Reastoc data versus the percentage of windows accepted as $\mathcal{Q}$ increases. 94

5.16 RMSE of RR estimates from the Smart Fusion method applied to the Reastoc data, versus the percentage of windows accepted as $\mathcal{Q}$ increases. 95

5.17 Main steps for a respiratory rate estimation system based on the CS-Med method and SRQI criterion. 100

List of Tables

2.1 Conventional pulse oximetry algorithm: R / IR: Absorption ratio of the red / infrared light waves. SpO2: Plethysmographic saturation in O2. [1] 14

3.1 Performances of RDT method compared to Sukor & *al.* $L = 12$s, $\gamma = 10^{-1}$ and $\tau = 4.5$ 39

3.2 Capnobase: Performances of adaptive RDT method compared to simple RDT $L = 10$s, $\gamma = 10^{-3}$ and $\tau = 2$ 41

3.3 Reastoc: Performance of adaptive RDT method compared to simple RDT $L = 10$s, $\gamma = 10^{-3}$ and $\tau = 1.5$ 41

4.1 Summary of main methods presented in the literature 48

4.2 Performance of the RR estimation algorithms, according to the age ranges, for windows of $64s$. In each age group, N is the number of records. For each method, we give the median and inter-quartile range ($25th$- $75th$) percentile of the RMSE expressed in breath/min. 64

4.3 Performance of the RR estimation algorithms with respect to the ventilation mode for window of $64s$. N is the number of records. For each method, the RMSE results are given in terms of median and inter-quatile range ($25th$- $75th$) percentile expressed in breath/min. 66

4.4 Performance of RR estimation methods. N is number of RR estimates. RMSE is calculated for each record and expressed in breath/min as median and inter-quartile range ($25th$- $75th$) percentile. We display the mean difference between the RR references and the estimates, with confidence intervals. We also give the correlation between the RR references and estimates. 67

4.5 Performance of RR estimation methods for instantaneous and global RR. N is the number of RR estimations over the whole dataset (% of accepted windows). The RMSE is calculated for each record and expressed in breath/min as median and inter-quartile range ($25th$- $75th$) percentile. We display the mean difference between the references and and the estimates, along with the confidence intervals and the correlations between the RR references and estimates. The percentage of accepted windows is calculated for each file: The mean and ± the standard deviation of these percentages is given by NF 69

4.6 Performance of RR estimation methods for estimating RR. N is the number of files where RR values were estimated. The RMSE is calculated for each record and expressed in breath/min as median and inter-quartile range ($25th$- $75th$) percentile. We display the mean difference between the references and and the estimates, along with the confidence intervals and the correlations between the RR references and estimates. 73

5.1 RMSE performances on Capnobase data when SRQI threshold is fixed to 0.99. 96

5.2 RMSE performances on Reastoc data when the SRQI threshold is fixed to 0.99. 96

Abbreviations

AC Alternating Current.

AF Atrial Fibrillation.

AM Amplitude Modulation.

BW Baseline Wander.

c-ICA Constrained Independent Component Analysis.

CS-Med Consensus Spectrum combined to Median filter.

CS-PPG Consensus Spectrum from PPG.

DC Direct Current.

EMD Empirical Mode Decomposition.

FFT Fast Fourier Transform.

FFT-Med FFT combined to Median filter.

FM Frequency Modulation.

HR Heart Rate.

IMF Intrinsic Mode Functions.

nPAW number of Pulses with Artifacts per Window.

NSR Normal Sinus Rhythm.

OSA Obstructive Sleep Apnea.

PACs Premature Arterial Contraction.

PPG Photoplethysmography.

PSD Power Spectrum Density.

PSG Polysomnography.

PtoP Peak to Peak.

PTT Pulse Transit Time.

PVC Premature Ventricular Contractions.

RDT Random Distortion Testing.

RMSE Root Mean Square Error.

RQI Respiratory Quality Index.

RR Respiratory Rate.

RSA Respiratory Sinus Arrhythmia.

SF-CS Spectral Fusion combined to Consensus Spectrum.

SF-Med Spectral Fusion combined to Median filter.

SPV Systolic Pressure Variation.

SQI Signal Quality Indices.

SRQI Spectral Respiratory Quality Index.

SVM Support Vector Machine.

VFCDM Variable Frequency Complex Demodulation.

CHAPTER 1 Introduction

1.1 Context 1
1.2 Motivation 2
1.3 Objectives 3
1.4 Outline of the thesis 3

1.1 Context

As in many fields, the health industry is strongly impacted by the advent of new technologies. New healthcare technologies present relevant solutions to better control the potential consequences of the life expectancy increase and chronic diseases growth. This digital revolution in medicine is termed by Hood & *al.* [4] the 4P medicine: personalized, predictive, preventive and participatory medicine. In fact, new system approaches will be centered on individual patients characteristics (genetic, environmental, etc.). New analytic, computational and mathematical tools combined with the available and potentially growing amount of health information can predict the risk of developing some diseases and then propose the most appropriate treatments, medications. Prevention and early detection of diseases are rendered possible by optimizing the wellness and improving the people's quality of life with new technologies. With the emergence of Internet and social networks, the patients become the actors of their healthcare and actively participate in the process of their health trouble treatment. New measurement technologies by smart, non-invasive and wearable sensor devices, permits real-time monitoring of patients, efficient and optimal management of resources, faster and more accurate diagnosis.

1.2 Motivation

One of the fields where the use of new technology is promising is the monitoring of physiological signals in emergency care. In fact, the quality of emergency care and resuscitation requires automated devices for anomaly detection and discrimination for the purpose of monitoring the evolution of various physiological signals.

Due to the important workload in emergency and intensive care, the automation of monitoring physiological signals reflecting the patient's condition is essential. Clinicians are thus in need for better decision aid tools. Principal clues for this system are cost efficiency, accuracy, reliability and effectiveness of the signal monitoring. The main difficulty in physiological signal monitoring is the multiplicity of forms that can take the signal from one case to another. In fact, this great variety of cases makes uneasy the use of standard detection methods because no known nominal model is available. In the same way, classification and discrimination approaches are very difficult to implement. Indeed, these approaches require large representative datasets. Despite the great development in automated data recording, several constraints make very difficult and costly the realization of such databases in emergency medicine. We mention for example: the multiplicity of cases encountered, the almost impossibility to label these cases in emergency situations, the security and confidentiality of the patient's information.

In emergency care, the main indicators of patient's health deterioration are the cardiac and respiratory activities. It turns out that performing reliable cardiac and respiratory monitoring from only one non-invasive sensor is possible by exploiting the Photoplethysmography (PPG) signal. PPG measurement is a simple non-invasive technique for measuring blood oxygen saturation through oximeter. New studies are interested in characterizing cardiac and respiratory activities from the temporal and spectral components of the PPG signal. Despite the great assets of such a technique, some constraints limit the spread of the PPG use in clinical setting. In fact, the PPG signal is very sensitive to artifacts resulting from patient's motions and environmental acquisitions. Other limit is the extreme inter- and intra- variability of the PPG signal. In fact, the PPG signal behavior is different from one subject to another and even for the same subject, depending on its health status. As a consequence, methods proposed in the literature suffer from a lack of universality because of this extreme signal variability and also because of the PPG's dependency to subject state and environmental conditions.

1.3 Objectives

This thesis investigates the feasibility of estimating accurate respiratory rate from only one non-invasive physiological signal. The considered signal is the PPG, which is widely used in clinical routine and also in remote home-care. The main aim of this work is to develop robust and universal methods for estimating respiratory rate regardless of the intra- and inter-individual variability that affects PPG features. To overcome the difficulties resulting from the signal variability, we focus on methods that require no hypothesis on the signal distribution. This fact illustrates the fundamental difference between the approach we follow and the usual approaches where distribution parameters are assumed to be unknown deterministic or random. To assess the reproducibility of our algorithms, two databases are used: the Capnobase database, which is an online database, standardly used in the literature to benchmark methods and that we will use in this respect; the Reastoc database, which contains physiological signal recordings from patients in intensive care at the University Hospital of Brest and which will be used to explore the versatility of our approach when facing another dataset.

1.4 Outline of the thesis

This thesis is organized as follows:

Chapter 2 introduces the PPG signal and its main waveform characteristics, by presenting the technical features of its measurement. Then, an overview of the main clinical uses of the PPG signal are exposed. In addition, the constraints inherent to the PPG signal and that limit its use are discussed. Finally, the databases considered in this thesis are also introduced.

In Chapter 3, an adaptive artifact detection method is presented. The method is based on template matching and hypothesis testing by Random Distortion Testing. This hypothesis testing is a novel statistical decision strategy introduced by Pastor & *al.* [5] for diverse applications in signal processing and telecommunication. The RDT is introduced for comparing the pulse template to the signal pulses and then detecting the PPG pulses with artifact. This method requires little parameterization and has the advantage of fitting the model to each considered signal. The results of this method are then discussed and compared to the literature.

In Chapter 4, an overview of existing methods for estimating RR is exposed. Then, we propose some novel spectral methods for Respiratory Rate (RR) estimation, which

aims to get the best spectral peak resolution corresponding to the respiratory rate. The proposed methods are tested on two different datasets Capnobase and Reastoc. The results are analyzed according to patient's conditions such as age and ventilatory mode.

In Chapter 5, we experimentally show that detection of artifacts is not enough to get accurate respiratory rate estimation. In fact, other physiological effects impact the PPG signal and limit the algorithms ability to get accurate respiratory rate. For this reason, we introduce a quality measurement attributed to respiratory rate estimates, in order that the clinician might select only RR values with a large confidence index. In this respect, the Spectral Respiratory Quality Index (SRQI) is defined to qualify the resolution and the dominance of the respiratory frequency peak in each spectrum. Results yielded by this novel criterion are then discussed and compared to existing methods.

Chapter 6 contains the conclusions where we summarize the main results of the thesis and present the possible extensions of the presented methods.

CHAPTER 2 The photoplethysmography signal

2.1 Introduction 6

2.2 PPG measuring characteristics 6

2.2.1 Technical facts about PPG 6

2.2.2 Sites and devices for measuring PPG 7

2.2.3 Measurement protocol and reproducibility 8

2.3 PPG waveform characteristics 8

2.3.1 Pulse characteristics 8

2.3.2 PPG waveform modulations 10

2.3.3 Factors affecting PPG waveform 11

2.3.4 PPG mathematical model 12

2.4 PPG for clinical physiological monitoring 14

2.4.1 Monitoring blood oxygen saturation 14

2.4.2 Monitoring heart activity 15

2.4.3 Monitoring respiratory activity 16

2.4.4 Monitoring hypovolemia 17

2.5 PPG database 17

2.5.1 Sukor Data 18

2.5.2 CapnoBase Data 18

2.5.3 ReaStoc data 19

2.6 Conclusion 20

2.1 Introduction

In recent decades, the desire for miniaturization, ease-of-use, robustness and noninvasive device are key factors for oximeter and Photoplethysmography (PPG) emergence. Many clinical physiological measurements and monitoring use henceforth PPG signals. In this chapter, we introduce general technical facts about PPG. Then, we present its physiological characteristics and how PPG-based techniques may replace heavy devices for monitoring cardiovascular and respiratory activities. Finally, a description of physiological signal databases is set forth. These databases contains annotated PPG signals recorded in different contexts that will be used later to assess the proposed algorithms in this thesis.

2.2 PPG measuring characteristics

2.2.1 Technical facts about PPG

Photoplethysmography is an optical technique introduced by Hertzman and Spealman in 1937 [6]. It measures local blood volume variations in the tissue at the surface of the skin.

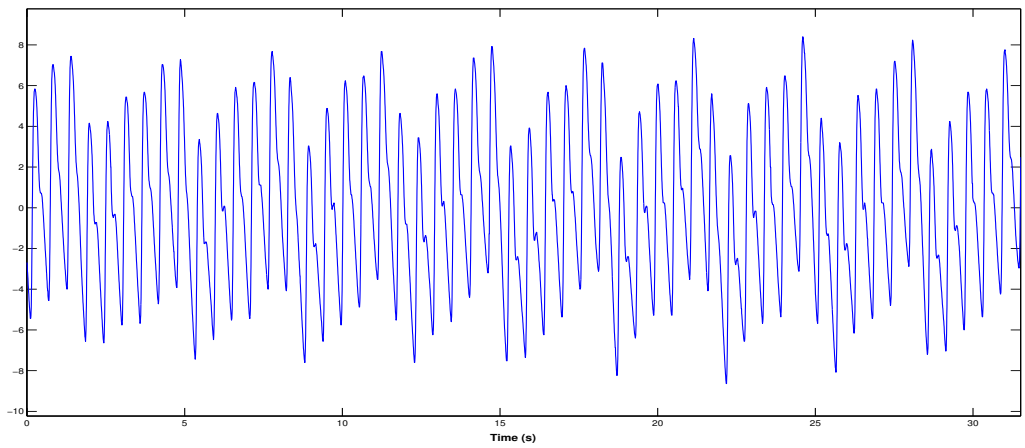


Figure 2.1 – An example of a raw PPG from the Capnobase database: pulsatile component AC and stationary component DC are easily discerned

The principle of this method is simple. A light source emits light which will be dispersed in the tissues and partially absorbed. A photosensitive sensor detects unabsorbed light. Since blood absorption coefficient is higher compared to other underlying tissues, blood flow changes are detected by sensors and reported as a photoplethys-

mogram (PPG)[7]. So, PPG is a representation of the intensity of the detected light (figure 2.1).

The PPG has two major parts:

- The Direct Current (DC) component is the stationary feature of the PPG. It reflects the optical character of the surrounding tissue and the venous blood flows. It fluctuates slowly due to respiration, vasomotor activity and vasoconstrictor waves.
- The Alternating Current (AC) component is the pulsatile feature of the PPG waveform. It corresponds to the arterial pulsations and has its fundamental frequency, typically around the heart rate.

PPG is easily measured from the skin. For wavelengths from 600 to 1300nm, the volume and depth of tissue reached by the light is important. Therefore, most devices use visible or near infrared light. In 1974, Takuo Aoyagi [8] discovered that deoxygenated hemoglobin Hb and oxygenated hemoglobin HbO2 have a distinct difference in absorption of light waves. In fact, HbO2 absorbs more in the infrared wavelengths (850 to 1000nm) than Hb. In the other hand, the Hb absorbs more in the spectrum of Red wavelengths (600 to 750nm) than HbO2.

Since the 1980s, with the developments in semiconductor technology and light emitting diodes, the use of PPG in pulse oximetry has become a standard device for measuring pulse oxygen saturation and providing valuable information about cardiovascular system [9].

The principle of the oximeter is as follows: two light-emitting diodes (LEDs) illuminate the tissue and a sensor detects the light reflected by the tissue at two distinct wavelengths: 660nm (red) and 940nm (infrared) [10] are the most used. Pulse oxygen saturation (SpO2) is calculated by measuring the difference in absorption between oxygenated hemoglobin and deoxygenated hemoglobin. The former tends to absorb infrared light and transmit red light, whereas the later tends to absorb red light and transmits infrared light.

2.2.2 Sites and devices for measuring PPG

In clinical setting, the most used devices for PPG acquisition are pulse oximeters. They are usually worn on the finger but also on the toe and the ear . Depending on the case, it may happen that blood flow to the periphery may be reduced due to hypothermia

or shock, resulting in PPG with no detectable cardiac pulse. In this case, a PPG can be obtained from the ear, nasal septum or forehead [9].

There are other emerging technologies like PPG imaging technology [11]. With the recent advances in digital signal processing technology, PPG sensors are also integrated into wearable devices such as smart watches, smartphones and fitness gadgets [12]. They are mostly used for estimating heart rate.

2.2.3 Measurement protocol and reproducibility

Reproducibility is an essential element to validate accuracy of clinical physiological measurements. Comparison in absolute numbers between PPG records acquired from different subjects or measurement sites is not possible. The reasons are the following ones.

Up to now, there are no recognized standards to normalize PPG clinical measurements. In fact, absorption of light by an oximeter depends on local factors such as sensor placement, subject posture, movement artifact, medical treatment and relaxation [9, 10]. The shape of AC differs depending on the device wavelength: longer wavelengths penetrate deeper into the tissue than shorter ones.

PPG signal characteristics, for example amplitude, should then be considered as arbitrary and only relative comparisons can be made. Published research tends to be based on very different studies and technologies of measurement protocols, thus limiting reliability and relevance of comparison between different methods provided by different research centers.

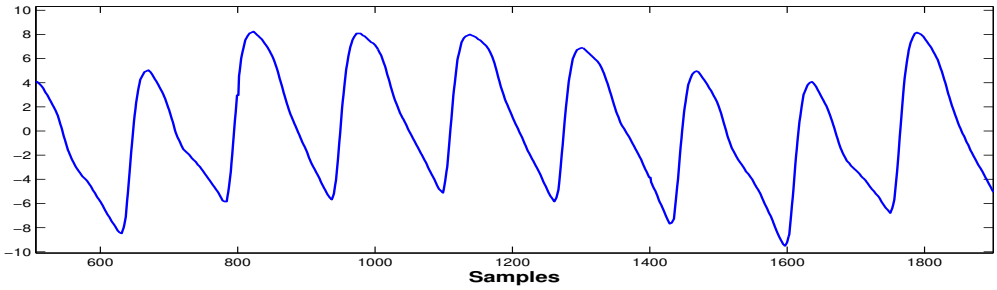
2.3 PPG waveform characteristics

With the development of signal processing techniques and the importance of automation of circulatory and respiratory monitoring, the analysis of the PPG waveform has become of increasing interest for researchers.

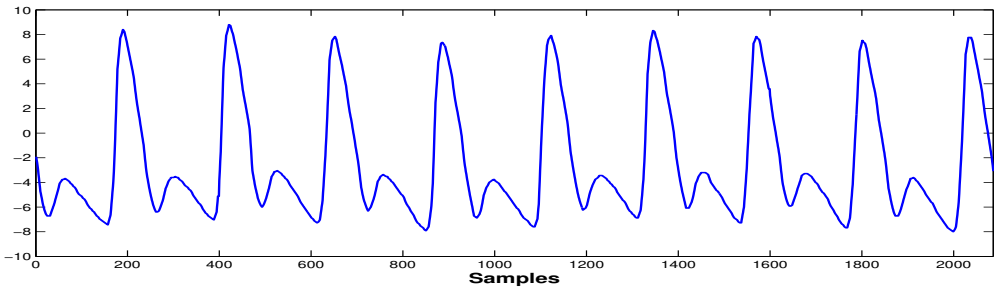
2.3.1 Pulse characteristics

The morphology of PPG pulses can be defined in two phases. The first phase is the rise of the pulse called the anacrotic phase, which corresponds to the systole. The second phase is the descent of the impulse, known as the catacrotic phase, which corresponds

to the diastole and the wave reflections from the periphery. Systole and diastole phases correspond respectively to the heart ventricles contraction and heart refill.



(a) PPG signal with less pronounced notch: systolic and diastolic phases combined.



(b) PPG signal with more pronounced notch: systolic and diastolic phases separated.

Figure 2.2 – Comparison between different shapes of pulses with notch

A dicrotic notch is usually observed in the catacrotic phase of subjects without problems of arterial compliance. It is a small downward deflection in the pulse, marking the end of the systole, which corresponds to the transient increase in aortic pressure following the closure of the aortic valve. Two examples of PPG signals with and without notches are presented in figure 2.2a and figure 2.2b.

From the literature, many features based on the PPG have been studied. Figures 2.4 and 2.3 summarize the main ones:

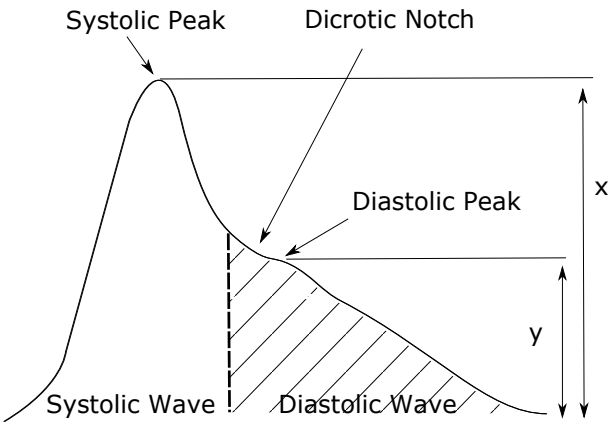


Figure 2.3 – Example of a PPG pulse and its parameters with area of anacrotic/systolic and catacrotic/diastolic phase. x and y are the amplitude of, respectively, systolic and diastolic peak.

- *Peak to Peak:* The distance between the maximum of two consecutive pulses is called Peak to Peak (PtoP) distance. This interval represents a cardiac cycle. Therefore, this feature is used to calculate the heart rate.

- *Amplitude:* Distance between the pulse maximum and the baseline.It is an indicator directly related to arterial blood flow around the measurement site. The amplitude varies according to several factors such as vasoconstruction and variation in blood volume.

- *Pulse Transit Time:*

  It is the distance between the maximum notch time and the maximum pulse time. It corresponds to the time required by the reflected wave to travel from the aorta to the measurement site. Pulse Transit Time (PTT) is highly correlated with the arterial stiffness and elasticity of the vascular walls.

- *Shape:*

  The PPG pulses have a similar shape. However this shape can undergo sudden changes due to movements. These artifacts require pre-processing in order to limit their impact on the information extracted from PPG.

- *Variability:*

  Variability of each of these parameters and their evolution are also important indicators for clinical monitoring.

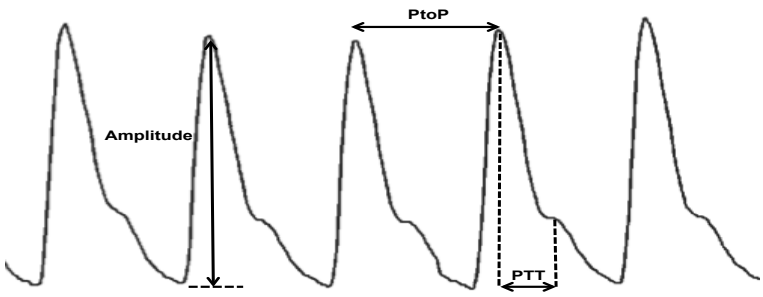


Figure 2.4 – Features of PPG pulse

2.3.2 PPG waveform modulations

Actually, physiologic effects related to respiratory activity and cardiac cycle induce several modulations of PPG (Figure 2.5). Albeit not fully understood, the mechanism mainly consists of three modulations that can briefly be described as follows:

- *Baseline Wander (BW):* also known as respiratory induced intensity variation. In fact, respiration causes blood volume fluctuations in the peripheral vascular bed. The intrathoracic pressure variates as a result of a change in the blood volume and contribution from abdominal and thoracic muscles. This fluctuation induces exchange of blood between the pulmonary and the systemic circulations. This results in a variation of the perfusion baseline.

- *Amplitude Modulation (AM):* also known as respiratory induced amplitude variation. PPG amplitude is also affected by the respiratory rate: it decreases with increased respiratory rate due to variation in peripheral pulse strength.

- *Frequency Modulation (FM):* also known as respiratory induced frequency variation. Heart rate changes induced by respiratory activity is known as Respiratory Sinus Arrhythmia (RSA). In fact, the heart rate increases during inspiration and decreases during expiration.

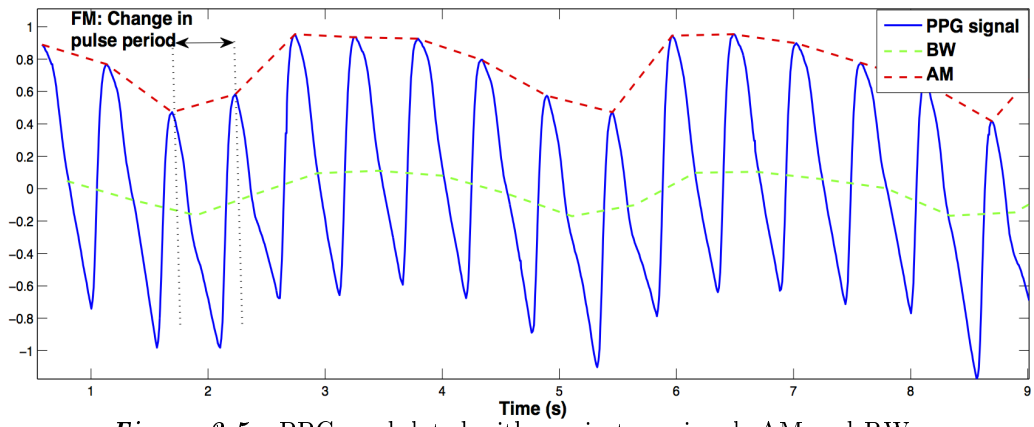


Figure 2.5 – PPG modulated with respiratory signals AM and BW

2.3.3 Factors affecting PPG waveform

PPG waveform could be characterized by several features. However, it should be noted that those features could variate from one subject to another or for the same subject due to health conditions. Arteriosclerosis, hypertension and some dermatoses are some diseases that affect the PPG waveform.

In addition to health conditions, PPG waveform variates with age and gender. In [13], a study analysis how age affects the dicrotic notch and the PTT. In fact, the notch amplitudes are larger for older subjects than younger ones. Also, experiments has shown that PTT is higher in the female group than in the male group.

Vasoconstriction and vasodilatation are physiological mechanisms corresponding to the decrease, inversely the increase, in the diameter of the blood vessels. Those

phenomena affect imperatively the PPG amplitude. In [14], an example of how PPG amplitude increases after infusion of vasodilator Nipride.

2.3.4 PPG mathematical model

Mathematical modeling of the PPG waveform is the center of interest of many researchers. Modeling helps researchers to assess algorithms for PPG processing. In the literature, there are mainly two approaches for PPG modeling, depending on the study purpose: temporal modeling and shape modeling.

2.3.4.1 Modeling based on respiratory modulation

As described above, PPG is modulated by three signals induced by respiratory activity. Using this information, PPG can be simulated as a sinusoidal signal shifted by the baseline with a variable amplitude and frequency.

$$x(t) = b(t) + a(t)cos(2\pi f_{HR}(t))$$

where $a(t)$, $b(t)$ and $f_{HR}(t)$ refer to AM, BW and FM modulation depending on the respiratory rate f_{RR}.

2.3.4.2 Modeling based on pulse morphology

This type of representation is generally used in the literature either in the framework of artifacts detection or for testing blood pressure estimation, since it specifically takes the pulse morphology and the presence of notches into account.

- *Gaussian model*

  In [15, 16], PPG pulses are modeled as a sum of two Gaussians. The first one represents the systolic part and the second one represents the diastolic part.

  $$x(t) = a_1 \exp(-\frac{(t-b_1)^2}{2c_1^2}) + a_2 \exp(-\frac{(t-b_2)^2}{2c_2^2})$$

  where a_1 and a_2 are the amplitude of the curves, b_1 and b_2 are the maxima location and c_1 and c_2 are the width parameters.

- *Logarithmic model*

Gaussian modeling has good accuracy for characterizing pulse maximum and notch but did not take the shape of pulse sides into account. Studies in [17], characterize pulses as long-tail. In fact, contrary to the Gaussian law, a long tail distribution has a large number of occurrences in the extreme ends of the distribution [18]. So, "tails" are longer than Gaussien law.

Every pulse is then decomposed into four long-tails. A Lognormal basis W_i is then used to fit the PPG signal and to improve the model accuracy. The lognormal basis W_i is defined as:

$$W_i(t) = \frac{\alpha_i}{t - t_i} \exp(\frac{(\ln(t - t_i) - \beta_i)^2}{\gamma_i}), \gamma_i < 0, i = 1, 2, 3, 4$$

The pulse waveform is then characterized by a feature vector:

$L = \{\alpha_i, \beta_i, \gamma_i, t_i, t_{end}\}, i = 1, 2, 3, 4$

Successive fitting functions are then used by minimizing the Mean Square Error MSE_i between W_i and original pulse.

2.3.4.3 Limits of mathematical models

PPG mathematical modeling is a tool for researchers to assess algorithms. But, the question remains to what extent these models are close to PPG acquired in real conditions. In fact, model based on respiratory modulation does not take PPG pulse shape parameters into account, which limits its use to only assessing RR estimation algorithm.

The studies around the temporal model of the pulse, ie how the parameters vary over time and how they are affected by respiratory activity, remain limited. In [15], authors combine both shape modeling and temporal modeling. A Gaussian model is applied to parametrize pulse shape and autoregressive moving average is applied for modeling temporal behavior of each pulse shape parameter. This method yields good results to synthesize missing segments from PPG signals and to derive probabilistic distributions of pulse shape parameter. However, the relationship between respiratory modulation, temporal pulse shape parameter evolution and patient conditions is still ambiguous. In fact, these models do not allow the tracking of the respiratory activity.

PPG model should include other parameters relative to subject conditions such as age and gender and combine both respiration and shape modeling. To our knowledge, such models are not studied yet.

2.4 PPG for clinical physiological monitoring

PPG has widespread uses in many clinical settings. The main direct application is the measurement of blood oxygen saturation by pulse oximeter. In the following, we will focus on the monitoring of blood oxygen saturation and other subordinate applications for PPG signal.

2.4.1 Monitoring blood oxygen saturation

As described in Section 2.2.1, pulse oximeters are based on absorption differences between Red (R) and infrared (IR) light waves. The photodetector continuously analyzes the R / IR ratio. From the equivalence Table 2.1 between this R / IR ratio and calibration values of oxygen saturation, the monitor displays the value of the measured oxygen saturation. The correspondence between the values of R/IR and those of SpO2 reported in Table 2.1 are obtained from a calibration algorithm. In fact, the pairs (R/IR,SpO2) with SpO2 measured between 75 and 100% are obtained from experiments on healthy volunteers. For SpO2 below 75%, the displayed values are obtained by extrapolation of the data between 75 and 100% [1].

Table 2.1 – Conventional pulse oximetry algorithm: R / IR: Absorption ratio of the red / infrared light waves. SpO2: Plethysmographic saturation in O2. [1]

R/IR	SpO2(%)
2,5	0
1,75	20
1,60	30
1,50	40
1,25	60
1,00	82
0,75	91
0,67	95
0,43	99
0,40	100

In recent years, monitoring of arterial oxygen saturation with oximeter has elicited a considerable interest essentially in operating room and in intensive care. The main application of pulse oximetry is the detection of hypoxemia, or more specifically, the decrease of oxygen carried in the blood. In operating room, hypoxemias early detection allows faster correction and so improved safety especially in the practice of anesthesia.

In intensive care unit, pulse oximeters are a common indicative tool for assessing patient oxygenation. It is used for adapting the fraction of oxygen inspired and detecting abnormalities in mechanical ventilation[19].

Outside hospital setting, pulse oximeters are used by pilots and mountain climbers operating in unpressurized environment where oxygen levels are low. Oximeters are used for reminding the need for supplement oxygen.

2.4.2 Monitoring heart activity

As described previously, the pulsatile component of PPG is synchronized with the cardiac cycle. Therefore, heart rate can be estimated by calculating the distance between consecutive pulses' maxima . This application of PPG is interesting in clinical setting where the ECG electrodes may be uncomfortable to wear, as for subjects under hemodialysis treatment.

Heart beat irregularity can be then detected using information about pulse morphology and heart rate. In the literature [20, 21, 22], researchers are interested in detecting some types of arrhythmia using a simple PPG with no need to ECG record.

In [22], Premature Ventricular Contractions (PVC) are detected by extracting features characterizing pulse amplitude, dicrotic notch and pulse interval. A linear classifier is then used to classify ventricular premature beats. PPG analysis has similar performance than ECG analysis performance with an accuracy of 99.3%.

PVC are most of the time benign but can be caused by some heart diseases or because of stress or intensified exercise. Atrial Fibrillation (AF) is the most common type of serious arrhythmia. This type of arrhythmia increases the risk for heart failure and stroke. Discriminating AF from other benign heart rhythm turbulence is a major challenge because detection of abnormalities during monitoring heart activity in early stage could help clinician for early diagnosis and enable better treatment.

Arrhythmia discrimination between Normal Sinus Rhythm (NSR), AF, PVC and Premature Arterial Contraction (PACs) were studied in [20], A smartphone application was introduced by classifying PPG time series exploiting pulsatil features (amplitude, frequency), their variability and Poincare plot. A flowchart in [20] details the discrimination procedure. The results show that the proposed algorithm detects NSR with specificity of 0.9886, and differentiates PVCs and PACs from AF with sensitivities of 0.9684 and 0.9783, respectively.

2.4.3 Monitoring respiratory activity

Respiratory Rate (RR) is a physiological signal widely used in clinical care including critical and neonatal care, sleep study evaluation and anaesthetics. It is the most sensitive vital sign marker of clinical deterioration when suspicious changes are detected. Continuous RR monitoring provides highly informative indicator about health status of patients in hospitals, in intensive or emergency care units or at home in case of remote monitoring of chronic diseases and postoperative rehabilitation via mobile sensors.

Despite the relevance of RR in assessing physiological state of ambulatory patients, existing methods for measuring and monitoring RR are still limited by their difficulty of use and their inaccuracy. In fact, placing a tight-fitting band around the patient's thorax is uncomfortable and a non reproductible measure. Other techniques like capnography or acoustic measurements are limited to intensive care units and operating rooms. Some new noncontact respiratory monitoring methods using optical and electromagnetic waves are developed. However, their efficiency is still debatable. In addition, they raise questions about the patients' safety because of the risk of interference with existing medical equipments.

Another approach for measuring RR is proposed: extracting RR from photoplethysmography (PPG). As described in 2.3.2, respiratory rate can be estimated from PPG signal. In addition, respiratory activity is tightly related to blood circulation. In fact, the intensity of modulations could be indicative of thoracic pressure changes that characterize breath effort.

So, not only respiratory rate could be extracted from PPG, but other characteristic features could be extracted from PPG to diagnose related respiratory pathologies. In [23], PPG signal is used to analyze respiratory effort by considering PPG modulations parameter and PTT features. Distinctive monotonic relationships were found between many of the PPG parameters and the coached breathing protocol developed in this experiment.

Another important application for monitoring breathing activity is the detection of sleep disorder. The most common is Obstructive Sleep Apnea (OSA), which is induced by complete or partial obstructions of the upper airway. Despite the effort exerted by the patient to breath, these obstructions are associated with repetitive episodes of superficial or interrupted breathing during sleep. These obstructions are also accompanied by some decrease in blood oxygen saturation. OSA has been widely identified as a major risk factor for many health anomalies causing mortality [24].

OSA is diagnosed by Polysomnography (PSG): It is a multichannel test that supervises physiological signal including heart rate, brain activity, eye movement and muscle activity during a full night in-lab sleeping. It is a costly and heavy procedure for patients and clinicians.

Many studies [10, 25, 26, 27, 28] cite the potential of PPG to replace PSG for detecting OSA. In fact, PPG signal contains many physiological indicators similar to PSG. PTT is inversely proportional to blood pressure. Respiratory rate and respiratory effort could be monitored from the PPG modulations. And finally heart rate could also be extracted from the PPG. Simultaneous experiments with PSG showed that PPG compares well in the diagnosis of OSA.

2.4.4 Monitoring hypovolemia

Non invasive measurement of blood volume is clinically very important. In a perioperative study, Shamir & *al.* [29] demonstrate that during moderate hypovolemia (10% subtraction of the blood mass), respiratory variability from the AM signal extracted from PPG is modified in the same direction and the same amplitude as arterial signals features such as Systolic Pressure Variation (SPV) which is an index used traditionally for detecting hypovolemia. Moreover, AM variation is noticed earlier than in SPV. Studies demonstrate the value of this method in assessing moderate hypovolemia in the perioperative period and during anesthesia.

2.5 PPG database

In this chapter, we present principle characteristics of PPG signals. As first observations, we notice that PPG waveforms depend on several factors. This dependence and the lack of standardization of measurement induce a large variability between subjects and also for the same subject in case of changes of physiological activities or measure circumstances.

As discussed in the following chapters, methods in the literature suffer from lack of generality. In other words, performance differs from one PPG record to another. For this reason, we use three different databases for assessing our algorithms.

2.5.1 Sukor Data

Sukor Data was collected during a study in the University of New South Wales, Sydney, Australia [2]. 13 healthy subjects (10 males and 3 females) aged between 24 and 32 year participated to the study.

Eight pulse oximetry measurements were recorded from each patient's finger. Records were then digitized to $1000Hz$. One minute was allocated to each recording. The protocol of the PPG acquisition is as follows: during the first and last $20s$ of the recording, the hand remained stationary, resting on a table top; in the middle of the recording the subject did movements for approximately $20s$. Eight different movements were conducted for each subject.

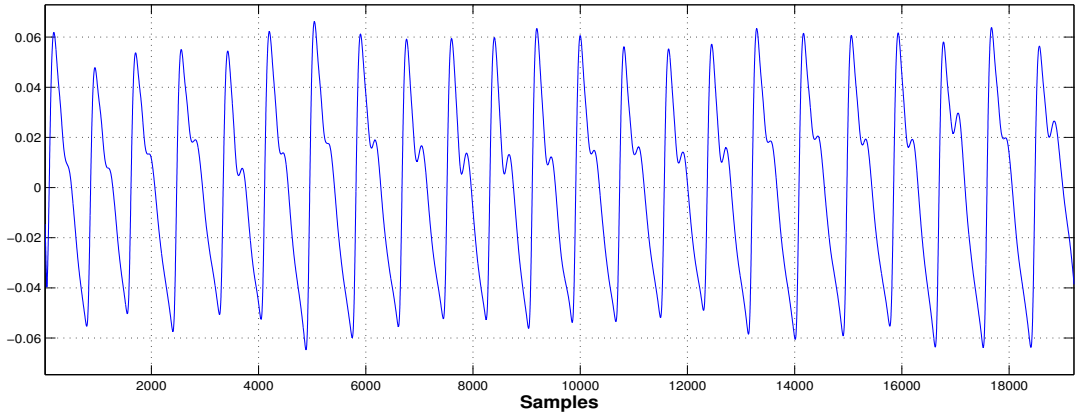


Figure 2.6 – An example of a clean PPG signal segment from Sukor Database

A total of 104 recordings of PPG signals was thus obtained. A manually annotated gold standard was constructed by two independent experts. For every pulse of the records, a score is attributed to describe the quality of the pulse: bad or good.

Sukor database is a good reference for assessing algorithms for artifact detection for the following reasons: It is a relatively large database well annotated; it contains several different types of induced movement artifacts.

2.5.2 CapnoBase Data

CapnoBase is a collaborative research project that provides an online physiological signals benchmark dataset [30]. Benchmark data were selected from physiological signals collected during elective surgery and routine anesthesia from St. Pauls Hospital and British Columbia Children's Hospital in Vancouver. It contains 42 PPG signals of 8

minute recording each one sampled at $300Hz$ acquired from 29 pediatric and 13 adults during spontaneous or controlled breathing.

The data set contains reliable labels about artifacts and simultaneous physiological signal acquisition. Each record includes: age, type of respiration (spontaneous or ventilated), PPG and artifacts label, ECG and artifact label, CO2 signal and instantaneous RR, instantaneous HR derived from PPG and ECG. All reference gold standards were validated by an expert. In addition, results from SmartFusion algorithm [31] were included for each record to allow other researchers to compare performances.

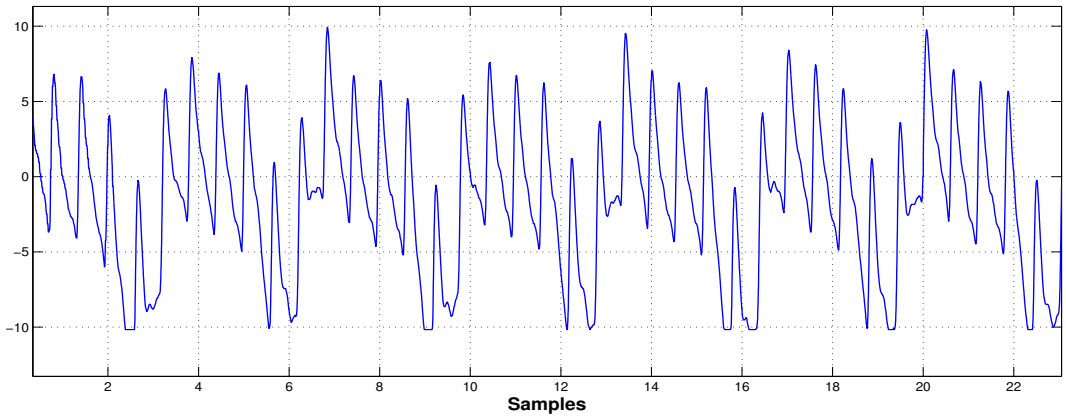


Figure 2.7 – An example of PPG signal segment from Capnobase

Unlike Sukor data, Capnobase is collected from clinical setting. It contains additional information about patients like age and respiratory mode. Records are longer and permit best analysis of PPG parameter evolution over time. Thanks to it labels, it could also be used as a reference for testing many algorithms: artifact detection, respiratory rate estimation, heart estimation.

2.5.3 ReaStoc data

ReaStoc is a research project, which aims to build a database of physiological signals within the intensive care unit of the Brest University Hospital. Signals were collected from 80 patients with different gender, age, cardiovascular anomalies and respiratory mode.

Each record contains ECG, PPG and arterial blood pressure. Data were collected using Synapse tool, which is a software developed by INSERM LTSSI unit in Rennes. Synapse is used to recover physiological signals records from the monitor. They are then filtered, resampled and stored. This database is referenced by the U.S. National Library of Medicine on their website `http:/www.clinicaltrials.gov` under the iden-

tifier (NCT02893462). It is a worldwide database of privately and publicly funded clinical studies.

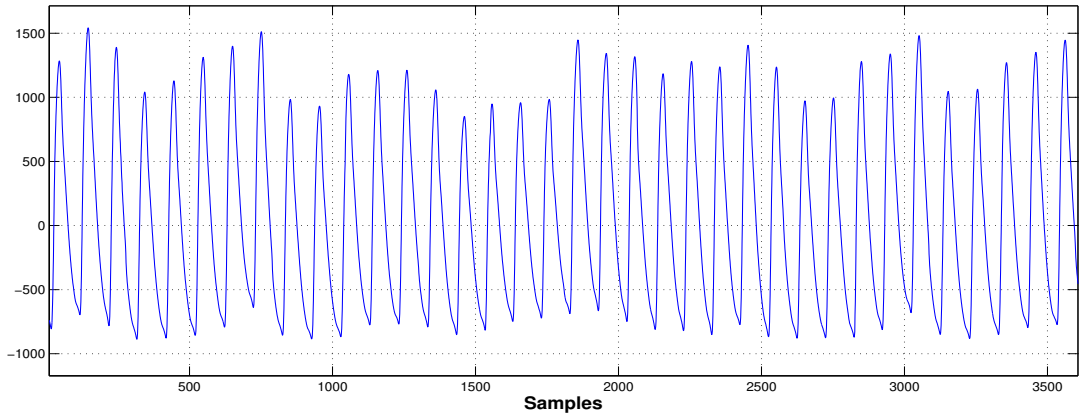


Figure 2.8 – An example of PPG signal segment from ReaStoc

In this study, we will focus on PPG records. Signals are sampled at a sampling rate of $125Hz$ and their lengths are between 5 and 25 minutes. Several records can exist for the same patient.

Every record contains references about patients' antecedent, illness, medical treatment and also average HR and RR for the two first hours. 14 records contain expert labels for artifacts.

This database is still under development and structuring. It aims to collect not only the physiological data but also the data specific to each patient. This data constitutes a great potential in particular for seeking the possible interaction between the evolution of physiological signals and each patient's specificities.

2.6 Conclusion

In this chapter we have introduced the most important characteristics of the PPG signal. In addition to its principal use for measuring oxygen saturation, PPG could be used in clinical settings for monitoring cardiac and respiratory activity. Despite its simplicity of use, the deployment of this technique is still limited by the non reproducibility between subjects, but also for the same subject, due to the signal sensibility to conditions. Several databases are presented to analyze PPG signals behaviour. Variability between records is easily noted when we observe records from different bases (2.6, 2.7, 2.8).

CHAPTER 3 Detection of artifacts in PPG signal

3.1 Introduction 21
3.2 Artifact causes and impact 22
3.3 State of the art 22
3.3.1 Filtering method with PPG restoration 23
3.3.2 Morphology analysis and artifact detection method 27
3.4 RDT for artifact detection 29
3.4.1 Simple artifact detection for short records 30
3.4.2 Adaptive RDT for artifact detection 35
3.5 Results and discussion 37
3.5.1 Simple RDT detection performance 37
3.5.2 Adaptive RDT performance 40
3.5.3 Possible improvement and extension to other physiological signals 42
3.6 Conclusion 42

3.1 Introduction

Heart rate (HR) and respiratory rate (RR) can be estimated from PPG signals. But motion and environmental artifacts cause distortions in PPG signals, which may induce erroneous estimations of HR and RR. So, it is essential to detect segments corresponding to artifacts. Reduced confidence for HR and RR values can then be assigned to these segments.

In this chapter, an overview of methods for artifact detection used in the literature is presented. Then, we will introduce an original method based on pulse template matching and a novel statistical decision strategy: Random Distortion Testing [5], which proved its efficiency in other signal processing applications. Finally, we will discuss results and performance of our method when applied to different databases.

3.2 Artifact causes and impact

As described in 2.2.1, PPG acquisition is based on an optical technique for measuring blood flow. Despite its efficiency for many clinical applications, acquisition of clean PPG is not a trivial task in clinical settings, due to its sensitivity to artifacts.

By observing the morphology of the PPG waveform, we notice that it is composed of pulses with similar forms and slight differences in amplitude and shape. In case of artifacts, the shape of the pulses suddenly changes, which entails signal distortion. Otherwise, when physiological activity varies, for instance in HR or RR, the pulse shape changes are slow and small.

Several factors affect the PPG waveform. Motion artifacts are the most common ones. They are caused by voluntary or involuntary patient's movements. Other factors related to measurement conditions such as misplacement of the oximeter, variation of light and temperature, lead to signal distortion.

Artifacts have unknown distribution and can have several forms. As most physiological signals, PPG is non stationary and its parameters change considerably overtime. So, comparison with a model or reference signals is not obvious. In addition, ordinary filtering methods are inadequate to eliminate artifacts, because there is an overlap between the frequency band of the artifacts and that of useful PPG signals.

3.3 State of the art

Limiting artifact effects on PPG has been a challenging task since the emergence of oximeters. Researchers proposed several approaches on the topic. Some methods [32, 33] use accelerometers to detect patient's movements but have still limited use because of the need for additional sensors. In addition, accelerometers detect motion artifacts only and do not consider other PPG signal distortion.

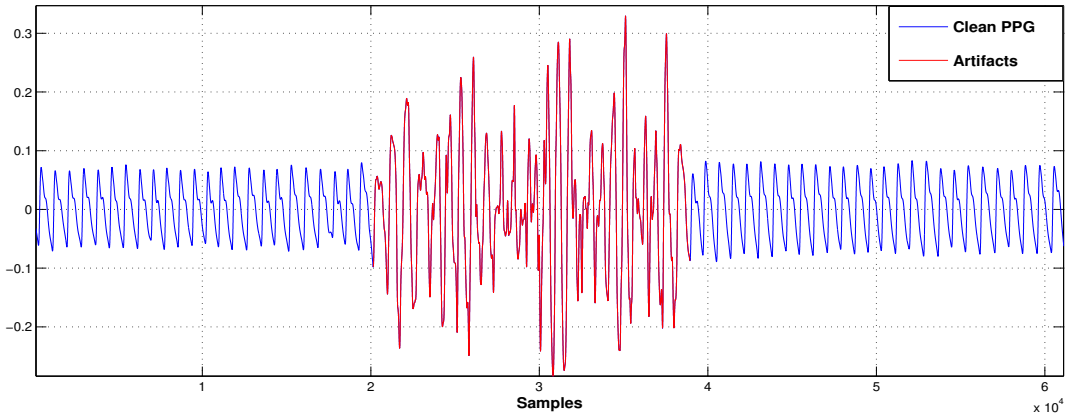


Figure 3.1 – An example of PPG signal with artifacts from Sukor Database [2]

Two strategies for limiting artifact effects using only PPG signals have been presented in the literature. The first one involves using filtering methods in the purpose of reducing artifacts effects and restoring PPG signals. The second one involves detecting segments with artifacts and just removing the polluted sequences. In the following, we present some of the algorithms for both processes.

3.3.1 Filtering method with PPG restoration

3.3.1.1 Adaptive filtering

Adaptive filtering is known to be a powerful tool for denoising signals, especially for in-band noise. It has the advantage of adjusting parameters when the signal changes. Least mean square adaptive filtering is the most common algorithm used in the case of artifact cancellation. Figure 3.2 presents the flowchart of the main steps for methods using adaptive filtering to denoise PPG. The difference between algorithms in the literature lies specially in the selection of the reference signal.

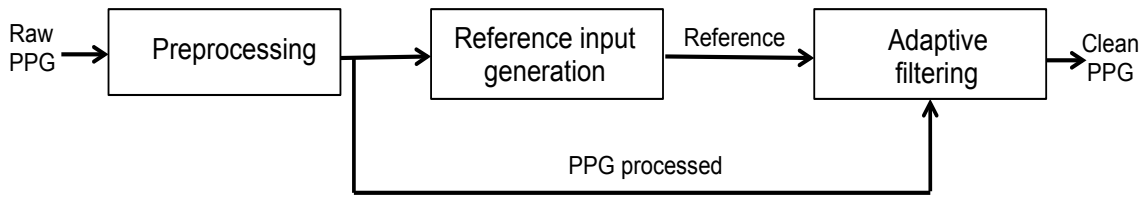


Figure 3.2 – Flow chart for removing artifacts with adaptive filtering method

More precisely, it is assumed that PPG clean signals and PPG artifacts are independent. The first step consists of a preprocessing of the PPG by filtering the signal by a passband filter $[0.1-5Hz]$. The second step consists of generating the reference signal.

In the third step, the reference and filtered PPG signals are introduced as inputs of the adaptive filter so as to obtain the clean PPG.

The critical step in adaptive filtering is the construction of the reference signal. Depending on studies, the reference can be noise or clean PPG. In [34], Ram & *al.* extract reference noise from the PPG spectrum. Peaks corresponding to heart frequency and respiratory frequency are canceled. All other peaks are considered as noise. Inverse FFT is then applied to generate the reference noise signal.

Nevertheless, in [35], Peng & *al.* proposed another approach in which the reference clean PPG is generated by Constrained Independent Component Analysis (c-ICA). In fact, as clean PPG and artifacts are considered as two independent sources, c-ICA can extract the desired signals. Unlike the standard ICA method, prior information is introduced in the c-ICA algorithm. In the case of PPG signals, this information concerns the periodicity of the PPG and it is introduced in the c-ICA algorithm by considering a rectangular signal with a period equivalent to heart rate as a reference .

The method proposed by Ram & *al.* [34] needs prior information of HR and FR, which limits the use of this method when these pieces of information are not available. Besides, the second method proposed by Peng & *al.* [35] used a fixed period for constructing the reference signal. Unfortunately, the PPG periodicity could change over time, which questions the performance of the method in real-time real-life applications.

Although the aforementioned algorithms demonstrate their effectiveness for recovering PPG pulsatile components, the effect of such filtering methods on respiratory modulations is not treated.

3.3.1.2 Wavelet filtering

Wavelet decomposition is a powerful processing tool for discriminating signal irregularities [36]. For this reason, it is a common robust alternative for denoising signals in signal processing field. Signal can be decomposed into different scales by wavelet filter banks: approximation coefficients are obtained after low pass filtering and detailed coefficients are obtained after high pass filtering. The same process is applied to approximation signal again depending on the decomposition level.

In the case of PPG, after applying wavelet decomposition, significant peaks appear in the detailed band in case of artifacts. Raghurami & *al.* [37] exploited this fact and presented a performance comparison of different wavelet types for detecting artifacts.

Denoising PPG signal consists of three key steps as illustrated in figure 3.3:

1. *Wavelet decomposition:* After selecting the wavelet to test and an appropriate level N, compute the wavelet decomposition of x at level N.

2. *Thresholding:* Search for the suitable threshold which will be applied to the detailed coefficients.

3. *Wavelet reconstruction:* Carry out the inverse transform by keeping original approximation coefficients and adding the modified detailed coefficients of levels from 1 to N.

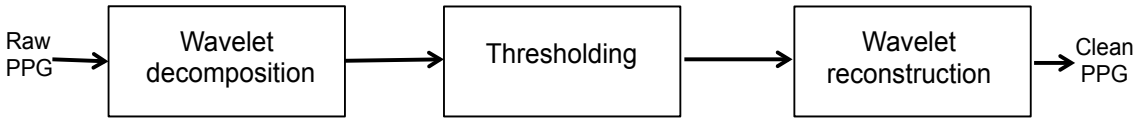


Figure 3.3 – Flow chart for removing artifacts with wavelet method

Experiments in [37], shows that Daubechies wavelets are the most efficient wavelet type for removing artifacts without altering PPG features. In this study, performance of the algorithm is shown by calculating SpO2 values and comparing them to a reference. But, no numerical test was carried out to prove the influence of such algorithm on respiratory rate estimation from PPG. However, the authors only provide us a subjective analysis by observing PPG. In addition, selecting the appropriate threshold is problematic and no indication has been given about the choice of the threshold and the level.

In [38], Nguyen & *al.* present a solution for a similar problem to calculate optimal threshold in the context of detecting abrupt change in the flow signal during assisted mechanical ventilation. After applying wavelet decomposition, the detailed band is considered as composed of gaussian noise and peaks. Peaks correspond to the irregularity in the original signal and should be detected. The thresholding function depends then on the noise standard deviation. The interest of the approach is that the threshold can be chosen so as peaks are detected optimally. The criterion for optimality involves possible deviations around 0. These deviations are resulted from phenomena that are poor of interest to detect.

Such approach could seemingly be a good alternative for thresholding detail coefficients at certain resolution levels for the purpose of detecting artifacts. However, the problem encountered with PPG is that artifacts cannot be restricted to specific frequency bands. So, the choice of the decomposition level and the detail coefficients could not be generalized to all cases.

3.3.1.3 Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) is an appropriate method for analyzing non-linear or non stationary signals. It has the advantage of preserving instantaneous frequency variability [39]. In this context, the PPG signal is considered to be a sum of oscillatory signals with different frequencies [40]. EMD decomposes signals into a

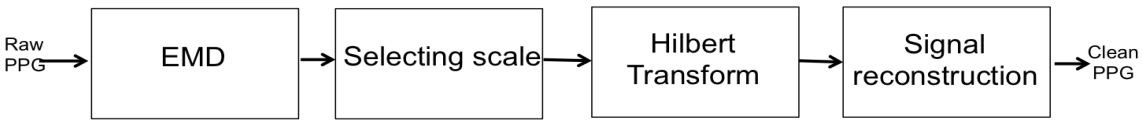


Figure 3.4 – Flow chart for removing artifacts with EMD method

finite number of functions of different scale resolutions called Intrinsic Mode Functions (IMF). An IMF represents a simple oscillatory function which satisfies two criteria: First, the difference between the number of zero crossings and the number of extrema should be one or zero. Second, envelopes defined by local maxima and minima should be symmetric. Unlike harmonic function, an IMFs can have different amplitudes and frequencies over time.

IMF can better identify local features of a signal and adjust it to pulse changes over time. In addition, IMFs are directly extracted from the signal unlike wavelet analysis where we need to select a predefined wavelet function. Figure 3.5 explains the process flow to extract IMF from a given signal S(t).

After extracting IMFs from PPG signals, the first three IMFs extracted from PPG signal are selected. An Hilbert Transform is then applied to each IMF, which permits to identify localized features and extracting the instantaneous frequency $f(t)$ and amplitude $a(t)$.

Clean PPG is then obtain by:

$$x(t) = \sum_{j=1}^{N} a_j(t) \exp(i \int f_j(t)dt)$$

where N is the number of selected IMF scales, $a_j(t)$ is the instantaneous amplitude of the selected IMFs and $\theta_j(t)$ is the instantaneous frequency of these same IMFs.

The results in [40] show a great potential of EMD to restore PPG signals, especially in the context of heartbeat estimation. However, tests, in this study, were conducted on a limited database. In fact, PPG records were collected from healthy subjects which were asked to bend their finger. The motion artifacts are then induced by such movements. In addition, the first three scales were chosen subjectively to reconstruct

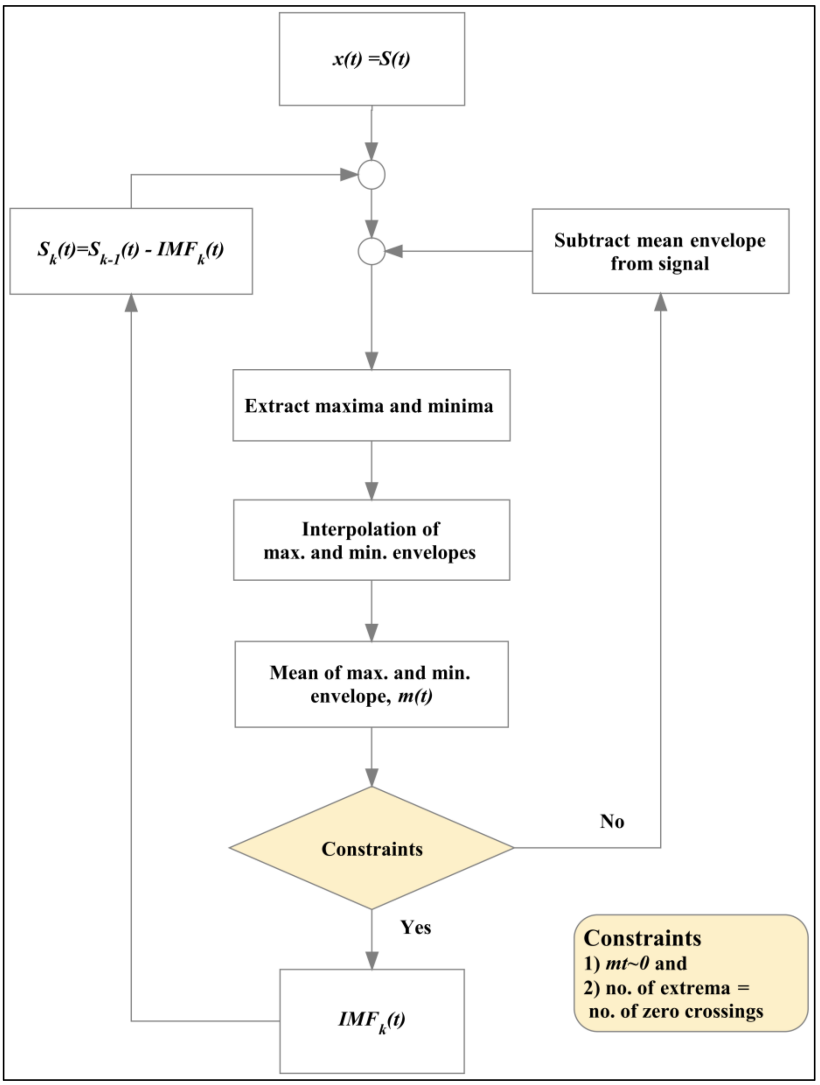


Figure 3.5 – Process flow for extracting IMF from a given signal S(t) (figure from [3])

signals. Performance could be affected in real conditions if the scale selection is not optimized.

3.3.2 Morphology analysis and artifact detection method

3.3.2.1 Morphology comparison

Signal quality metrics are assigned to pulses so that only good quality pulses are kept. Another approach based on a decision tree and predefined thresholding detection is introduced by Sukor & *al.* [2]. Every pulse is compared to a reference by computing the Euclidean distance. Many predefined thresholds are then used to decide if a given

pulse can be accepted or rejected. This method cannot be integrated in a standalone system because templates and thresholds are manually fixed.

3.3.2.2 Classification

Classification methods that combine several waveform morphology features are used in the literature. Pulse morphology can be defined by many features such as amplitude, width, PtoP distance as described in 2.3.1. These features are generally similar and vary slightly. But large variations occur when PPG is corrupted.

From this observation, researchers in [41] extract standard deviation of pulse parameters to characterize pulse variability: PtoP intervals, PtoP amplitude, PTT and pulse shape. Features are then introduced into Support Vector Machine (SVM) to build boundaries, in the purpose to discriminate clean pulses from corrupted ones. Algorithm shows good performance with an accuracy of 94% when tested on a labeled database of 11 healthy volunteers in case of controlled movements and daily activity movements.

In [42], Li and Clifford proposed a method based on template matching and classification. A template is calculated by averaging the pulses extracted from the first $30s$ seconds of each record. Comparison between new pulses and the template is carried out by defining four metrics for the signal quality:

- *Direct matching:* The length of the pulses is fixed equal to that of the template. The first sample of each pulse begins at some fiducial point, for instance, the instant when some noticeable pulse feature (minimum, maximum, etc) occurs. The correlation between the pulse and the template is then calculated.
- *Linear resampling:* Pulses are selected in their entirety. They are then resampled so that the number of samples equals that of the template. The correlation is then calculated.
- *Dynamic time warping:* Pulses are realigned to template using DTW, which is a technique for searching for the optimal matching between two time series. Correlation between pulse and template is then calculated.
- *Clipping detection:* Beats are clipped within a predefined threshold. The percentage of the remained beat is defined as clipping detection.

These four metrics, their fusion, and the number of pulses in each analysis window were then introduced into a multi layer perceptron neural network classifier. This

method yields good results with an accuracy of 95% for detecting pulses with artifacts when tested on a database of 1055 expert-labeled beats. The PPG data were recorded from 104 patient in critical care.

Classification methods have good results for detecting bad pulses on databases on which they were trained. Retraining the classifier to fix its weight must however be remade when new data is acquired from different subjects in different environments are added. So, labeling data should be systematically carried out. This seems hardly feasible for practical purpose because of the extreme variability between individuals and environment measures as well as the long time needed for experts to label signals.

3.4 RDT for artifact detection

Despite methodological consistence, the filtering methods presented previously have limited performance for real time application. In fact, there is no statistical model either for PPG or artifacts and their variations over time. Even worse, useful information about respiratory or cardiac activities can be omitted when applying filtering, even on clean portions of PPG. The second approach based on classification has also its limitations. Classification methods need large scale databases for training issues, which seems complicated. Indeed, until now, there is no standardization or measurement protocol. So, there is a large inter variability between records. In addition, the dynamic nature of physiological behavior causes non stationarity and unknown distribution of the PPG signal. So, there is a need for a standalone system with not many predefined parameters, which could be adapted for every record and could update its parameters when substantial changes occur.

For this purpose, we introduce a novel method for detecting artifacts. This method is based on a pulse morphology comparison and a decision test, with no need for either preliminary information on the observation distribution or any training on a large database.

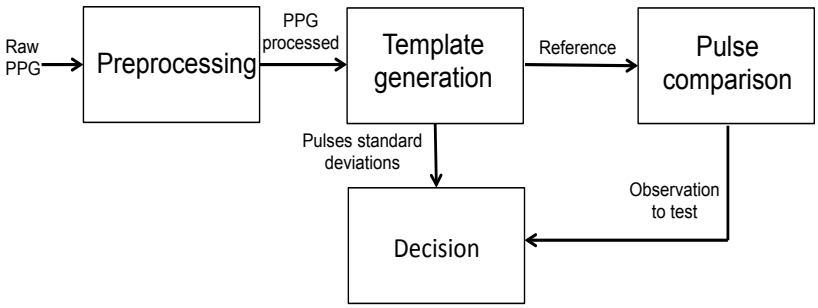


Figure 3.6 – Flow chart for removing artifacts with RDT method

Basically, from the existing literature, template matching methods are best suited to monitor PPG signals with artifacts without damaging useful information. But the problem is still how to choose the most appropriate threshold. As in [2], the general idea involves computing a pulse template and using a metric to compare pulses to this template. In this respect, we introduce a universal strategy to have a standalone system that could be adapted to any record, with no necessity for fixed parameters. The general flowchart of the method is described in figure 3.6 and described in the next section.

3.4.1 Simple artifact detection for short records

3.4.1.1 Preprocessing

Useful information pertaining to the PPG signal is localized in the band $0.05 - 5Hz$. PPG signals are then filtered with a Butterworth pass band filter corresponding to the same band.

3.4.1.2 Pulse segmentation

The pulse to pulse period T_{PtoP} is known to be approximatively the inverse of the heart frequency, which may vary over time. This frequency matches with the maximum peak between 0.5 and $3Hz$ when applying the Fourier transform to PPG. A window of length L is used to have a local estimation of the heart rate frequency f_{HR}. To avoid errors related to local maximum peaks due to signal noise or diastolic peaks [43], blocks of interest are generated using moving averages that demarcate heartbeat areas. More specifically, block limits are estimated by calculating the minima of the resulting signal x_{MA} (figure 3.7).

$$x_{MA} = \frac{1}{W}\left(x\left(\frac{n-(w-1)}{2}\right) + .. + x(n) + .. + x\left(\frac{n+(w-1)}{2}\right)\right)$$

where $W = \frac{T_{PtoP}}{2} = \frac{1}{2f_{HR}}$ is the length of the moving average window, n is the number of samples and x is the PPG signal.

The maximum values of the pulses cannot be estimated directly from x_{MA} because of the delay introduced by moving average. Pulse maxima are then the maximum values of $x(n)$ in each block of interest marked out by calculating the minima of x_{MA}.

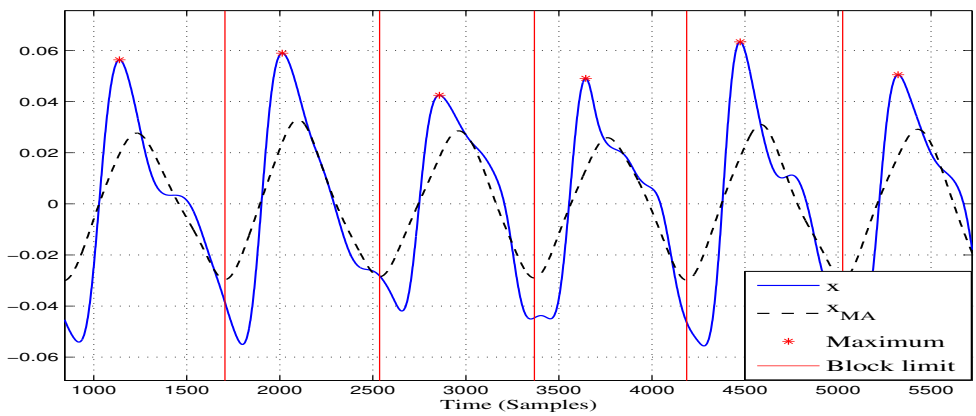


Figure 3.7 – Detecting blocks of interest by applying moving average on x then estimating mimima on the resulting signal x_{MA}. Pulse maximum is the maximum value on x in the selected block of interest.

3.4.1.3 Pulse matching

In this study, the pulse length is calculated on the basis of the Heart Rate (HR). A global estimation of HR by Fast Fourier Transform (FFT) is then carried out to get an average of $PtoP$ by:

$$T_{PtoP} = \frac{1}{f_{HR}}$$

where f_{HR} is HR frequency.

In order to compare pulses, pulses are extracted by maintaining the same length and the same time feature characteristics. We choose to take the pulse maximum as a fiducial point which will mark the middle of each pulse. Pulse lengths are fixed to T_{PtoP} in such a way that, for every pulse P_n considered below, the pulse maximum is localized at $\frac{T_{PtoP}}{2}$.

Pulse matching is then carried out by corresponding every pulse at their maximum point as shown in figure 3.8.

3.4.1.4 Template generation

PPG pulses can have slight variable length during recording. This difference is a result of heart and respiratory variability.

For every record, the first one minute segment is taken to compute the initial pulse template. The choice of a one minute segment follows from our experience with the databases described in Section 2.5. Indeed, within this time frame, sufficient information about pulse morphology variations can be gathered to detect artifacts without

altering the detection performance. This segment is divided into N windows of length L. This is for the purpose to find the most stable window without artifacts.

In each window $i \in \{1, \ldots, N\}$, a local pulse template P_{tp_i} is generated by averaging all the pulses extracted in the window i (figure 3.8).

The correlation coefficient $C_i(n)$ between the pulse template P_{tp_i} and each pulse P_n extracted from window i is calculated. The most stable window then corresponds to the one that yields the maximum correlation between the local pulse template and the pulses within this window. The final pulse template P_{tp} is then defined as the local template corresponding to this most stable window. This computation can be summarized as:

$$j = \underset{i}{argmax}(\overline{C_i} : \overline{C_i} = mean(corr(P_{tp_i}, P_n)), i = 1..N)$$
$$P_{tp} = P_{tp_j}$$

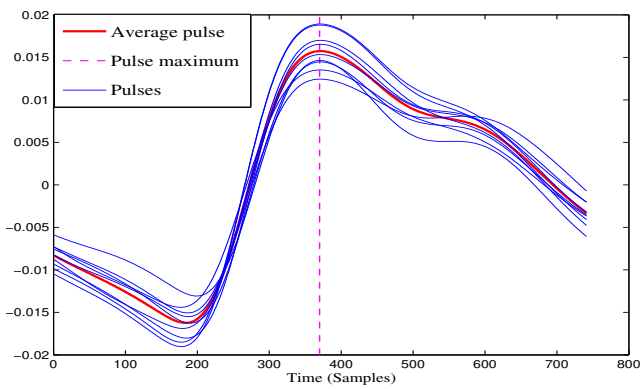


Figure 3.8 – Superposing pulses and calculating pulse template by averaging.

3.4.1.5 Pulse comparison

After selecting the optimal template P_{tp}, a comparison between P_{tp} and all the pulses of the record is carried out. The comparison is performed as follows. T_{PtoP} is now fixed to the length of P_{tp}. The same template matching as described previously is applied: for a given pulse P_k of the record, we match its maximum with that of the template. From either side of this maximum, we select $\frac{T_{PtoP}}{2}$ samples. The correlation coefficient c_k is then calculated between the template P_{tp} and the part of the pulse P_k so extracted.

When an artifact occurs, the correlation coefficients of the pulses affected by this artifact have low values compared to those in clean signals. Thresholding is necessary

to decide which correlation coefficient is acceptable to discriminate good from bad pulses.

However, physiological variability between subjects affects PPG pulse characteristics as described in 2.3.3. Therefore, having a fixed threshold for all PPG signals is not feasible. Threshold should be adapted to every record and depends on pulse variability for each subject.

3.4.1.6 Decision test

Let us consider a correlation coefficient $Y(k)$ between a given clean pulse P_k and the template P_tp. This correlation coefficient can be considered as a sum of a nominal model C_0 and some noise $X(k)$ resulting from measure errors and physiological effect. When artifacts occur, a distortion with unknown distribution $\Delta(k)$ will affect pulse and therefore $Y(k)$. So, $Y(k)$ becomes:

$$Y(k) = C_0 + X(k) + \Delta(k)$$

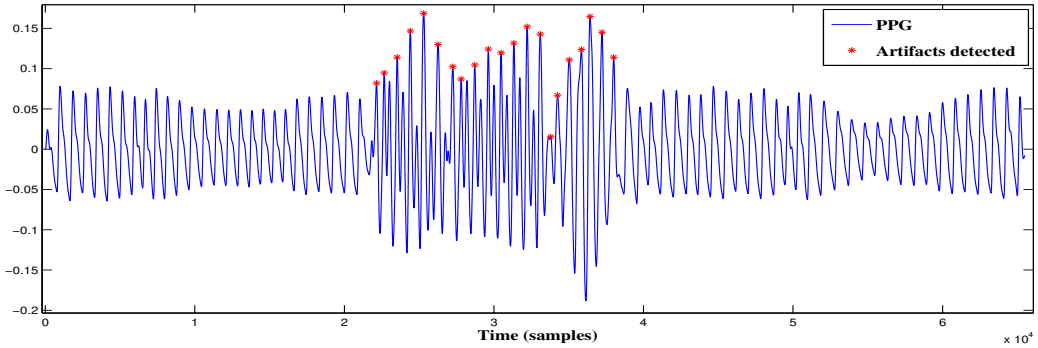
The question now is to know if this distortion $\Delta(k)$ is sufficiently high to consider pulse as a bad pulse. The detection of artifacts then amounts to knowing whether there exists k where the amplitude of the distortion $\Delta(k)$ is sufficiently large to mean the presence of some artifacts.

In order to quantify the amplitude above which a distortion is considered as resulting from the presence of some artifacts, we introduce a non-negative real value τ, called tolerance. If we define the noiseless observation:

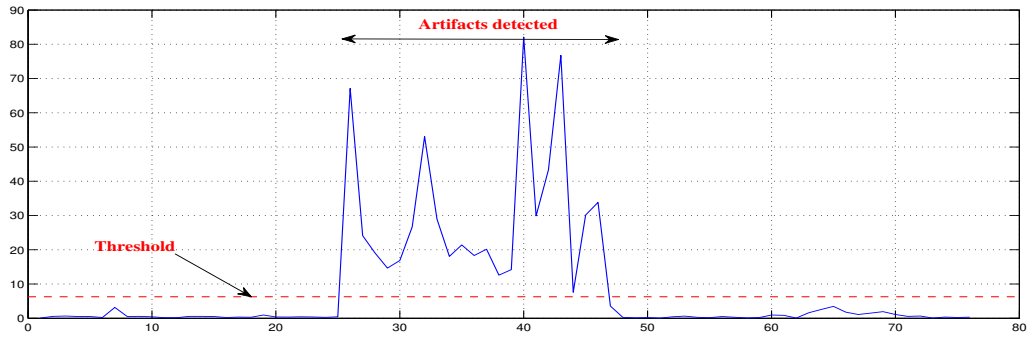
$$\Theta(k) = C_0 + \Delta(k)$$

the problem is then the testing of $|\Delta(k)| \leq \tau$ versus $|\Delta(k)| > \tau$. Let $g : \mathbb{R} \to \mathbb{R}$ be the map defined for every $x \in \mathbb{R}$ by $g(x) = 2C_0 - x$. Basically, we have $|g(\Theta(k)) - C_0| = |\Theta(k) - C_0|$ and $|g(Y(k)) - C_0| = |Y(k) - C_0|$. In addition, $g(\Theta(k) + X(k)) = g(\Theta(k)) + W$ with $W = -X$. Therefore, if W has same distribution as X, the problem of testing $|\Theta(k) - C_0| \leq \tau$ against $|\Theta(k) - C_0| > \tau$ remains unchanged whether we observe $Y(k)$ or $g(Y(k))$. This basic invariance leads to consider the class of all the tests that return the same decision whether we observe $Y(k)$ or $g(Y(k))$. Of course, in this class, we are then interested by finding an optimal test with respect to a certain criterion. Under the additional assumption that noise X is Gaussian distributed with $X \sim \mathcal{N}(0, \sigma^2)$, the solution to this question is provided in [5, 38, 44] where such problem is resolved by a

novel statistical decision strategy introduced by Pastor & *al.* for diverse applications in signal processing and telecommunication. This approach is called Random Distortion Testing (RDT).



(a) PPG signal with gold standard annotation in red, (*) indicate pulses where artifacts were detected by RDT algorithm.



(b) Values of $|Y(k) - C_0|$ compared to the RDT threshold η.

Figure 3.9 – Results of RDT artifact detection on PPG signal

It is a robust non-parametric hypothesis testing which requires no prior information on the signal distribution and could evaluate signal of interest by comparing with a given model.

RDT is a good alternative for physiologic signal processing. In fact, real signals often do not correspond strictly to a nominal model. Independently of noise, there is always some distortion due to physics fluctuation. In addition, in our PPG context, the nominal case is that when there is no artifact because the measure is supposed to be carried out when the patient is motionless. An artifact represents an anomaly. So the purpose is to guarantee a maximum value for the false alarm probability — that is, the probability of erroneously detecting an artifact —, while guaranteeing an optimal detection in case of an artifact.

In our context of artifact detection, the RDT decision test is then:

$$D_{artifact} = \begin{cases} 1 \text{ (artifact) if } |Y(k) - C_0| > \eta \\ 0 \text{ (no artifact) if } |Y(k) - C_0| \leq \eta \end{cases}$$

In this equation, η is the threshold calculated as:

$$\eta = \sigma \lambda_\gamma(\frac{\tau}{\sigma})$$

in which $\lambda_\gamma(\rho)$ is the unique solution in x to equation $1 - [\Phi(x - \rho) - \Phi(-x - \rho)] = \gamma$, $\Phi(.)$ is the standard normal cumulative distribution function, γ is the probability of false alarm and σ^2 is the variance of the additive noise [5].

The window, in which the pulse template Win_{tp} is computed, is considered as the reference signal for estimating nominal model. More specifically, the mean correlation coefficient $\overline{C_{Win_{tp}}}$ computed in this window Win_{tp} is considered as the true correlation coefficient C_0. The slight pulse variability due to physiological effect is characterized in this study by calculating the variance σ^2 of pulse correlation coefficients in a clean segment. This segment corresponds to Win_{tp} where template were generated.

3.4.2 Adaptive RDT for artifact detection

By observing the morphology of the PPG waveform, we notice that it is composed of pulses with similar forms and slight differences in amplitude and shape. This difference is due to physiological effect. In the method introduced previously, the template is automatically estimated on a selected stable segments in order to not skew comparison between pulses. This method is suitable for short records of approximately one minute.

However, a small shape variation is noticed when cardiac rhythm and/or respiratory rhythm change. So using a fixed template and a fixed threshold for long record seems irrelevant. Some pulses could be detected as artifacts although they derive from a little change due to physiological effect. In the following we introduce a new method to automatically update parameter for long PPG record.

In case of artifacts, the shape of the pulses suddenly changes. Otherwise, when physiological conditions change, the pulse shape variation is slow and small. We then extend the method described in Section 3.4.1 by introducing a new strategy that consists in updating the pulse template and sigma used for threshold calculation.

3.4.2.1 Initial template and RDT parameters generation

For every record, the first one minute segment is taken to compute a first pulse template and initialize RDT algorithm. As in Section 3.4.1, P_{tp_1} is selected from the most stable window Win_{tp}. After selecting the optimal template P_{tp_1}, we compute μ_1 and σ_1 which are, respectively the mean and the variance of all correlation coefficients in the window Win_{tp}.

3.4.2.2 Template and RDT parameter updates

A comparison between P_{tp_1} and every pulse k of the record is carried out. The comparison is performed as follows. For a given pulse of the record, we match its maximum with that of the template. From either side of this maximum, we select $\frac{T_{PtoP}}{2}$ samples, where T_{PtoP} is the length of P_{tp_1}. The correlation coefficient c_k is then calculated between the template and the part of the pulse P_k so extracted.

Thresholding is then necessary for decision. As described in the previous method in Section 3.4.1.6, RDT is used for detecting pulses with artifacts. However, for the adaptive case, the nominal model C_0 and the threshold η are adjusted gradually as artifact detection is carried out. In the following, we will note μ_i the nominal model, which represents the mean of the correlations between the template and the clean pulses. σ_i^2 represents the variance of the correlation values between the template and the clean pulses.

The RDT decision test then becomes:

$$D_{artifact} = \begin{cases} 1 \text{ (artifact) if } |Y(k) - \mu_i| > \eta_i \\ 0 \text{ (no artifact) if } |Y(k) - \mu_i| \leq \eta_i \end{cases}$$

In this equation, η_i is the threshold calculated as:

$$\eta_i = \sigma_i \lambda_\gamma(\frac{\tau}{\sigma_i})$$

in which $\lambda_\gamma(\rho)$ is the unique solution in x to equation $1-[\Phi(x-\rho)-\Phi(-x-\rho)] = \gamma$. As above, $\Phi(.)$ is the standard normal cumulative distribution function, γ is the probability of false alarm and σ_i^2 is the additive variance of the noise [5].

If the decision is that the given pulse is free from artifacts, parameters are updated using the following equations where i represents the number of clean pulses detected and used for updating P_{tp_i}, μ_i and σ_i^2:

- *Pulse template update:*

$$P_{tp_i} = ((i-1)/i)P_{tp_{i-1}} + P_k/i$$

- *Nominal model update:*

$$\mu_i = ((i-1)/i)\mu_{i-1} + c_k/i$$

- *Variance update:*

$${\sigma_i}^2 = ((i-1)/i)(\sigma_{i-1}^2 + \mu_{i-1}^2) + c_k^2/i - \mu_i^2$$

The iterative process is summarized in algorithm 1.

Algorithm 1 Updating P_{tp_i}, μ_i, ${\sigma_i}^2$

Require: $i > 2, k > 2, N > 2$ (number of pulses)
  for $k = 1 : N$ **do**
    if decision (P_k) is NO ARTIFACT **then**
      $P_{tp_i} \leftarrow ((i-1)/i)P_{tp_{i-1}} + P_k/i$
      $\mu_i \leftarrow ((i-1)/i)\mu_{i-1} + c_k/i$
      ${\sigma_i}^2 \leftarrow ((i-1)/i)(\sigma_{i-1}^2 + \mu_{i-1}^2) + c_k^2/i - \mu_i^2$
      $i \leftarrow i + 1$
    end if
  end for

3.5 Results and discussion

As already mentioned above, methods described in the literature suffer from lack of generality. In other words, performance differs from one circumstance to another. To assess our algorithm , different databases with different lengths. In fact, performance of simple RDT is tested on short records of Sukor data. However, the adaptive RDT is carried out on long records from Capnobase and ReaStoc. A comparison between the two processes is then presented.

3.5.1 Simple RDT detection performance

After pulse segmentation of records from Sukor data, we obtain 7550 pulses in which 2771 are labeled as bad pulses due to artifacts.

Figure 3.9 shows an example of how the correlation coefficient permits to track high pulse morphology variations. The RDT threshold takes into account low morphology variations resulting from respiratory and vasomotor activities. These variations are involved via the way σ is estimated on clean PPG segments.

The window length L, used to calculate the pulse template, is an important parameter that influences accuracy. In fact, if the window is too small, pulses in the window will be almost the same. In this case, σ will be low and, in turn, the threshold will be even smaller. As a result, the false positive rate will be very high. However, if the window is too large, there is high probability that artifacts affect some pulses. So, the pulse template will be biased, σ will be too high and then, the true positive rate will be too low. Figure 3.10 shows the ROC curve behavior when changing L for $\gamma = 10^{-1}$ and $\tau = 4.5$. The best accuracy point is for $L = 12$s.

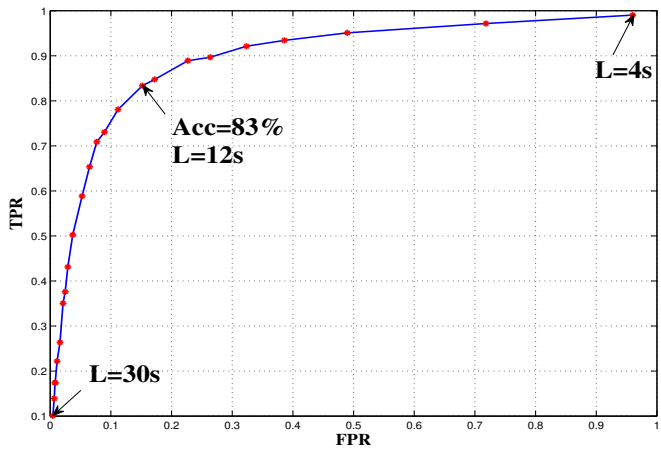


Figure 3.10 – ROC curve of the RDT algorithm for different values of $L = 4 : 30$s, $\gamma = 10^{-1}$ and $\tau = 4.5$

Table 3.1 details performance based on a comparison between the manually annotated gold standard and the output of our algorithm vs. Sukor & *al.*, tested on the same database (104 records). The sensitivity and specificity results of Table 3.1 indicate that, for raw PPG signals, the algorithm has correctly classified about 83% of actual pulses with artifact and 85% of good pulses.

Tolerance τ is an empirical value which can be fixed by the clinician. In theory, as proved in [5], the false alarm rate should be lower than γ. This theoretical result is not satisfied in this study for two possible reasons. First, there are some errors in the database labeling as shown in figure 3.11. Second, assumptions about the presented model are not fulfilled in practice. As a consequence, there is an unavoidable bias between the theoretical result and the practical one. However, as detailed below, the results obtained pinpoint the added-value brought by the method. Indeed,

Table 3.1 – Performances of RDT method compared to Sukor & *al.* $L = 12$s, $\gamma = 10^{-1}$ and $\tau = 4.5$

Performance metrics	Simple RDT	Sukor & *al.*
Sensivity	$84\% \pm 16\%$	$89\% \pm 11\%$
Specificity	$85\% \pm 12\%$	$77\% \pm 19\%$
Accuracy	$83\% \pm 8\%$	$83\% \pm 11\%$

with a reduced number of parameters and a very limited adjustment of these ones, it achieves performance similar to that of methods significantly more demanding in terms of tuning.

The total accuracy of the algorithm is approximately 83%, which is approximately the same result as Sukor & *al.* But we notice that standard deviation obtained is less than that yielded by the approach of Sukor & *al.*, which demonstrates the robustness of our algorithm regardless of the PPG characteristics. In the developed algorithm, only one automatic threshold has been introduced in contrast to the six predefined thresholds of [2].

Moreover, the thresholds used in [2] are empirically chosen so as to achieve the best results on the whole database, whereas the thresholding propounded in our method is adaptively adjusted to the PPG signal under consideration. It must also be noticed that the scores are negatively impacted by questionable annotations of the gold standard. In fact, there is no clear definition of what is a PPG pulse corrupted by an artifact. Even expert human scorers have difficulty in differentiating such pulses. This is clearly illustrated by Figure 3.11, where the gold standard is that artifact is present, whereas the decision made by the test that no artifact is actually present seems correct.

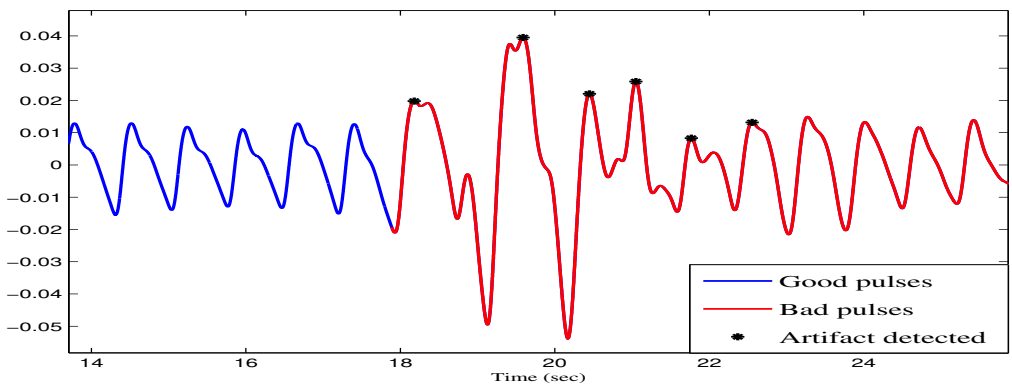


Figure 3.11 – In this example, RDT detects no artifacts however the label indicates that there are bad pulses (red). If we observe the shape of these questionable pulses, we notice no actual artifact.

Other studies using classification methods suffer from lack of generality. Li & *al.* [42] uses a multi-layer perceptron (MLP) neural network to combine several signal quality features. A database of 1055 expert-labeled pulses was used. The weights of the trained MLP were specific to the type of database on which it was trained. If we change the database, for example, when another type of oximeter is used or when the environment where data are recorded is different, the MLP must be retrained. This is hardly acceptable in real conditions such as those encountered in intensive care units, because annotating pulses is a cumbersome task for professional. In addition, as cited before, the labeling may differ from one expert to another. Finally, having a representative database of all possible cases seems impossible due to the variability between individuals.

3.5.2 Adaptive RDT performance

To evaluate the performance of the adaptive RDT method, two databases are used in the purpose to validate the universality of the method, whatever the acquisition devices, the environment and the subject.

3.5.2.1 Capnobase

After pulse segmentation of the whole database (42 records), we obtain 28282 pulses in which 105 are labeled as bad pulses due to artifacts. The portion of bad pulses is quite small compared to that of good pulses. Therefore, the records that contain the largest number of artifacts are selected to check how the algorithm behaves. In this case, we have 4 records with 2297 pulses in which 59 are labeled as bad pulses due to artifacts.

The window length L, used to calculate the pulse template Ptp_i is fixed to $10s$. The RDT parameters are fixed by experiments on a portion of the database as follows: $\gamma = 10^{-3}$ and $\tau = 2$.

Table 3.2 details the performance measurements based on a comparison between the manually annotated gold standard and the output of adaptive RDT vs. simple RDT. Specificity and accuracy are remarkably enhanced when using adaptive RDT. This is explained by the decrease of the false positive rate. In fact, fixed template parameters are too severe when they are used for long records.

Table 3.2 – Capnobase: Performances of adaptive RDT method compared to simple RDT $L = 10$s, $\gamma = 10^{-3}$ and $\tau = 2$

	Portion of the database		Total database	
Performance metrics	adaptive RDT	simple RDT	adaptive RDT	simple RDT
Sensitivity	95%	96%	88%	86%
Specificity	84%	25%	91%	74%
Accuracy	85%	27%	91%	74%

3.5.2.2 ReaStoc

From ReaStoc database, we select 14 artifact-labeled records from patients who have no arrhythmia troubles. After pulse segmentation of the PPG signals, we obtain 19818 pulses in which 626 are labeled as bad pulses. As in the latter case, the records with the most artifacts are selected. In this case, We have 5 records with 8206 pulses, in which 518 are labeled as bad pulses.

The window length L, used to calculate the pulse template Ptp_i, is fixed to 10s. The RDT parameters are fixed experiments on a portion of the database as follows: $\gamma = 10^{-3}$ and $\tau = 1.5$.

Table 3.3 – Reastoc: Performance of adaptive RDT method compared to simple RDT $L = 10$s, $\gamma = 10^{-3}$ and $\tau = 1.5$

	Portion of the database		Total database	
Performance metrics	adaptive RDT	simple RDT	adaptive RDT	simple RDT
Sensitivity	89%	88%	88%	88%
Specificity	90%	74%	92%	80%
Accuracy	90%	75%	92%	80%

As for the Capnobase data, the adaptive RDT method outperforms simple RDT method (table 3.3). In fact, the template is adjusted step by step to signal characteristics without being biased by possible artifacts. We also notice that although the acquisition condition difference of the PPG signals, performance for both databases is still within the same range.

Otherwise, the method of Sukor & *al.* is not tested on Capnobase and Reastoc because of the lack of information about the values of the thresholds used to compare pulses. A training step may be used to fix those thresholds but it is a costly step that cannot be repeated for every new type of data.

3.5.3 Possible improvement and extension to other physiological signals

The physiological signal contamination by artifacts is a common problem in the biomedical engineering field. Without a phase of artifacts detection, the estimates and predictions made on these signals are often biased. If we take signals like electrocardiograph, blood pressure signals and electroencephalograph, we notice common points with PPG. They are quasi-periodic signals. Therefore, there is a pattern which is repeated periodically and which is deformed in case of artifacts.

The three main steps, as used for PPG, remain operative.

- *Template initialization:* choosing the most stable signal segment to compute initial parameter template.
- *Comparison metric:* choosing the metric (correlation, distance..) that will be used to compare the template to the other patterns
- *Adaptive RDT:* using adaptive RDT to make decision if the pattern is good or bad and then updating the template.

3.6 Conclusion

In this chapter, we have introduced the problem of detecting artifacts affecting PPG signals. In this respect, we have presented different approaches encountered in the literature to solve this problem. More specifically, we have introduced our methods based on template matching and decision by RDT testing. In the next chapter, we will focus on respiratory rate estimation and we will then explain how artifact detection could ameliorate RR estimation.

CHAPTER 4 Respiratory rate estimation from PPG

4.1 Introduction 44

4.2 State of art 44

4.2.1 RR estimation from raw PPG 44

4.2.2 RR estimation from derived PPG signals 45

4.2.3 General limits of the existing methods 49

4.3 Consensus spectrum for RR estimation 49

4.4 RR from PPG modulations 52

4.4.1 Extracting PPG modulations 52

4.4.2 Extraction of respiratory modulations 53

4.4.3 RR estimation 55

4.5 Results 60

4.5.1 Results on Capnobase 60

4.5.2 Results on Reastoc 72

4.6 Discussion 75

4.6.1 Comparison between the proposed algorithms 75

4.6.2 Age impact on algorithms performance 76

4.6.3 Comparison with others methods 77

4.7 Conclusion 79

4.1 Introduction

One of the most promising use of PPG signal is the estimation of respiratory rate in different contexts as in clinical setting or in e-health applications for remote monitoring of patients. A wide range of methods from the literature and from our experiments are presented. This chapter set out the behavior of different RR estimation algorithms when applied to different datasets with different age ranges and different respiratory modes. An analysis of the results is then presented to understand why the use of PPG for estimating RR remains skeptical for clinicians.

4.2 State of art

Respiratory rate (RR) is a physiological signal widely used in clinical care. It is the most sensitive vital sign marker of clinical deterioration when suspicious changes are detected. As described in 2.3.2 and 2.4.3, PPG signals obtained from pulse oximeter are a good alternative to traditional methods for measuring RR thanks to the simplicity of use and non invasiveness of pulse oximeter techniques.

In the literature, many methods are proposed for PPG-based RR estimation. These methods can be classified into two categories: RR estimation from raw PPG and RR estimation based on making decision from derived PPG signals.

4.2.1 RR estimation from raw PPG

Cardiovascular and respiratory activity can be characterized by spectral analysis. In fact, HR frequency is known to range in $0.5 - 3Hz$. RR frequency is by the range of $0.15 - 1Hz$. In some cases with "perfect" PPG records, simple FFT is sufficient to get HR and FR. HR and FR, then, correspond to the maximum peak in their respective frequency range. By perfect we mean that signals are free from artifacts and contain distinct modulations. However, in most of the cases, peaks are most of the time drowned in noise especially for RR which is localized in low frequency bands and yields peaks with small amplitude. Other unknown physiological effects can also impact the PPG modulation signals. Advanced techniques are then proposed in the literature.

In [45], researchers introduce the so-called Correntropy Spectral Density to estimate HR and RR. The correntropy function is a nonlinear similarity measure used especially

for non-gaussien signal processing. It is defined by:

$$V(m) = E[\mathcal{K}(x(n) - x(n-m))]$$

where $E[.]$ is the statistical expectation and $\mathcal{K}$ is the gaussian kernel defined by:

$$\mathcal{K}(x(n) - x(n-m)) = \frac{1}{\sqrt{(2\pi\sigma)}} \exp(-\frac{(x(n) - x(n-m))^2}{2\sigma^2})$$

where σ is the kernel parameter. Such function is particularly useful in the impulsive noise environment for detecting nonlinear characteristics in signals. The Correntropy Spectral Density is the Fourier Transform of the Correntropy function.

$$P(f) = \sum_{m=-\infty}^{+\infty} V(m) \exp(-i2\pi fm)$$

HR and RR correspond then to the maximum peak in their corresponding frequency range.

This method is particularly interesting for its robustness in case of artifacts. In fact, $\mathcal{K}(x(n)-x(n-m))$ tends to 0, when either $x(n)$ or $x(n-m)$ is an artifact, whereas the other one is from a clean PPG segment, because of the big difference between the two samples. The median Root Mean Square Error (RMSE) obtained from this method for estimating RR was $0.95(breath/min)$ when tested in capnobase data. However, some outliers still limit the algorithm performance because of harmonic peaks that are sometimes higher than peaks corresponding to RR. An other limitation of this method is the choice of sigma. A calibration step preceded the RR estimation step is necessary when changing database.

4.2.2 RR estimation from derived PPG signals

As described in 2.3.2, physiologic effects related to respiratory activity and cardiac cycle induce several modulations of PPG signals. Based on this fact, most of the algorithms, in the literature, propose as a preliminary step to extract PPG signal modulations. Then, RR values are estimated from these extracted signals. Table 4.1 resumes some of the proposed methods on the topic. Three major steps compose the process of establishing RR:

1. Extracting signals from PPG

2. Estimating RR

3. Analyzing RR and calculating final RR

4.2.2.1 Variable Frequency Complex Demodulation

In [46], Chon *al.* present a method based on Variable Frequency Complex Demodulation (VFCDM). It is a time frequency approach with high resolution that permits to get instantaneous accurate values of signal amplitude and signal frequency. A bank of low pass filters is used to decompose PPG signals into a series of band limited sinusoid modulations. An Hilbert transform is then applied to get instantaneous frequency, amplitude and phase for every modulation. The instantaneous frequency and amplitude are then respectively the FM and AM signals. The RR is then estimated by the maximum peak in the RR frequency band of the Power Spectrum Density (PSD).

Tested on 15 healthy subjects, the presented algorithm reveals that accuracy for estimating RR depends on frequency range. Good performance is observed for low frequencies, corresponding often to RR in the normal range between $0.2 - 0.3Hz$. However, for fast breathing rates $0.4 - 0.6Hz$, the performance is limited but still acceptable. The RR estimated from FM outperforms the ones from AM. So, the authors kept only estimations from FM for the final RR values. Although the method has reliable performance, we believe that the database used for testing the algorithm is limited and not very representative in terms of number of breaths and health situations. In fact, it was tested only on healthy subjects under laboratory conditions.

4.2.2.2 Wavelet transform

In [47, 48, 49], authors introduce a method based on the wavelet transform. AM and FM signals are extracted by detecting ridges corresponding to respiratory band from scalogram, which is the energy density function of the wavelet transform. A weighted averaging method is used to provide confidence metrics for the extracted RR values from AM and FM. This confidence metrics depends on the presence or not of artifacts.

Good performance of this method has been related. A good agreement was found between the reported RR and the reference with a mean difference of $-0.23br/min$ and standard deviation of $1.14br/min$ in case of healthy subjects. However, for test in clinical floor, results are a little less good: $-0.48br/min$ and standard deviation of $1.77br/min$.

The details of the decision process on how weights are attributed to the RR values are not mentioned in the publications of this research study. As a matter of fact, this study is realized in an industrial context, which might explain the lack of details of the proposed method.

4.2.2.3 Smart fusion

In [31], Karlen & *al.* present a method for estimating RR from three respiratory-induced modulations from PPG. The first step involves estimating the maximum and minimum of every pulse. The second step is the estimation of every signal AM, BW and FM as follows. The pulse amplitude is the height between the maximum and the minimum of a given pulse. AM is then the time series of amplitudes of all PPG pulses. BW is the envelope of the PPG. It corresponds to the time series of the pulse maxima. FM is the instantaneous heart frequency. It corresponds to the time series of inverses of peak-to-peak distances. We remind the reader the PtoP distance is the distance between the maxima of two consecutive pulses.

Every modulation is then sampled at $4Hz$. After that, maximal spectral power approach is carried out for each modulation to estimate RR In other words, the respiratory frequency corresponds to the frequency with maximal spectrum power. The final RR is the mean of the three RR estimated from AM, BW and FM. Indeed, a decision process, named Smart Fusion, is introduced to limit errors due to aberrant values. In fact, if the standard deviation of the three RR is greater than $4breaths/min$, the final RR value is rejected.

The smart fusion improves substantially the final RR value compared to simple fusion as well as individual estimation from every signal modulation. Simple fusion involves just calculating the mean of the three RR estimations. The RMSE was by $3 \pm 4.7breaths/min$ when tested on Capnobase data. However, we believe that the reject rate (38.5% of RR values were rejected) of this decision process is too large and then the method is too strict for RR monitoring application. Another drawback of this algorithm is that aberrant values are founded when estimating low frequency respiration rate, especially those below $7breaths/min$.

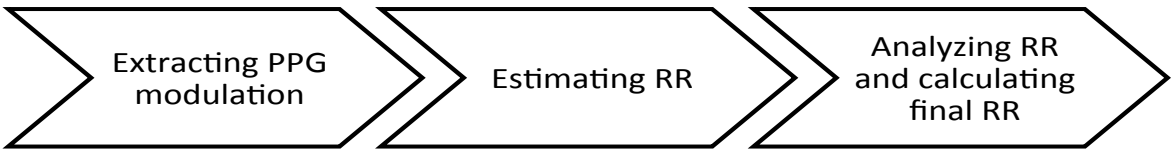


Figure 4.1 – Typical steps of most algorithms for estimating RR

Table 4.1 – Summary of main methods presented in the literature

Method reference	Derived signals	Extracting method	Estimating RR method	Decision making
Chon *al.* [46]	FM ; AM	VFCDM and Hilbert transform	PSD	Only FM is kept
Addison *al.* [47, 48]	FM ; AM	Wavelet	Wavelet	Weighted averaging
Karlen *al.* [31]	FM ; AM ; BW	Pulse segmentation	FFT	Fusion and Smart fusion
Orphanidou *al.* [50]	FM ; AM ; BW	Pulse segmentation	Autoregressive spectral analysis	Pole ranking criterion
Lazaro *al.* [51]	Pulse width variability	Pulse boundary detection	Welch periodogram	Temporal smoothing $RR_i = 0.2RR_{est} + 0.8RR_{i-1}$
Pimentel *al.* [52]	AM, BW, FM	Pulse segmentation	Autoregressive models with multiple model order	Median value depending on model orders

4.2.2.4 Other methods for RR estimation from PPG modulations

Methods based on RR estimation from PPG modulations are widespread. In the literature, slight differences exist between methods presented for estimating RR. The differences reside essentially on how to estimate the RR from the modulation signals. We summarize these methods in table 4.1.

In [53], a review of several methods for estimating RR from ECG and PPG is presented. It is worth noting that RR could be estimated from the ECG. In fact, respiratory activity has analogous effect on ECG features as on PPG ones. So, methods used for ECG can be also applied to PPG. In this respect, the study conducted by Charlton & *al.* is interesting because it assessed every possible combination between RR algorithms for extracting and fusing RR estimations. Comparison is then carried out on the same database and with the same statistical performance measure. The database used in this study was collected from healthy participants.

Results show that algorithms using time domain RR estimations generally perform better than frequency domain techniques. This can be explained by the criterion of quasi-stationarity of the respiratory signal, which is not always ensured. The smart fusion method remains the best way for fusing RR estimation despite of its high rejection rate.

4.2.3 General limits of the existing methods

Existing studies for estimating RR privilege more methods based on PPG modulations than those based on raw PPG. We cannot generalize that the methods of the first category are more reliable than those of the second, up to now, there has been no comparative study in this respect, involving the same database and the same error statistical measures. The fact that the exploitation of the PPG modulation for estimating RR is the most widespread method can be explained intuitively by the fact that exploiting several signals reduces the probability of making errors. In addition, the same methodology applied on ECG has been proven to have good performance [53].

Despite of the broad spectrum study presented in [53], some limitations prevent from having general conclusions about the methods considered in this review. In fact, the methods were tested on one database of healthy volunteers only. Indeed, robustness of the proposed algorithms against artifacts and arrhythmia cases is not presented.

Common limitations have been found for most algorithms for both types of methods. In general, the estimation of RR in very low frequency band is not reliable. Another limitation is that these algorithms do not allow apnea detection, which is a very important vital indicator. Also, universality of the methods, whatever the base, is not demonstrated, especially for the methods which require a predefined parameter like the order for AR methods or the level of decomposition for wavelet transforms. Another critical point is the robustness of a given method in case of artifacts. In fact, the methods presented either work on clean signals or require additional preprocessing block to remove artifacts .

4.3 Consensus spectrum for RR estimation

In this section, we present a method for estimating RR from raw PPG signal. With this method, there is no need to extract the modulations of the PPG signal.

By observing the PPG spectrum, many facts are revealed. Two main peaks are observed in the PPG spectrum. The highest one is localized in the frequency band between 0.5 and $3Hz$. This band corresponds to the heart frequency range and, therefore, the peak corresponds to the subject heart rate. The second peak is localized in the frequency band between 0.015 and $1Hz$, which corresponds to the actual respiratory rate range and the peak corresponds to the subject respiratory frequency. Some other peaks can appear, which may result from artifacts or other physiological activity.

Amplitudes of peaks due to artifacts may exceed or hide those of HR and RR. However, the peaks corresponding to RR and HR are quasi-stationary over time, unlike the other peaks that have a variable behavior.

Based on this observation, we introduce a novel method for smoothing spectra so as to minimize the impact of artifacts on PPG spectrum and accentuate peaks of interest, namely RR and HR.

This method was originally introduced in computational molecular biology and is known as Consensus Spectrum. The method is based on multiple cross spectrum of Fourier transform in the purpose of finding common frequencies between different spectra [54, 55].

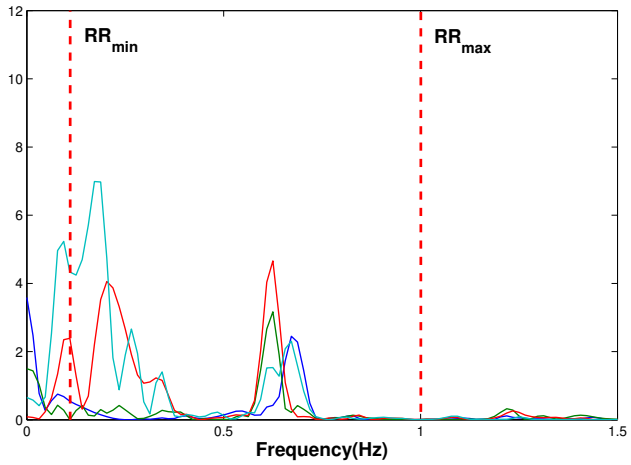
In the case of PPG signals, Consensus Spectrum is the result obtained by multiplying k consecutive PSD:

$$C = \prod_{i=1}^{k} |P_i|$$

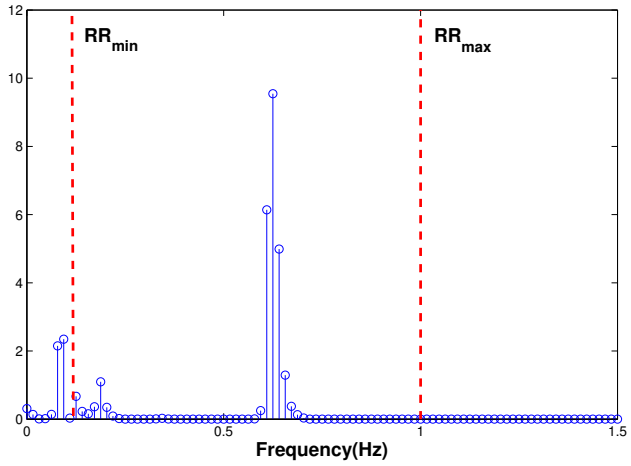
where P_i is the PSD of PPG signal for the window i.

As it can be noticed in figure 4.2, resolution of stationary peaks is increased, whereas other peaks induced by artifacts are reduced. RR then corresponds to the maximum value of the resulting spectrum C in the frequency band corresponding to the specified RR range.

This method has interesting advantages in comparison to those cited in the literature. In fact, the consensus spectrum method needs very few predefined parameters. Indeed, only the length of the FFT and the overlap between 2 consecutive FFT windows are required. Another advantage is that there is no need of pulse segmentation or modulation extraction, which minimizes the algorithm implementation cost. In addition, with this type of smoothing, little peaks induced by artifacts tend to disappear. Therefore, the RR estimation performance is less affected by the presence of localized artifacts.



(a) Illustration of consecutive spectra of a PPG before smoothing where respiratory band is delimited with red lines: many irregular peaks appear, however, peaks around $0.6Hz$ are more stable over time



(b) Spectrum of a PPG after smoothing on respiratory band : irregular peaks noticed in figure (a) are reduced. However a well distinguished peak is observed around $0.6Hz$, which corresponds to RR

Figure 4.2 – Comparison between simple PSD vs. Consensus of four spectra: Peaks are well differentiated in (b).

4.4 RR from PPG modulations

The general flow of the proposed method is based on three principal steps as in 4.2.2 and figure 4.1: extraction of respiratory modulations, RR estimation, analyzing RR and calculating final RR. No artifact detection is taken into account in the proposed method. In the following, an optimization of every step is proposed to get a robust stand-alone system for RR real time monitoring.

In the sequel, the following notation is used to distinguish between the local values of estimated RR from the induced respiratory signals and the final RR values given to the clinician.

- RR_{AM} is the value of RR estimated from AM signal.
- RR_{FM} is the value of RR estimated from FM signal.
- RR_{BW} is the value of RR estimated from BW signal.
- RR_{Fin} *is the final value of RR*

4.4.1 Extracting PPG modulations

4.4.1.1 Preprocessing

Similar to artifact detection in chapter 2, PPG signals need to be filtered in the useful signal range band. All records are filtered with a Butterworth passband filter with frequency band equal to $0.05 - 5Hz$.

4.4.1.2 Pulse detection

Classic methods for peak detection using simple derivative have limits when applied to PPG signals. In fact, pulse shape may differ from one subject to another, depending on age and cardiovascular dysfunction. PPG notches (second peak) tend to be less pronounced when aging. To avoid the detection of notches instead of pulses, the method based on blocks of interest for detecting peaks, described previously in 3.4.1.2, is also applied in this section.

In brief, the pulse to pulse period T_{PtoP} is known to be approximately the inverse of the heart rate (HR). When applying the Fourier transform to PPG, the heart rate can then be estimated by seeking the maximum peak between $0.5 - 3Hz$. The PPG signal

x is then smoothed by applying a moving average filter. Peaks cannot be immediately detected from the smoothed signals x_{AM} because of the delay introduced by the moving average filter. So, we propose to demarcate pulses approximately by calculating the minimum of the resulting signal x_{MA}. These demarcated limits are blocks of interest. Maximum of every pulse corresponds then to the maximum of signal x in each block of interest.

4.4.2 Extraction of respiratory modulations

Respiration waveforms modulation AM, FM and BW are extracted from the PPG. Let us consider peaks as a series of pairs,(t_{pk_i},x_{pk_i}) where i is the pulse index. Let N_{pk} be the number of peaks. Similarly, let us consider the pairs (t_{tr_i},x_{tr_i}) of troughs and denote by N_{tr} the number of troughs. The number of troughs and peaks, N_{tr} and N_{pk}, are equal or differ at most by one. For simplification reason, we will denote by N the number of pulses with $N = min(N_{pk}, N_{tr})$.

- *AM* is the value of every pulse amplitude. It corresponds to the difference between the maximum and its corresponding minimum.

  $$x_{AM} = |x_{pk_i} - x_{tr_i}|, i = 1, 2, .., N$$

  A time series t_{AM} is defined for the signal x_{AM} as follows

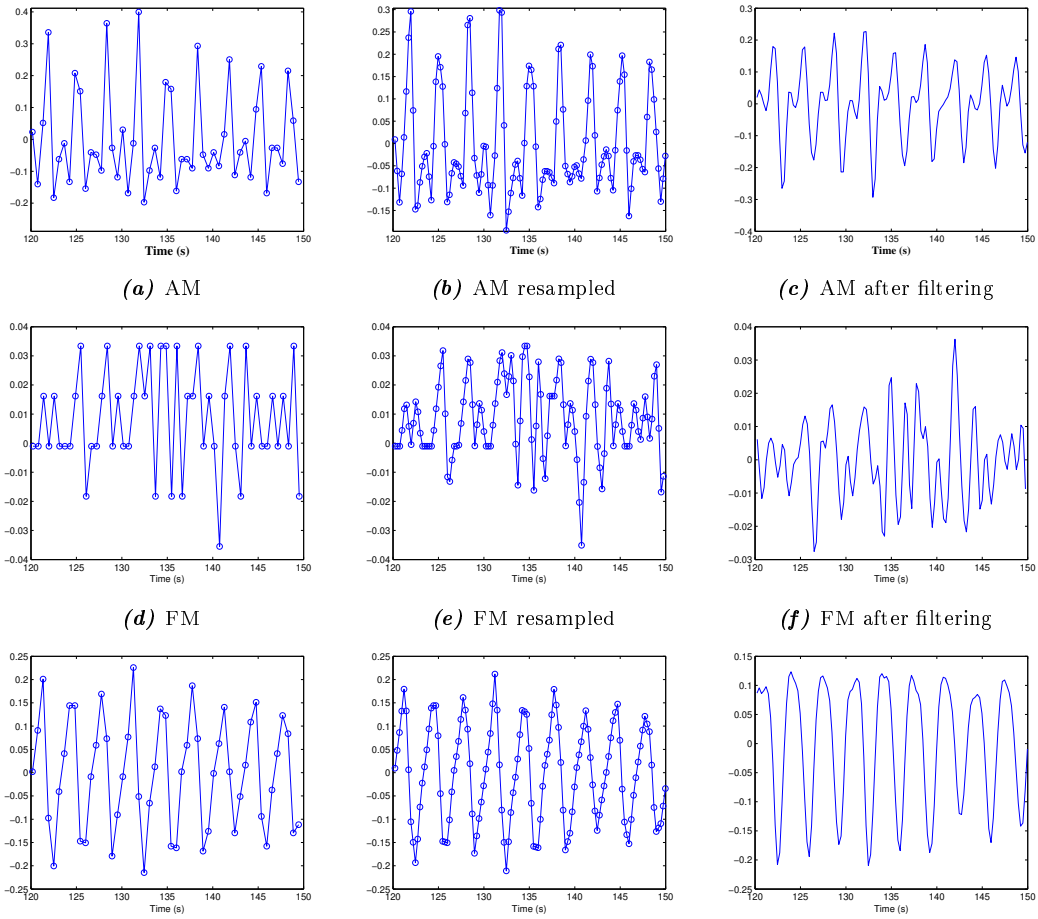
  $$t_{AM} = \frac{t_{pk_i} + t_{tr_i}}{2}, i = 1, 2, .., N$$

  This time series will be useful later for comparison purpose with the reference signal.

- *FM* is the instantaneous heart frequency. It corresponds to the inverse of the peak to peak distance. The pair (x_{FM}, t_{FM}) is thus defined by:

  $$x_{FM} = |x_{pk_{i+1}} - x_{pk_i}|, i = 1, 2, \ldots, N-1$$
  $$t_{FM} = \frac{t_{pk_i} + t_{pk_{i+1}}}{2}, i = 1, 2, .., N-1$$

- *BW* is the time series of the means between the amplitude of each pulse peak and the amplitude of its corresponding trough, which reflects the changes in the signal



(a) AM ***(b)*** AM resampled ***(c)*** AM after filtering

(d) FM ***(e)*** FM resampled ***(f)*** FM after filtering

(g) BW ***(h)*** BW resampled ***(i)*** BW after filtering

Figure 4.3 – Each signal AM, FM and BW (a,d,g) is interpolated and oversampled at frequency $4Hz$ to get signals over time (b,e,h). Finally, every signal is filtered to eliminate noise resulted from sampling (c,f,i)

baseline. The pair (x_{BW}, t_{BW}) is therefore defined by:

$$x_{BW} = \frac{x_{pk_i} + x_{tri}}{2}, i = 1, 2, .., N$$

$$t_{BW} = \frac{t_{pk_i} + t_{tri}}{2}, i = 1, 2, .., N$$

The obtained signals AM, FM and BW are by construction heterogeneous. In addition, samples are not regularly spaced, which prevents the proper use of some processing tools, especially the Fourier transform that need evenly sampled data. Each signal is then resampled to $4Hz$ using linear interpolation. Some high and low frequencies can arise due to physiological effects and linear interpolation. To reduce the imperfections resulting from the previous process, each signal is filtered with a 5th-order Butterworth bandpass filter between 0.83 and $1Hz$, which corresponds to the respiratory frequency

band. Figure 4.3 presents an example of the impact of each step on the three modulations AM, FM and BW. The length of the three signals can be slightly different. Signals are then truncated to the minimum length of the AM, BW and FM signals.

4.4.3 RR estimation

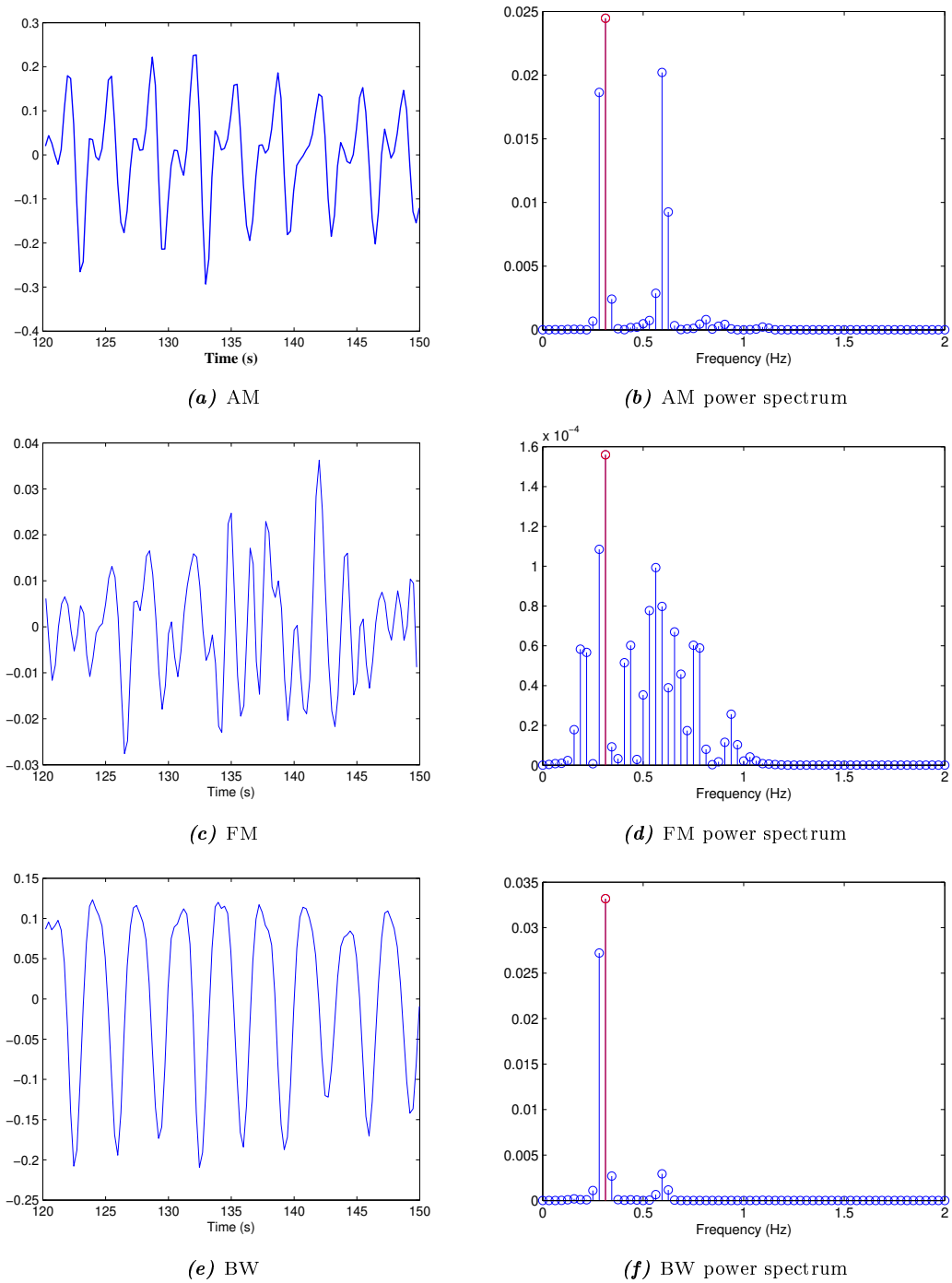
The estimation of the respiratory rate from the modulation signals presents two major problems: the first is how to calculate the respiratory frequency from every modulation and the second is how to fuse RR_{AM}, RR_{FM} and RR_{BW} and making decision on the final value of the respiratory rate. In the following, we present three different methods for estimating the final RR from the respiratory modulation signals.

4.4.3.1 FFT and median filtering for RR estimation

First, a spectral method is used in this step to extract RR. A Hanning window with length L is applied to signals. An FFT is applied to every window then power spectrum is calculated. The same strategy is then used to extract RR from spectrum: Respiratory frequency is the maximum between 0.1 and $1Hz$. In some cases, the heart frequency can be less than $1Hz$. Physiologically, it is not possible to have RR superior than HR. In this case, the respiratory frequency range is then between $0.1Hz$ and the heart frequency.

In the study presented in Karlen & *al.* [31], the mean of the three signal modulations is used to fuse values. For smart fusion, if the standard deviation of the three RR extracted from the three signal modulations is more than $4breath/min$, the final RR estimation value is considered as aberrant. Using simple mean, as Karlen & *al.*, seems to bias final values. Indeed, the standard deviation criterion for smart fusion is too strict. In fact, experiments pinpoint that respiratory rate estimation from the three signals can include outliers (aberrant values of RR) from one signal or several of them.

For example, we have a PPG signal which reference RR value is $20breath/min$. RR_{AM}, RR_{BW} and RR_{FM} are estimated as described previously. We obtained the three following RR values from the modulation signals: 20, 20 and $10breath/min$; the mean of these values is $16.66breath/min$ and the standard deviation is $5.77breath/min$. Using the criterion of Karlen & *al.*, this final RR estimation is rejected because the standard deviation is above $4breath/min$. Intuitively, the use of the median is more robust. In fact, the median is better suited to calculate central tendency for RR



(a) AM

(b) AM power spectrum

(c) FM

(d) FM power spectrum

(e) BW

(f) BW power spectrum

Figure 4.4 – Example of a $32s$ window used for RR estimation from the induced respiratory signals resampled and filtered. The power spectrum is calculated for each respiratory modulation signal (right column) and the maximum power (red line) is selected. In this example, the three respiratory frequencies coincide at the same frequency $0.3125Hz$, which corresponds to $18.75breath/min$

values. If we take the latter example, the median of the RR values is $20breath/min$ which corresponds to the real RR given by the reference.

4.4.3.2 Consensus spectrum and median for RR estimation

In this section, the consensus spectrum method, presented in 4.3, is used for every respiratory induced signal AM, BW and FM. A sliding window with a predefined overlap is applied to the signals. With this method, the respiratory peak in the spectra is accentuated, whereas other peaks resulting from artifacts are masked.

To estimate the final value of RR, the median is used to fuse the three values obtained from the respiratory induced signals. In some cases like the example presented in figure 4.5, the peak of the harmonic in the FM consensus spectrum is higher than the peak corresponding to the real value of RR. In the other two spectra, the respiratory peak is successfully localized. By calculating the median of the three RR, the final RR estimation RR_{fin} corresponds to the reference value.

4.4.3.3 Spectral fusion for RR estimation

As noticed in the precedent example in figure 4.5, the peaks resulting from harmonics are a challenging problem. In fact, sometimes these peaks are higher than that corresponding to the actual respiratory frequency. So, keeping only maximum peak from every AM, BW and FM spectrum can hide some useful information about other potential peaks that may also correspond to respiratory frequency.

The question is then to test if the fusion of spectra could be more efficient than fusing values estimated from the three spectra. In this section, we propose to fuse spectra of the three modulation signals in order to keep all the potential peak candidates that may correspond to the actual RR peak.

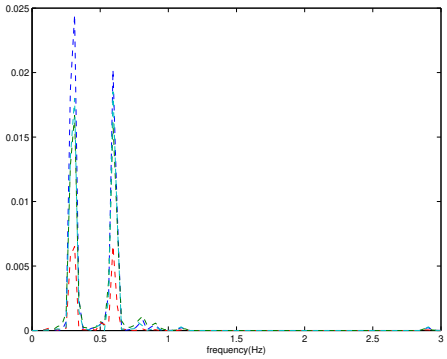
Like previous methods, the FFT is calculated by using a sliding Hanning window for every modulation signal. The spectral amplitudes from AM, FM and BW are not on the same magnitude order. So, a normalization step is carried out. Every spectrum is normalized as follows:

$$P_{XN} = \frac{P_X}{P_{Xmax}}$$

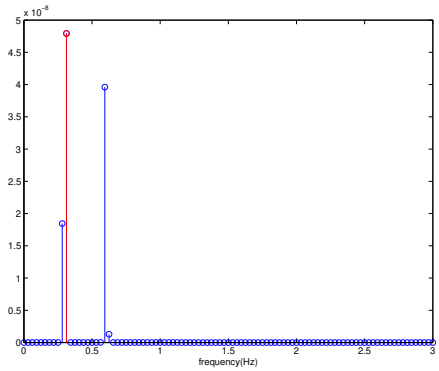
where P_{Xmax} is the maximum of the given power spectrum P_X of signal X, which is AM, PM or BW.

A spectral fusion step is then performed to take the three different estimates into account. We propose two techniques for this purpose.

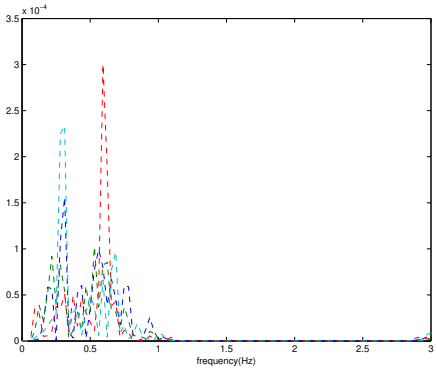
- Fusion with median filter



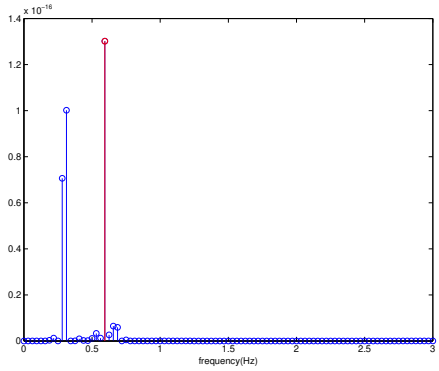
(a) Four successive FFT of AM



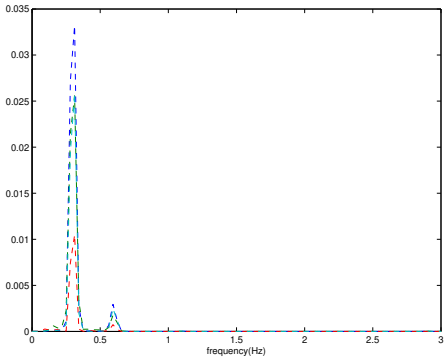
(b) AM consensus spectrum



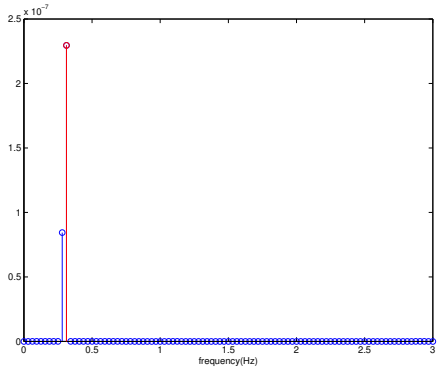
(c) Four successive FFT of FM



(d) FM consensus spectrum

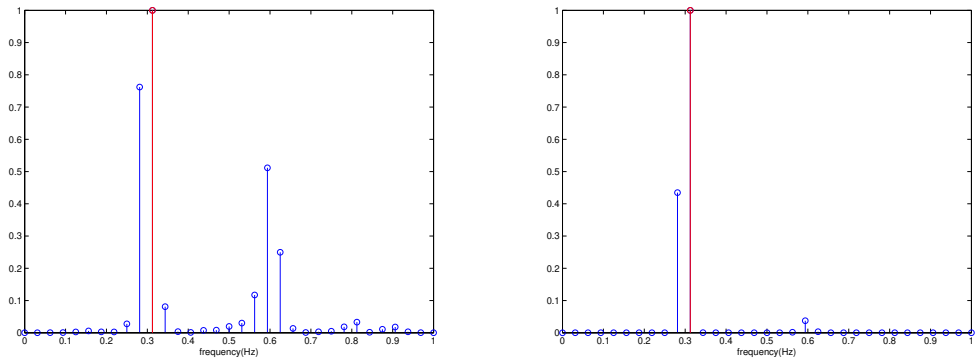


(e) Four successive FFT of BW



(f) BW consensus spectrum

Figure 4.5 – Example of four spectra from four $32s$ sliding windows with an overlap of $16s$ (left column). The resulted consensus spectrum for each signal modulation is presented in the right column and the maximum peak (red line) is selected. In this example, two respiratory frequencies coincide at the same frequency $0.3125Hz$, which corresponds to $18.75breath/min$. However, in the FM signal, the maximum coincides with an harmonic at $0.6Hz$.



(a) Spectral fusion with median filter

(b) Spectral fusion with consensus spectrum

Figure 4.6 – Example of a fused spectrum from the windowed FFTs of the AM, FM and BW signals. The respiratory frequency is estimated at 0.3125 Hz, which corresponds to 18.75 breath/min. The harmonic observed in Figure 4.5d is dimmed in this case

The median filter can eliminate outliers, in contrast to the mean, which can be contaminated with outliers. For every frequency from each of the three spectra, a median value is calculated as follows:

$$P_{\text{Fus}}^{\text{med}}(i) = \text{median}(P_{AM}(i), P_{FM}(i), P_{BW}(i)), \text{ for } i = 1..L$$

with L is the length of the spectrum and i is the frequency bin index.

- Fusion with consensus spectrum

  Consensus spectrum is another way for seeking common frequency between the three spectra. The three spectra from AM, FM and BW are multiplied by each other, frequency per frequency.

$$P_{\text{Fus}}^{\text{cons}}(i) = (P_{AM}(i) \times P_{FM}(i) \times P_{BW}(i)), \text{ for } i = 1..L$$

with L is the length of the spectrum and i is the frequency bin index.

The final estimate RR_{fin} is then given by the frequency with the maximum amplitude in the resulting spectrum P_{fus}. Figure 4.6 shows the resulting spectra from the two methods. A well distinguished peak is observed at frequency of $0.312Hz$, which corresponds to the real respiratory rate 18.75breath/min for this example.

4.5 Results

In this section, performance measurements of the algorithms introduced above are given and commented. Several tools are presented to compare the estimated RR to the reference RR values. Two different databases are considered in this study: Capnobase and Reastoc. These databases are described in Section 2.5.

4.5.1 Results on Capnobase

Capnobase is a collaborative database collected from clinical setting (section 2.5.2). For assessing the proposed algorithms, we consider the whole Capnobase except one file, namely file $'0031_8min'$ of which RR reference file contains many aberrant values. In summary, we used 41 PPG signals. Each signal was sampled at $300Hz$ and lasts 8 minutes. The data were acquired from different subjects with different ages, during spontaneous or controlled breathing. For each record, the RR reference values were extracted from Capnography. In fact, the capnography is the plot of the concentration of CO2 in the air exhaled by a patient. Two anesthsiologists annotate the begin and the end of every respiratory cycle. A time coordinate was attributed to each measured RR value. Obtained RR measurements were not performed on a regular basis. A reference value is available every 4 seconds on average.

Performance of the algorithms Fusion and Smart Fusion developed by Karlen & *al.* [31] are included in Capnobase, which makes it possible to perform a comparative study between these algorithms and our methods. In what follows, We always use the following notation to designate the methods proposed in this work:

- Consensus Spectrum from PPG (CS-PPG) (section 4.3)
- FFT combined to Median filter (FFT-Med) (section 4.4.3.1)
- Consensus Spectrum combined to Median filter (CS-Med) (section 4.4.3.2)
- Spectral Fusion combined to Median filter (SF-Med) (section 4.4.3.3)
- Spectral Fusion combined to Consensus Spectrum (SF-CS) (section 4.4.3.3)

4.5.1.1 Comparison with reference gold standard

A synchronization step is needed to have the same number of RR references and RR estimates and also to match the RR reference and RR estimate time scales. Depending

on the file, the number of estimations can be greater or less than the number of RR references:

- If the number of RR estimations is less than the number of references, the final estimate is taken as the median of the reference RR values corresponding to the time interval where RR is estimated from PPG:

$$RR_{ref} = median(RR_{ref}(t)) \text{ with } t_1 \leq t \leq t_2$$

  where t_1 and t_2 are the instant of beginning and end of the corresponding interval.

- Otherwise, a linear interpolation of the RR references is carried out.

4.5.1.2 Parameter choice

The methods for RR estimation presented in this chapter use fewer parameters than methods presented in the literature. Some preliminary tests are thus required to fix the length of the FFTs, the number of FFTs needed for Consensus Spectrum and the size of the overlap between FFT windows. Note that the choice of the overlapping rate depends on how often clinicians wish to obtain an RR estimation.

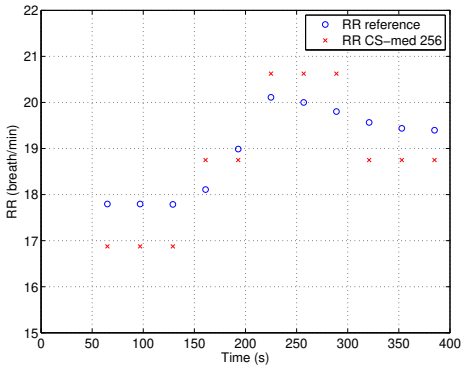
Window length test The problem with the Fourier Transform is how to find a compromise between the frequency measurement accuracy, on the one hand, and the processing time for real time application, on the other hand. Four window sizes without zero padding are tested { $32s$, $64s$, $128s$, $256s$}. These window sizes are appropriate to the physiological characters of the respiratory activity and meet the clinician's requirements.

Figure 4.7 gives an example of how the CS-Med algorithm evolves as the length of FFT window changes. The short windows $32s$ and $64s$ enable a real-time monitoring of the respiratory frequency, even when small changes occur. However, for the windows with lengths $128s$ and $256s$, RR estimates are more accurate.

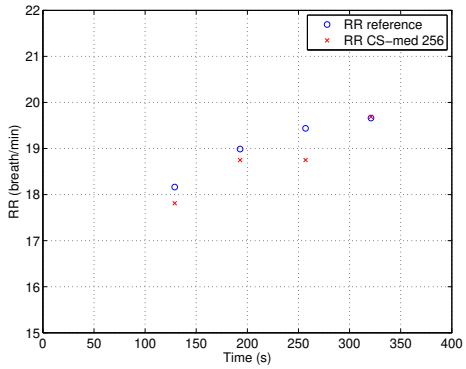
The methods were assessed by using the root mean square error (RMSE) defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(RR_{ref}(k) - RR_{Gen}(k))^2}$$

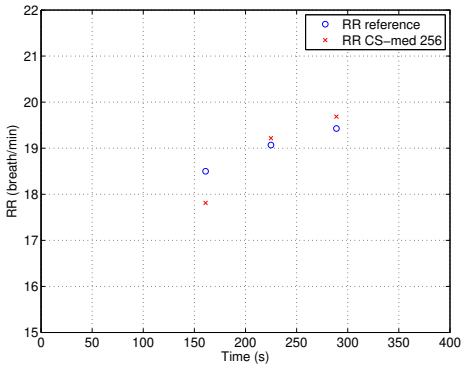
where n is the number of RR estimations and, for the kth estimation, $RR_{ref}(k)$ is the RR reference and $RR_{Gen}(k)$ is the final RR estimated by a given algorithm.



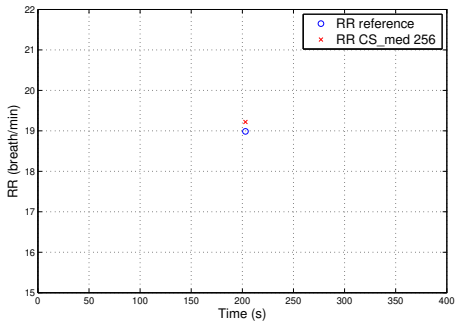
(a) RR estimated with CS-Med method using a window of $32s$



(b) RR estimated with CS-Med method using a window of $64s$



(c) RR estimated with CS-Med method using a window of $128s$



(d) RR estimated with CS-Med method using a window of $256s$

Figure 4.7 – Example of RR estimated with CS-Med method using different window sizes compared to the RR values from reference.

RMSE are calculated for each file from Capnobase for different window sizes $\{32s, 64s, 128s, 256s\}$ in Figure 4.8. The central mark of each box is the median, the edges of the box are the lower quartile q_1 (25th percentile) and the upper quartile (75th percentile) q_3. The whiskers extend to the most extreme data points defined by 1.5 times the interquartile range $(q_3 - q_1)$. Values greater than $q_3 + 1.5(q_3 - q_1)$ or less than $q_1 - 1.5(q_3 - q_1)$ are considered as outliers and represented by red crosses.

RMSE median values obtained for all methods are between 0.11 and $2.38 breath/min$. The larger the window size, the smaller the RMSE. However, the upper quartile and the number of outliers differ from one method to another. In fact, high values of RMSE are also observed for large windows of $256s$. The best performance compromise found in this context is for the CS-Med algorithm with a window of $64s$, where the RMSE median is $0.3 breath/min$ and the upper quartile value of the RMSE is about $0.7 breath/min$.

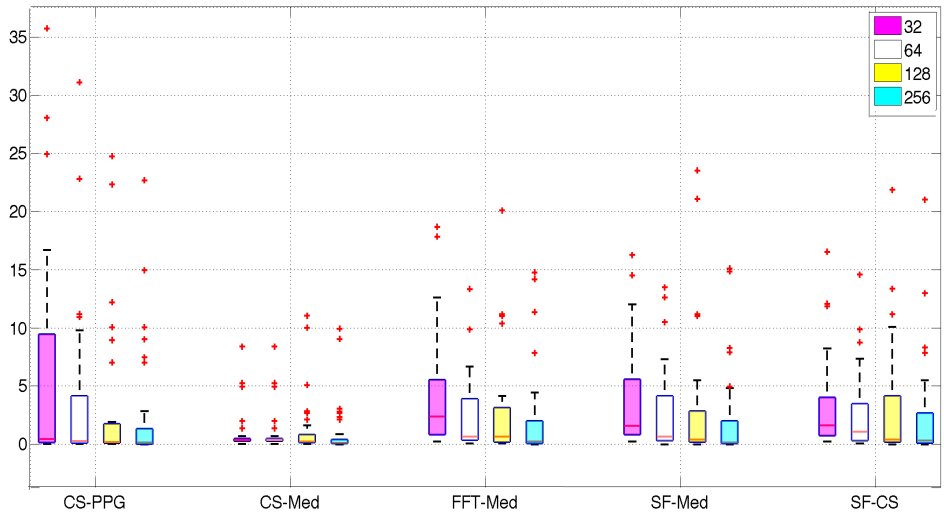


Figure 4.8 – Results from RMSE analysis of capnobase data to select the best FFT window size $\{32s, 64s, 128s, 256s\}$.

Number of FFTs for Consensus Spectrum The number of FFTs needed to optimize error rates for the CS-Med and CS-PPG methods is studied in this paragraph (4.9). According to the results above, the window length is fixed to $64s$ with an overlap of $16s$. The higher the number of FFTs, the better the RMSE median. There is no significant difference between the values of the RMSE median obtained with either CS-Med or CS-PPG (between 0.27 and $0.53 breath/min$ for CS-PPG method and between 0.35 and $0.57 breath/min$ for CS-Med method). However, like in the previous figure 4.8, the CS-Med method is more stable: upper quartile values for the CS-Med method are less than those obtained with the CS-PPG method. The best performance is observed by selecting 8 FFTs but, from a computational point of view, it is a costly choice, especially because it induces no significant performance gain. For this reason, in the following, we will fix the number of FFTs to 4.

4.5.1.3 Performance by age

Anatomical and physiological changes with age affect the respiratory system of subjects. In this section, we aim to study the behavior of the respiratory rate estimation algorithms with respect to to the patients' age.

Participants of the Capnobase project are from different ages including pediatrics and adults. PPG records are classified into 3 groups according to age. The RMSE of the proposed algorithms within each age range is presented in Table 4.2 and Figure 4.10.

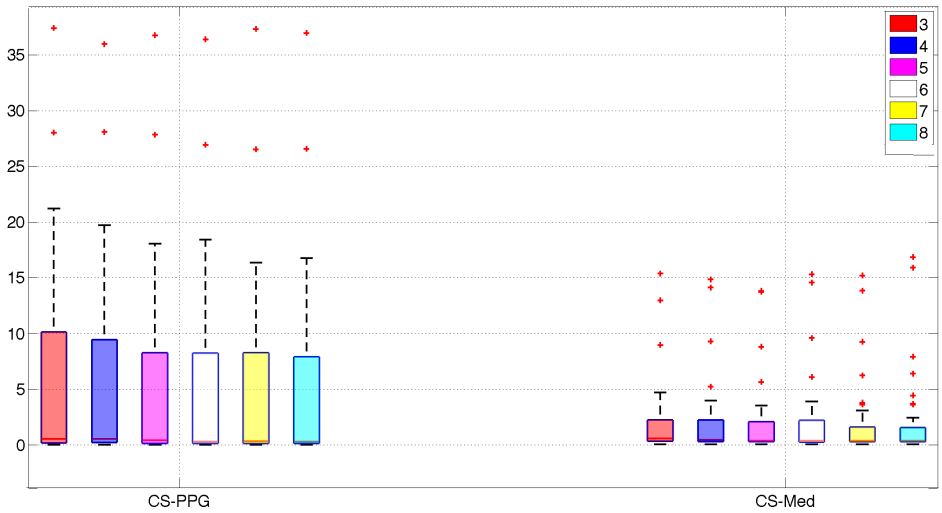


Figure 4.9 – Results from RMSE analysis of capnobase data to select the best number of FFTs needed for Consensus Spectrum calculation in CS-PPG and CS-Med

Parameters are fixed to $64s$ and 4 ffts for both CS-Med and CS-PPG. The best performance measurements are observed for the CS-Med algorithm for all age classes. Comparable RMSE results are achieved by the CS-Med method for all age classes. Another observation is that the CS-PPG algorithm outperforms the CS-Med algorithm for ages between 12 and 29 years. However, the behavior of the CS-PPG algorithm is unstable for the other age classes. For the other methods, high RMSE values are more observed for young subjects ($<$ 12 years). This difference in performance can be explained by the respiratory rate range and the quality of the respiratory modulation signals, which can differ according to the patients' age.

Table 4.2 – Performance of the RR estimation algorithms, according to the age ranges, for windows of $64s$. In each age group, N is the number of records. For each method, we give the median and inter-quartile range ($25th$- $75th$) percentile of the RMSE expressed in breath/min.

Age range	N	CS-PPG	CS-Med	FFT-Med	SF-Med	SF-CS
$<$ 12	16	0.31 (0.17-3.56)	0.37 (0.25-0.62)	0.95 (0.47-6.2)	0.75 (0.32-6.77)	1.11 (0.39-6.51)
12-29	11	0.21 (0.11-0.32)	0.31 (0.21-0.36)	0.42 (0.29-3.75)	1.03 (0.29-3.36)	0.43 (0.25-2.49)
$\geqslant$ 30	14	0.88 (0.11-9.14)	0.33 (0.31-0.5)	0.57 (0.37-3.12)	0.56 (0.32-3.68)	1.27 (0.31-2.71)

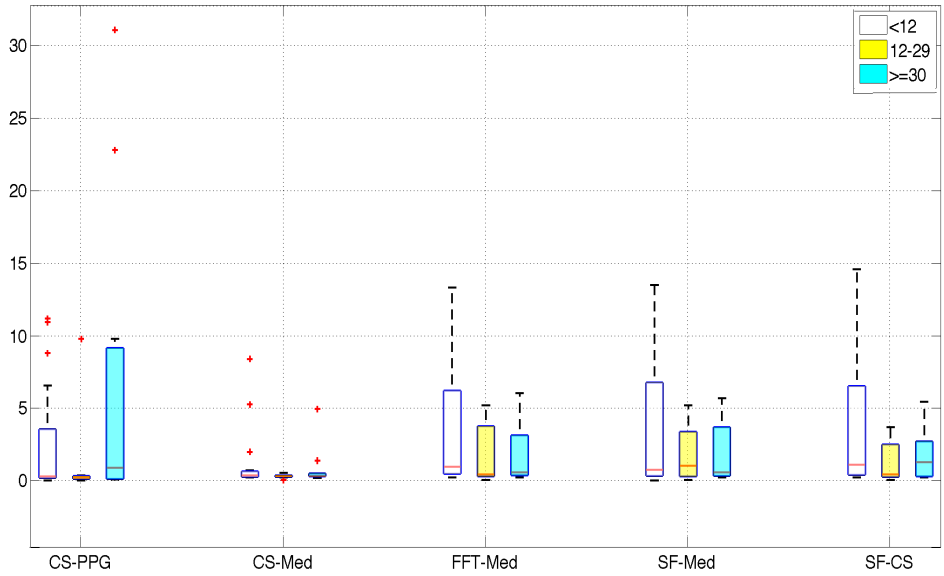


Figure 4.10 – Results from RMSE analysis of capnobase data using 64s window. Comparison of all methods according to age range

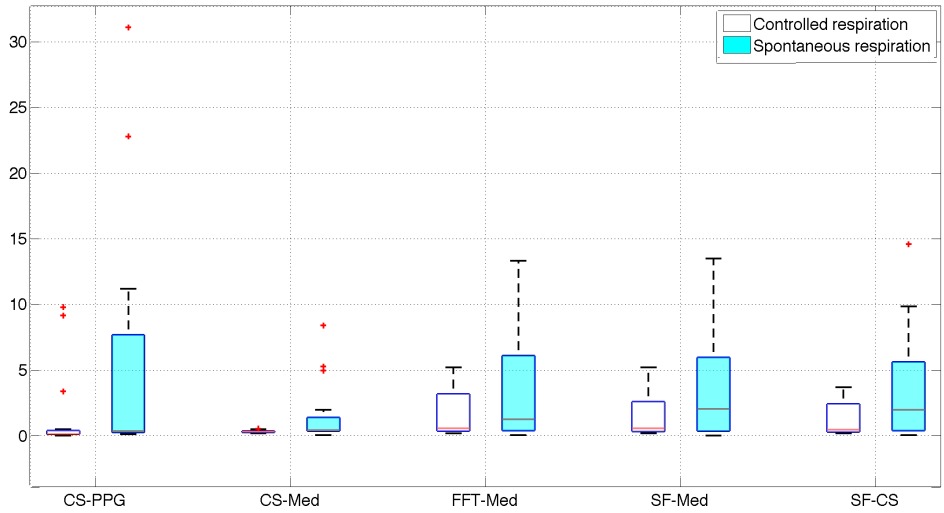


Figure 4.11 – Results from RMSE analysis of capnobase data using 64s window. Comparison of all methods according to ventilation mode

4.5.1.4 Performance by ventilation mode

Under different ventilatory conditions, the pulmonary volume and the intra-thoracic pressure are modified. So, respiratory induced signals change behavior. In this section, the proposed methods are analyzed according to the available ventilation modes.

Capnobase data contains 21 subjects under controlled ventilation and 20 subjects with spontaneous respiration.

Figure 4.11 and table 4.3 present the RMSE analysis for the two respiration modes. A substantial difference is observed for the two respiration modes. The RMSE values are lower for patients under controlled ventilation than for patients with spontaneous respiration. This can be explained by the fact that respiration activity is more stable and with less variations under controlled ventilation.

It is also worth noting that the CS-PPG method outperforms slightly the CS-Med method for patients under controlled respiration. However, the CS-PPG method yields strong outliers for patients with spontaneous respiration, which questioned the reliability of the method.

Table 4.3 – Performance of the RR estimation algorithms with respect to the ventilation mode for window of $64s$. N is the number of records. For each method, the RMSE results are given in terms of median and inter-quatile range ($25th$- $75th$) percentile expressed in breath/min.

Ventilation mode	N	CS-PPG	CS-Med	FFT-Med	SF-Med	SF-CS
Controlled respiration	21	0.11 (0.09-0.37)	0.31 (0.23-0.35)	0.55 (0.36-3.17)	0.55 (0.30-2.58)	0.46 (0.28-2.40)
Spontaneous respiration	20	0.36 (0.24-7.66)	0.42 (0.34-1.37)	1.23 (0.39-6.10)	2.04 (0.34-5.95)	1.96 (0.39-5.60)

4.5.1.5 Global comparison

In addition to the RMSE analysis, several statistical tools are presented to assess the methods.

- **Bland and Altman:** It is a data plotting method that compares a technique of parameter measuring to the reference technique. It involves computing a bias by calculating the mean differences (MeanDiff) obtained between the two techniques and fixing a confidence interval with limits $MeanDiff - 1.96 \times SD$ and $MeanDiff + 1.96 \times SD$ where SD is the standard deviation of the differences. The lower are the bias and the SD, the closer is the measurement method to the reference method [56].

- **Scatter plot and correlation** A scatterplot is a graphical tool for showing the relationship between two variables. In our context where we want to compare an RR estimate and an RR reference, the abscissa of each point is the RR reference

and the ordinate is the RR estimate. The best fit line is the identity line $x = y$. The closer the points to the best fit line, the closer the correlation to 1 or -1.

Table 4.4 – Performance of RR estimation methods. N is number of RR estimates. RMSE is calculated for each record and expressed in breath/min as median and inter-quartile range ($25th$- $75th$) percentile. We display the mean difference between the RR references and the estimates, with confidence intervals. We also give the correlation between the RR references and estimates.

Method	N	RMSE (br/min)	Mean difference (br/min)	Correlation
CS-PPG	164	0.26 (0.11-4.15)	-2.25 (-15.89 - 11.37)	0.75
CS-Med	164	0.34 (0.27-0.47)	-0.24 (-3.77 - 3.28)	0.97
FFT-Med	287	0.66 (0.37-3.91)	-0.45 (-7.81 - 6.89)	0.87
SF-Med	287	0.66 (0.31-4.14)	-0.38 (-8.82 - 8.05)	0.83
SF-CS	287	1.06 (0.30-3.46)	-0.38 (-8.05 - 7.27)	0.86

The performance measurements of the proposed methods with respect of the forgoing criteria are detailed in table 4.4. There are significant differences between the CS-MED method and the other proposed methods. Although the CS-PPG method has the lowest median RMSE, the corresponding inter-quartile distance is the highest. Considering this fact, the CS-Med method has the best median RMSE with the smallest distance.

For the mean difference criterion evaluated by the Bland-Altman plot, the CS-Med method has the lowest bias and also a narrower confidence interval comparing to the other methods. This result is illustrated in Figure 4.12. The confidence interval is between -3.77 and 3.28 breath/min, which is an acceptable RR error range for real applications. Indeed, only four outliers are observed outside the confidence interval.

The same behavior is observed by using the correlation. CS-Med has the highest value of correlation. This fact is also confirmed by Figure 4.13. In fact, the scatter plot shows that RR values estimated with the CS-Med method are the closest to the best fit line, with fewer outliers than the other methods.

4.5.1.6 Comparison to Karlen methods

In this section, CS-Med is compared to methods from literature whose performance results on the Capnobase data are available. In most literature reviews, the Smart

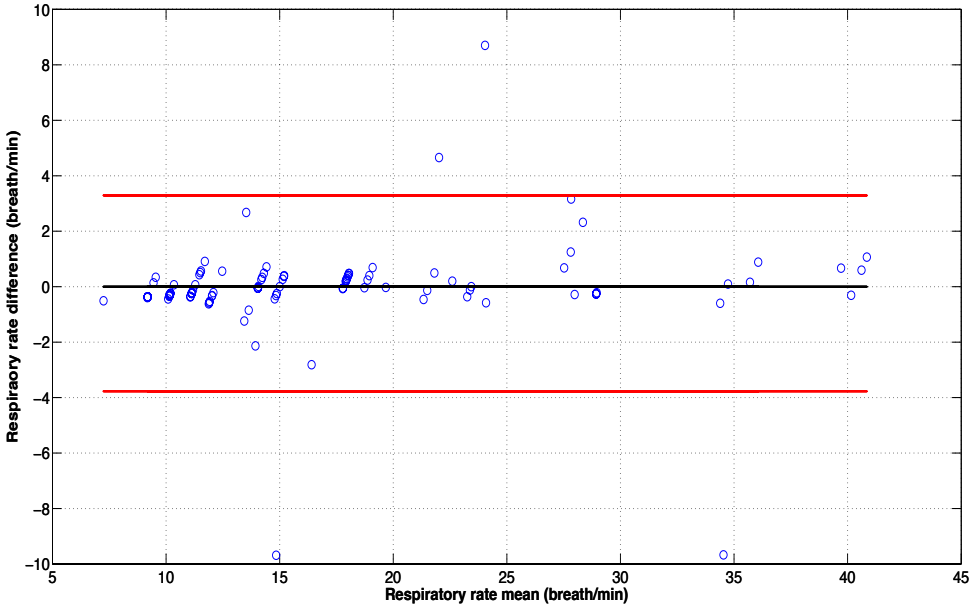


Figure 4.12 – Bland-Altman plot of the comparison between RR reference and RR estimate from CS-Med. Mean Difference line is in black at -0.24, Confidence interval limits are between -3.77 and 3.28

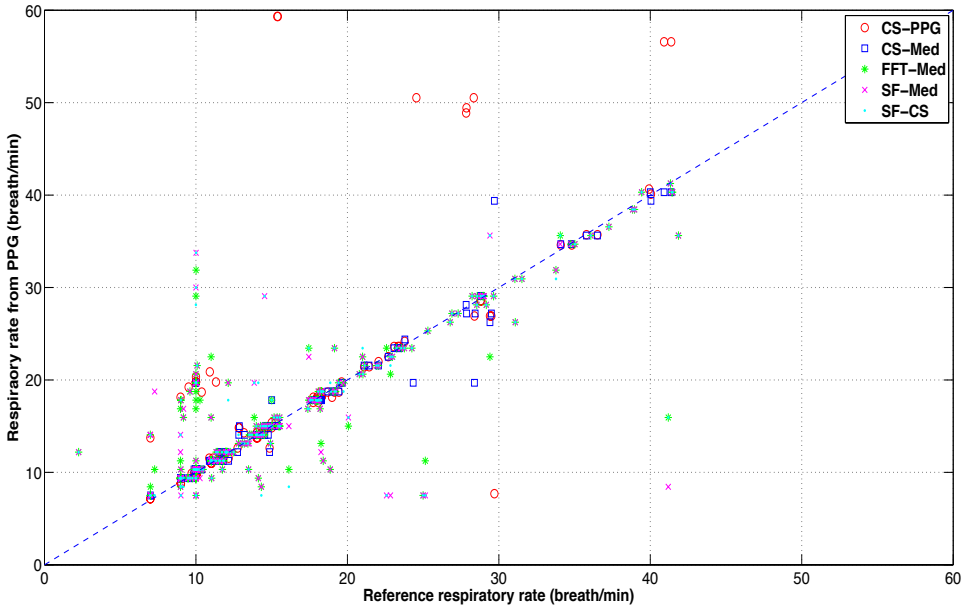


Figure 4.13 – Scatter plot comparing RR references with RR estimated from PPG by using CS-PPG, CS-Med, FFT-Med, SF-Med and SF-CS methods. The best fit line is represented by a blue broken line.

Fusion method proposed by Karlen [31] outperforms most proposed algorithms. Details of the Karlen method were already described in Section 4.2.2.3. A brief reminder for the reader: the Fusion method involves estimating RR_{AM}, RR_{FM} and RR_{BW} from the

induced respiration signals; then, the final RR is calculated by computing the mean of the three RR values. An additional preprocessing module rejects windows where artifacts are detected. The Smart Fusion eliminates aberrant RR values by rejecting RR estimates when the standard deviation of RR_{AM}, RR_{FM} and RR_{BW} is more than 4 breath/min.

The Karlen methods give an RR estimate every $3s$. The FFT window length is set to 64s. In order to compare our proposed method CS-Med to the Fusion and Smart Fusion from Karlen study, the overlap of the CS-Med method is adjusted to have a constant occurrence of RR estimates. For the CS-Med method, no preprocessing for selecting windows is used. All PPG signals are analyzed. Results are detailed in Table 4.5. Two comparison approaches are used in this study:

- The instantaneous RR, which is estimated every $3s$, is compared to the instantaneous reference RR.
- The global RR, which is the median of the instantaneous RR values per file, is compared to the global reference RR.

Table 4.5 – Performance of RR estimation methods for instantaneous and global RR. N is the number of RR estimations over the whole dataset (% of accepted windows). The RMSE is calculated for each record and expressed in breath/min as median and inter-quartile range ($25th$- $75th$) percentile. We display the mean difference between the references and and the estimates, along with the confidence intervals and the correlations between the RR references and estimates. The percentage of accepted windows is calculated for each file: The mean and ± the standard deviation of these percentages is given by NF

Instantaneous RR				
Method	N	RMSE (br/min)	Mean difference (br/min)	Correlation
CS-Med	5542 (100%)	1.75 (0.59-3.63)	-0.59 (-9.48, 8.29)	0.81
Fusion	5231 (94,3%)	3.37 (1.91-4.90)	-0.6032 (-10.43 , 9.22)	0.75
Smart Fusion	3083 (55,63%)	1.16 (0.57-3.17)	0.59 (-5.10, 6.30)	0.93
Global RR				
Method	NF	RMSE (br/min)	Mean difference (br/min)	Correlation
CS-Med	100%	0.31 (0.17-0.31)	-0.41 (-3.74 , 2.90)	0.97
Fusion	94% (± 1%)	0.62 (0.18-2.65)	-0.50 (-7.30 , 6.29)	0.89
Smart Fusion	55,63% (± 3%)	0.31 (0.11-1.92)	0.44 (-4.27 , 5.16)	0.95

The RMSE analysis of instantaneous measures show that CS-Med and Smart Fusion yield comparable results. In contrast, a more significant difference is observed

between CS-Med and Smart Fusion (figure 4.14). In fact, the confidence interval is more restricted when using the smart fusion method (figure 4.15). The smart fusion method has also the highest correlation value.

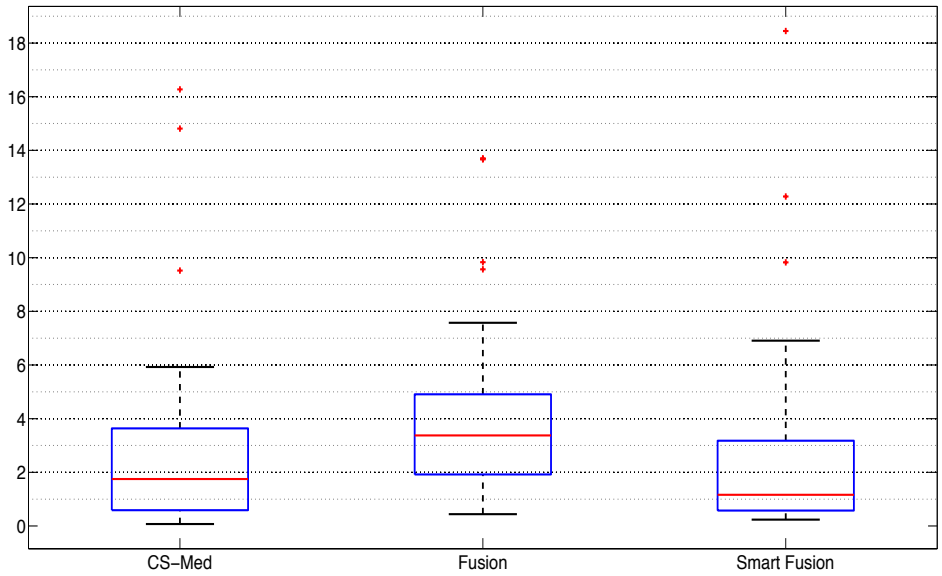
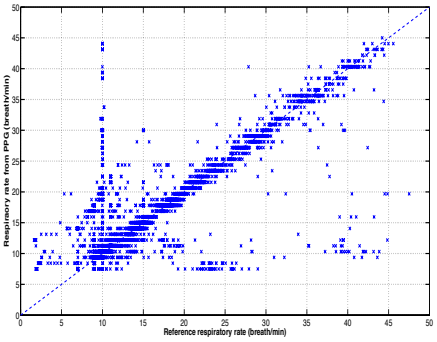


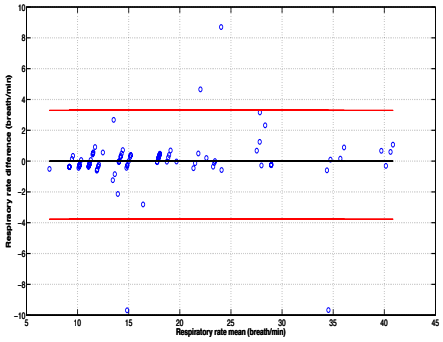
Figure 4.14 – Results from RMSE analysis when comparing instantaneous RR references to estimates from Karlen methods: Fusion and Smart Fusion and our proposed method CS-Med

However, when considering global measures, the CS-Med method outperforms all methods according with respect to each comparison tool. The CS-Med has not only the best error rate comparing to Karlen methods but also it has the best estimation rate. In other words, all the PPG windows are considered when using CS-Med, independently of the presence or not of artifacts. This is a significant advantage compared to smart fusion which rejects 45% of the PPG signals.

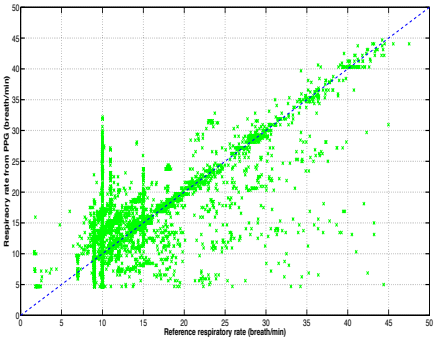
In the scatter plots of Figure 4.15, it is noteworthy how Smart Fusion discards aberrant RR estimates comparing to Fusion. But, even if Smart Fusion is a good tool for selecting best RR estimates, it cannot ameliorate measure accuracy. This fact can explain the difference between instantaneous and global performance for the CS-Med method and the Smart Fusion method. In addition, median filter applied to the reference permits to compensate measure errors and also the lag between the RR reference measurement time from the capnography signal and the RR estimation time from the PPG signal. In fact, synchronization between PPG signal and the Capnography signal on which RR reference is measured is not always accurate and some delays are often introduced. On the other hand, median filter applied to RR



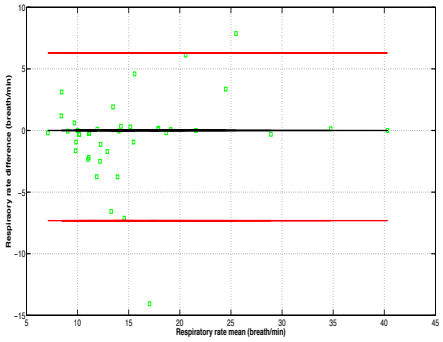
(a) Scatter plot comparing instantaneous RR references with RR estimates from CS-PPG



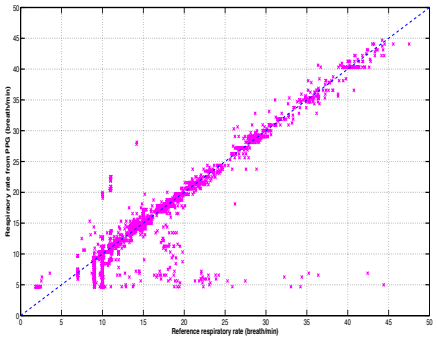
(b) Bland-Altman plot of the global RR reference and global RR estimates from CS-Med.



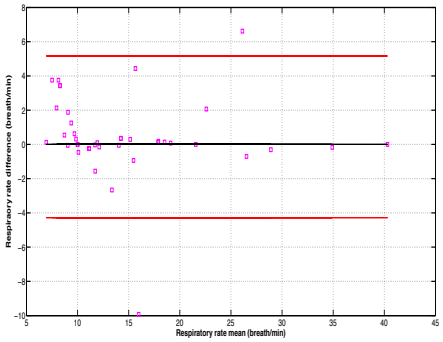
(c) Scatter plot comparing instantaneous reference RR with RR estimates from Fusion



(d) Bland-Altman plot of the global RR reference and global RR estimates from Fusion.



(e) Scatter plot comparing instantaneous RR references with RR estimates from Smart Fusion



(f) Bland-Altman plot of the global RR references and global RR estimates from Smart Fusion.

Figure 4.15 – Performance comparison of the CS-Med method, Fusion method and Smart Fusion method with respect to instantaneous and global measures.

estimated by the CS-Med method eliminates aberrant RR values; which explains the performance improvement for the CS-Med method when considering a whole 8 minute record.

4.5.2 Results on Reastoc

In this section, we evaluate in a clinical setting the performance of our proposed methods for estimating RR and that of the Smart Fusion method proposed by Karlen & *.al.* Tests are carried out on a portion of the Reastoc database (section 2.5.3). Reference RR values on Reastoc were measured manually by medical staff during the first two hours of the recording for each patient. Only one reference value is available for each patient. Unlike the Capnobase dataset whose instantanous RR references are available, instantaneous comparison between RR references and RR estimates from Reastoc data is not possible. In addition, the reference RR value for the Reastoc data does not reflect possible variation of the patient's respiratory activity. Although this RR variation can be enhanced by estimating RR through the proposed algorithms, we have not sufficient reference RR values to be able to assess this feature of the proposed methods. Therefore, only patients under controlled ventilation are analyzed in this study. In fact, mechanically ventilated patients have generally stable breathing rate adjusted on the ventilator by the clinician. In some cases, patients under mechanical ventilation can still have unstable respiratory activity. So, we have selected only patients whose RR reference measured by the clinician is equal to the RR fixed on the ventilator system. In this way, patient respiratory activity is more likely to be stable and then we can guarantee an accurate reference. For the same reasons, patients suffering from arrhythmia are also excluded from this study. Consequently, 26 patients are selected. The length of each record is variable from 5 to 25 minutes.

The number of patients may be considered as too limited for valuable clinical assessment of RR estimation. But, our goal is to assess, as much as possible with respect to Reastoc constraints, the behavior of the methods on different data and contexts. Of course, for clinical validity of such methods, it will be necessary to complete the assessment through experiments conducted according to a well defined clinical protocol involving more patients' contexts.

As for Capnobase data, the FFT window length is fixed to $64s$. The number of FFTs needed for consensus spectrum for the CS-Med and CS-PPG methods is fixed to 4. The median of all RR values estimated per file is then calculated for each record. Performance results are detailed in Table 4.6.

The behavior of the proposed algorithms for the Reastoc database is different from Capnobase data. The CS-PPG method outperforms all other methods according to the RMSE analysis (Figure 4.16). The same results are also observed according to the

Table 4.6 – Performance of RR estimation methods for estimating RR. N is the number of files where RR values were estimated. The RMSE is calculated for each record and expressed in breath/min as median and inter-quartile range ($25th$- $75th$) percentile. We display the mean difference between the references and and the estimates, along with the confidence intervals and the correlations between the RR references and estimates.

Method	N	RMSE (br/min)	Mean difference (br/min)	Correlation
CS-PPG	26	0.35 (0.18-2.11)	1.03 (-4.68 ; 6.75)	0.84
CS-Med	26	0.34 (0.12-5.62)	2.44 (-6.21 ; 11.09)	0.64
FFT-Med	26	0.4 (0.15-5.15)	2.36 (-6.12 ; 10.86)	0.62
SF-Med	26	0.34 (0.12-4.68)	2.29 (-5.67 ; 10.26)	0.67
SF-CS	26	0.43 (0.25-2.96)	2.11 (-5.07 ; 9.30)	0.71
Smart Fusion Karlen	23	0.5 (0.17-6.16)	3.02 (-6.29 ; 12.34)	0.61

Bland Altman analysis (table 4.6). In fact, the CS-PPG method yields the least mean difference and the most restrained confidence interval.

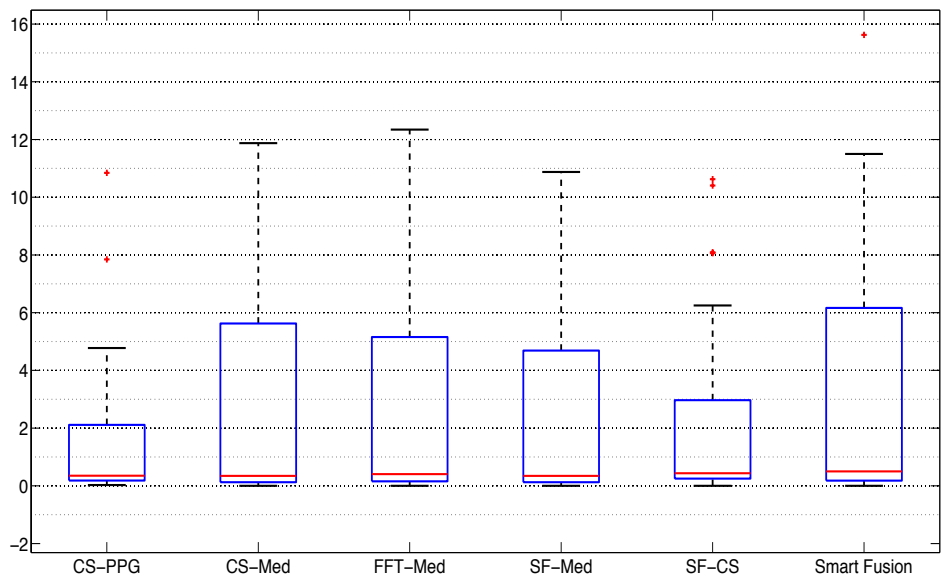


Figure 4.16 – Results of the RMSE analysis when comparing RR references to RR estimates for the methods: CS-PPG, CS-Med, FFT-Med, SF-Med, SF-CS and Smart Fusion

Scatter plot of all the methods is given by figure 4.17. We note that there are some cases where most methods underestimate RR. This fact can be the result of low frequency non respiratory fluctuations due to other physiological processes. The true

RR frequency is then hidden by these spurious frequencies. This phenomenon is known as Mayer waves which are arterial blood oscillations that occur at a frequency lower than respiration frequency (about $0.1Hz$) and might result from sympathetic nervous activity.

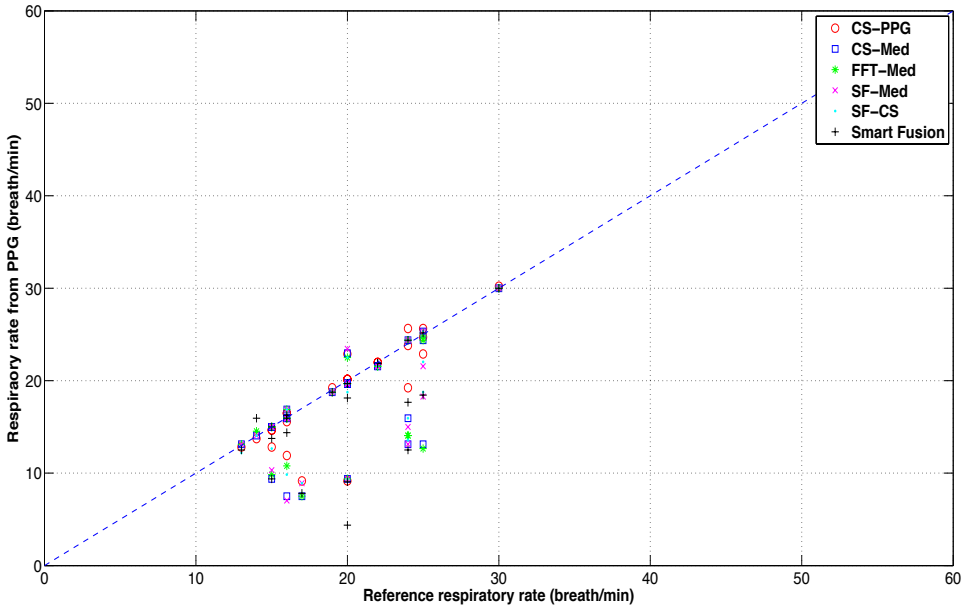


Figure 4.17 – Scatter plot comparing RR references with RR estimates. The best fit line is represented by a blue broken line

Another noticeable fact is that performance measurements of the Smart Fusion are poor comparing to the other methods. In addition, 3 records have been eliminated by the Smart Fusion. In fact, the criterion of RR selection of the Smart Fusion has discarded all possible RR estimations for these records.

To further analyze RR estimation on the Reastoc dataset, we propose to observe the behavior of the proposed algorithms with respect to age range. In figure 4.18, we notice that there is a considerable difference in performance for all methods between patients over 60 years and patients under 60 years. The respiratory frequency is not discernible for patients over 60 years. This results from the metabolism change for elderly persons, which affects the respiratory activity.

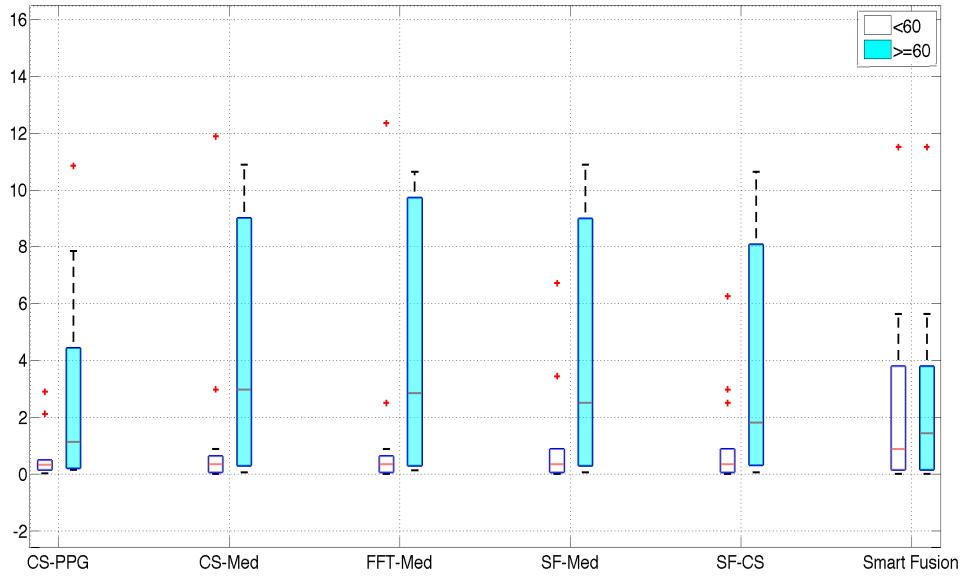


Figure 4.18 – RMSE analysis according to age range

4.6 Discussion

4.6.1 Comparison between the proposed algorithms

The proposed algorithms were tested in two different contexts: Capnobase data and Reastoc data. By this approach, we aimed to investigate if there is a generic method for estimating RR regardless of the patient's state. We proposed to compare several types of methods. The first method was the CS-PPG method based on the spectral analysis of the whole raw PPG signal. Then, we were interested in the analysis of respiratory induced signal AM, BW and FM and how to estimate the final RR which will be communicated to the clinician. The first approach, given by the CS-Med and FFT-Med methods, was to fuse RR values estimated from the modulation signals by median filter. The second approach, given by the SF-Med and SF-CS methods, was to fuse spectrum of modulation signals AM, BW and FM then estimating RR from the resulted spectrum.

A summary is given in figure4.19. Spectral fusion methods have limited performances on Capnobase comparing to the CS-Med. In fact, fusing power spectra has generated secondary frequency peaks that have hidden the respiratory frequency. However, it is worth to note that the SF-CS method has interesting performance for the

Reastoc data but instable when observing the confidence range of the Bland Altman performance.

For methods based on combining RR_{AM}, RR_{FM} and RR_{BW}, the CS-Med method surpasses the FFT-Med method for both Capnobase and Reastoc. In fact, the spectral smoothing engendered by consensus spectrum eliminates fluctuations in the modulation signals. Compared to the CS-PPG, which estimates RR from raw PPG, the CS-Med outperforms the CS-PPG for the Capnobase data. On the contrary, for the Reastoc data, the CS-PPG method gives better results than the others. But, when selecting patients under 60 years from the Reastoc data, the gap between the performance of the two algorithms is less important. The performance difference between the two methods CS-Med and CS-PPG when changing database remains a critical point that limits the universality of the proposed algorithms. The most likely hypothesis for this performance difference could be the quality of the extracted respiratory signals. . In fact, some oximeters contain internal modules for filtering DC components in order to have a better representation of the PPG signal. It is recalled that the AM and BW modulations are closely related to the baseline of the PPG signal. So,these signal processing stages can corrupt mainly the AM and BW modulation signals. Further tests are necessary to construct a decision process to choose the right method depending on the PPG acquisition mode and on patient characteristics.

4.6.2 Age impact on algorithms performance

The strength of the respiratory induced signals are likely to be deteriorated by several factors as described in section 2.3.3. The most obvious finding in the presented results is the age impact on respiratory induced signals. The medical explanation of this phenomenon is that the respiratory system undergoes several physiological changes with age. In [57], authors present how aging impacts respiratory activities. Some factors like chest wall deformation and muscle atrophy of the lung and diaphragm, are the principal causes of lung compliance decrease. By lung compliance we mean the lung ability to change its volume in response to a change in pressure. According to [57], this decline of lung compliance is accelerated after 60 years old. As AM and BW signals are mainly related to the intrathoracic pressure and FM signals are related to the respiratory sinus arrhythmia, the quality of those signals is thus deteriorated for aged patients whose intrathoracic pressure becomes low. In this case, the RR estimation accuracy from PPG cannot always be guaranteed for patients over age 60. Indeed, within this age class in the Reastoc data and although PPG signals are free

from artifacts, all algorithms underestimate RR in most cases. However, the patients of the Reastoc data are under mechanical ventilation. Therefore, their intrathoracic pressure tends to be high, which should favor good RR estimation from the modulation signals. We have two contradictory facts. Unfortunately, the number of patients in this age class is not enough to explain the increase of the number of errors that we have noticed for aged patients within the Reastoc dataset. A thorough understanding of how age impacts respiratory signals from PPG is necessary to have a robust RR estimator. Adding weights to the signal modulations according to their strength could maybe ameliorate the performance for this age class.

4.6.3 Comparison with others methods

The performance of the algorithms was compared to that yielded by some methods of the literature, principally the Smart Fusion method proposed by Karlen & *al.*, which are reference methods for Capnobase data. These methods are also generally used in many comparative studies. For Capnobase Data, our proposed algorithm CS-Med behaves well compared to the Fusion algorithm. When comparing to Smart Fusion, according to error statistics, Smart Fusion has the best performances at the expense of decreasing the estimation rate. In fact, only 55% of the windows are retained. Nevertheless, unlike Karlen's methods, for the CS-Med, no preprocessing for eliminating artifacts was carried out and 100% of the signal was analyzed. The difference between the accuracy of Smart Fusion and CS-Med is not very significant. So, the CS-Med remains a good compromise between robustness, accuracy and estimation rate. Besides, the Smart Fusion method failed to estimate acceptable RR values for the Reastoc data. Consequently, the criterion of standard deviation comparison introduced by the Smart Fusion method to eliminate RR aberrant values is not sufficient.

Much of the available literature presents a plethora of methods for estimating RR in different contexts: healthy subjects, ambulatory patients or patients in in general care floor. Thus, a quantitative comparison of all methods seems not possible for many reasons. First, performance of methods depends on the characteristics of the subjects in the used database. Second, replication of algorithms of the literature is not accurate to the original algorithms because of the lack of information given by authors about parameter adjustment. Third, RR estimation is presented in some studies as a module of a whole process aimed at monitoring physiological signals where PPG is coupled to other signals.

However, we can present a qualitative comparison between our proposed methods and some others. Compared to the algorithms listed in Table 4.1, the paramount advantage of the CS-Med and CS-PPG methods is that few parameters are needed regardless of the data characteristics. In fact, wavelet and autoregressive methods need a calibration step to fix order parameters. However, these order parameters could change when the algorithms are applied to another database. In addition, the proposed methods are relatively simple and take the physiological fluctuation of the PPG signal into account. Consequently, the computational requirements of the algorithms are not costly and are appropriate for implementation in real-time setting.

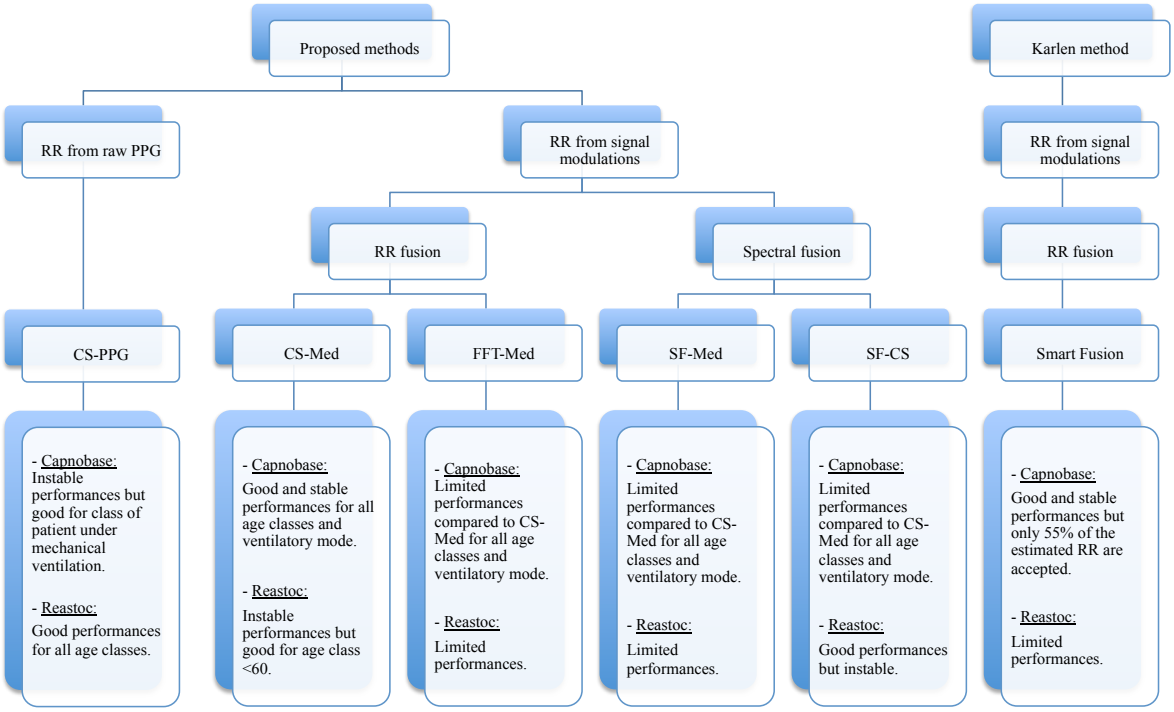


Figure 4.19 – Summary of the RR estimation methods analyzed in this chapter

In a recent study [58], Pimental and *.al* presented a new method for estimating RR by fusing multiple autoregressive models of different orders from the three signal modulations. The algorithm was tested on Capnobase and on BIDMC dataset. BIDMC dataset contains records of patient in intensive care unit. It has many characteristics similar to Reastoc: the same sampling rate (125 Hz) and the same median age (about 64 years). It is worth noting that the algorithm proposed by Pimental and *.al* has comparable behavior to the CS-Med and CS-PPG, especially when CS-Med and CS-PPG are applied to the Reastoc dataset. In fact, the Mean Absolute Error is higher when the algorithm is applied to BIDMC dataset. On the scatter plot, the algorithm presented in the cited study underestimates also RR for the BIDMC dataset. Authors have also presented performance of Smart Fusion on the BIDMC data and similar

ascertainment is determined: the Smart Fusion estimates RR only for 34% of the windows and the results are worst than the proposed algorithm.

From this study and the previous exposed results, it can be concluded that Capnobase is a good online dataset to benchmark studies about estimating RR but cannot be considered as an absolute reference because of its lack of representativity of different age classes and patient states. Therefore, algorithms tend to have many errors when tested on patients in intensive care. The proposed algorithms CS-Med and CS-PPG perform well compared to the other proposed algorithms and algorithms from the literature for a large proportion of the data. But, in some cases, the results obtained by CS-PPG are extremely limited, while CS-Med performs well and inversely. The CS-PPG method seems more adapted for patient under mechanical ventilation. However, the CS-Med method seems more adapted for young patients. A new strategy should be investigated to understand in-depth the performance variability then developing a system that takes the advantages of every method. In addition, there is a need for developing an index that defines the quality of signals and considers factors that could deteriorate performances.

4.7 Conclusion

In this chapter, we analyzed spectral methods for estimating RR from PPG signals. RR can be estimated from raw PPG signals with the CS-PPG method or also from the respiratory induced signals using the CS-Med method. The two methods perform well on both Capnobase and Reastoc, but with some differences. In fact, the results depend on several pulse oximeter factors, such as the sampling frequency and PPG processing modules, and also on patient conditions such as age and ventilatory mode.

Future work with large scale tests is necessary to validate the algorithms in order to generalize this technique of estimating RR from PPG signals in clinical practice. In the following chapter, we investigate the universality and robustness of the proposed algorithms in case of artifacts and we study the contributions of a quality index for determining cases where RR could be estimated with a large confidence scale.

CHAPTER 5 Optimization of respiratory rate monitoring from PPG

5.1 Introduction 81
5.2 Artifact detection impact on respiratory rate estimation 82
5.2.1 Motivation 82
5.2.2 Analysis methodology 83
5.2.3 Results on Capnobase 83
5.2.4 Results on Reastoc 85
5.2.5 Limits of artifact detection 86
5.3 SRQI impact on respiratory rate estimation 88
5.3.1 SRQI definition 88
5.3.2 Analysis methodology 89
5.3.3 Results on Capnobase 90
5.3.4 Results on Reastoc 93
5.4 Discussion 96
5.5 Conclusion 99

5.1 Introduction

There is a growing interest for noninvasive respiratory rate monitoring in diverse clinical applications. Estimating RR from PPG signal presents a promising alternative but still has limited use in real conditions because of the non robustness of the methods

proposed until today. In fact, as established in the precedent chapter, there is an important variability in performance of the algorithms depending on the databases and the subjects characteristics.

In this chapter, we present an analysis of factors that affect respiratory signals and we propose some tools to display to clinicians only relevant respiratory rate values. Results are compared to Karlen methods. Then, we present the advantages of the proposed method compared to some other studies in the literature.

5.2 Artifact detection impact on respiratory rate estimation

5.2.1 Motivation

Artifacts are a limiting issue for monitoring physiological signals. In the literature, artifact detection in PPG signals has been studied in various scopes and contexts. Previously, in chapter 3, we presented different algorithms from the literature for artifact detection and then we proposed the adaptive RDT algorithm for artifact detection, which is especially appropriate for long records. Studies in the literature about artifact detection usually assess their algorithm by comparing the detected contaminated pulses to a gold standard. In some other studies [59, 60], authors discuss the impact of artifact detection on heart rate estimation. Significant improvement of HR estimation is then noted. This result is expected because of the close interaction between the HR and the PPG pulsatile component.

However, researches about respiratory rate estimation do not explicitly discuss the impact of artifact detection for improving RR estimation accuracy. In fact, studies on respiratory rate estimation [31, 53, 58] often present a preprocessing module for eliminating contaminated segments. These modules for identifying segments with artifacts are always referenced as Signal Quality Indices (SQI). Heuristic thresholds are used to fix the SQI value above which the PPG segments are considered to be clean from artifacts. Segments with low SQI are rejected based on the assumption that they are too contaminated by artifacts to provide a good RR estimate. However, the interest of artifact rejection with respect to RR estimation improvement is not discussed in the literature. Specifically, the performance of the algorithms is presented without a focus on the actual relevance of the artifact detection module on the improvement of the RR estimation accuracy.

Based on these remarks and the previous experiments presented in chapter 3 and chapter 4, we present a novel contribution by analyzing the impact of artifact detections on RR estimations for both the Capnobase and Reastoc datasets.

5.2.2 Analysis methodology

In what follows, adaptive RDT for artifact detection is applied on PPG records for detecting pulses with artifacts. The number of Pulses with Artifacts per Window (nPAW) is introduced to characterize how much a given PPG segment is corrupted with artifacts. nPAW is simply the number of pulses that are detected as artifacts for a given window. As described in figure 5.1, when nPAW exceeds some threshold value $\mathcal{T}$, the RR estimation in the corresponding window is not considered as accurate enough and is thus rejected.

The proposed algorithms CS-Med and CS-PPG are used for RR estimation. These two algorithms gave the best performance in the experiments of chapter 3. Afterwards, the RMSE values are calculated for each record for different values of $\mathcal{T}$. We then discuss the effect of artifact detection on these RMSE values.

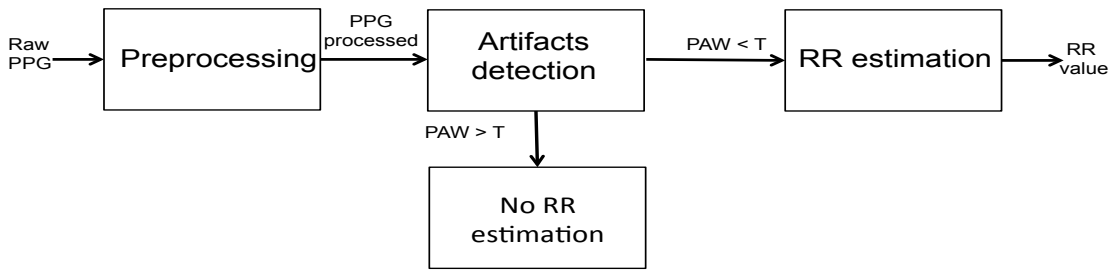


Figure 5.1 – Flowchart of the process for RR estimation coupled to artifact detection

5.2.3 Results on Capnobase

The RDT parameters are fixed as validated in Section 3.5.2.1: $\gamma = 10^{-3}$ and $\tau = 2$. The window length for FFT is equal to $64s$, the number of windows for consensus is fixed to 4 windows and a shift of $3s$ is applied for every FFT window. Overall, for a given RR estimate, a segment of length $73s$ is needed. This segment contains on average about 85 pulses. In fact, we remind that the number of PPG pulses per minute is equal to the heart rate which is on average about 70 beat/min for adults. However, the heart rate changes with age. So, an heuristic interval is fixed for nPAW which then takes a value from 1 to 60. For example, if $nPAW = 1$, all windows with at least one

artifactual pulse are rejected. Similarly, for $nPAW = 60$, all windows with more than 60 artifactual pulses are discarded.

The results are given in figure 5.2 and 5.3. For each value of nPAW, the median, the 25th percentile and the 75th percentile of the RMSE are given. The percentage of accepted RR values is defined as the number of accepted RR values divided by the number of RR values obtained when no artifact detection method is carried out and no segment is rejected.

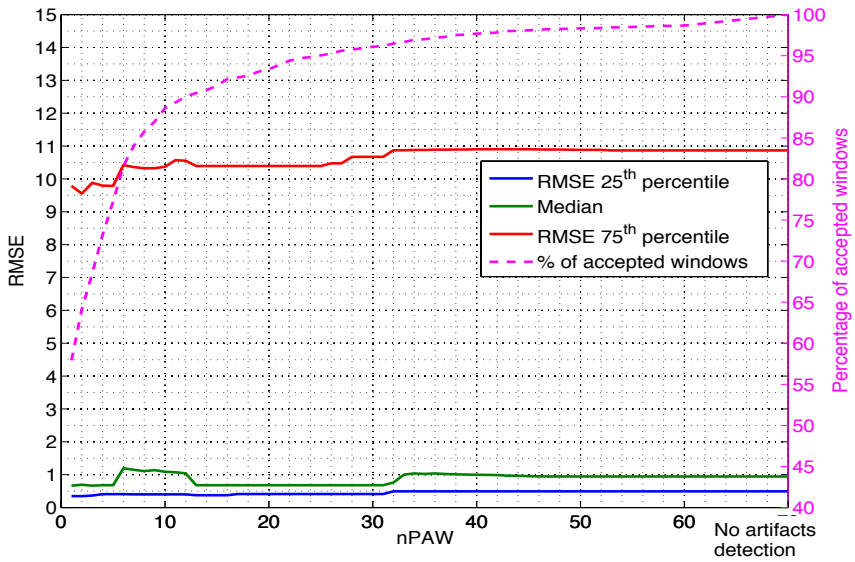


Figure 5.2 – RMSE of RR estimates by CS-PPG applied to the Capnobase data, versus the percentage of windows accepted as nPAW increases.

For the CS-PPG method (figure 5.2), a minor improvement of the RMSE median and 75th percentile is observed, compared to the RMSE when no artifact detection is used. The RMSE 75th percentile remains however too high, despite the elimination of almost all windows containing artifacts. Consequently, artifact detection has no significant impact on the performance of the CS-PPG method. Other factors have therefore caused the bad RR estimates.

For the CS-Med method (figure 5.3), a slight progress is observed when discarding all windows with $nPAW = 1$. This improvement is not substantial either. In particular, only 58% of the RR estimates are accepted. Little improvement of the RMSE median is observed. However, no significant amelioration is observed for the 75th percentile curve. The interquartile distance is still large.

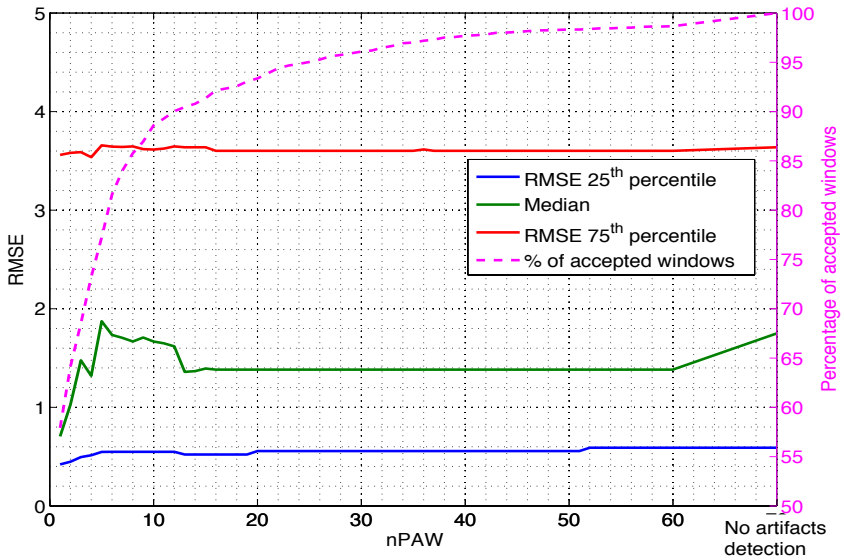


Figure 5.3 – RMSE of RR estimates by CS-Med applied to the Capnobase data, versus the percentage of windows accepted as nPAW increases

5.2.4 Results on Reastoc

The same strategy is adopted for the Reastoc dataset. The RDT parameters are fixed as in Section 3.5.2.2: $\gamma = 10^{-3}$ and $\tau = 1.5$. For the RR estimation algorithms, the window length is fixed to 64s ans the number of windows for consensus spectrum is fixed to 4.

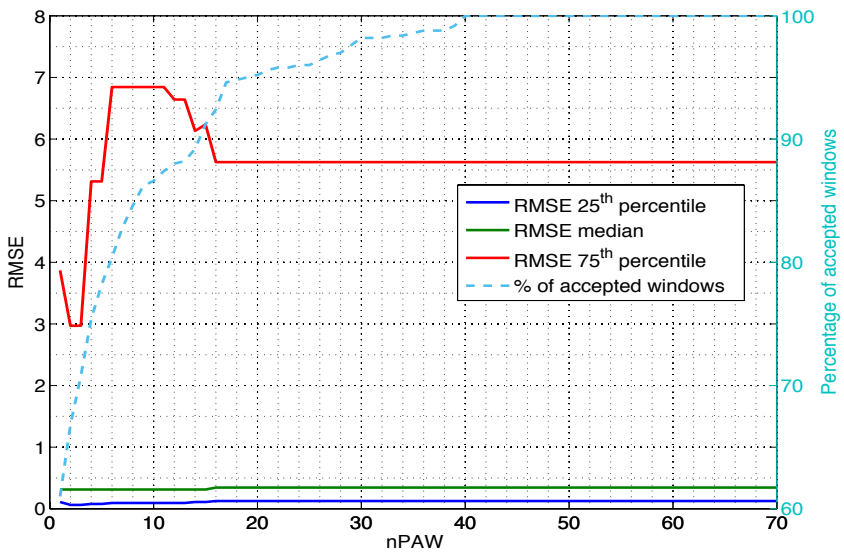


Figure 5.4 – RMSE of RR estimates by CS-PPG applied to the Reastoc data, versus the percentage of windows accepted as nPAW increases

For both the CS-PPG and CS-Med algorithms (figure 5.4 and figure 5.5), The median value of the RMSE is not impacted by the reject of the windows with artifacts. However, the difference between the 75th percentile and the median of the RMSE is reduced for the CS-PPG method for values of nPAW between 1 and 3. For nPAW=3, the percentage of accepted windows is about 71% with RMSE= 0.3 (0.14-1.17) breath/min. For the CS-Med method, the gap between RMSE values is reduced, but still high and fluctuating compared to the performance of the CS-PPG.

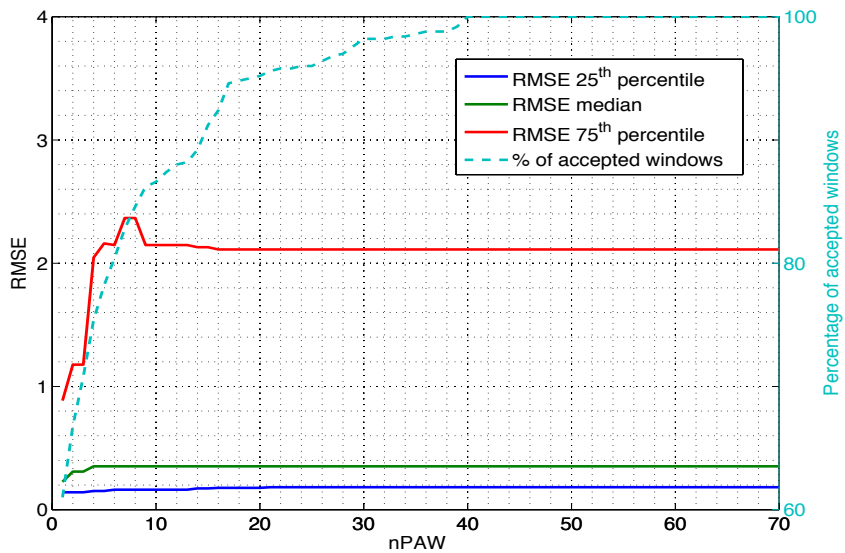


Figure 5.5 – RMSE of RR estimates by CS-Med applied to the Reastoc data ,versus the percentage of windows accepted as nPAW increases

In summary, artifact detection could ameliorate the RR estimation in some instances. However, this processing step is not sufficient to guarantee a robust RR estimation from PPG signals. In fact, RR estimation depends on several factors like age and pre-existing health conditions. Besides, the results presented in chapter 4 show that there is a close connection between RR estimation algorithm performance and age class in one hand, and also between RR estimation algorithm performance and ventilation mode in the other hand.

5.2.5 Limits of artifact detection

Previous researches conducted in the literature have focused on artifact detection for more accurate RR estimation. However, the analysis conducted in this chapter confirms that pulse with artifacts are not the only cause for bad estimation of RR. Figure 5.6 shows an example of a clean PPG signal with no artifact; but RR estimation for this

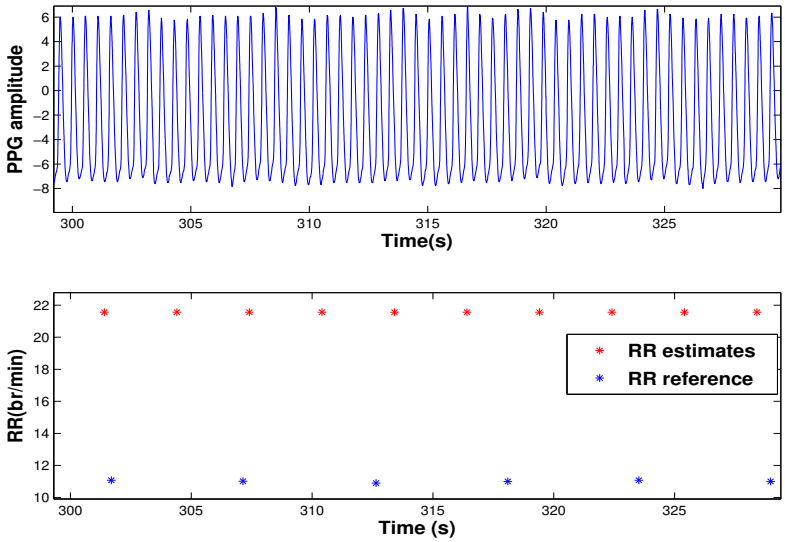


Figure 5.6 – Example of a PPG segment where there is no artifact but the CS-Med method fails to estimate good RR.

signal is most of the time erroneous. Indeed, artifact detection methods, especially those based on comparing pulse morphology, concern only temporal characteristics of the PPG signal. In fact, these methods handle pulsatile PPG waveform components with no concern to the signal modulations and the spectral component quality.

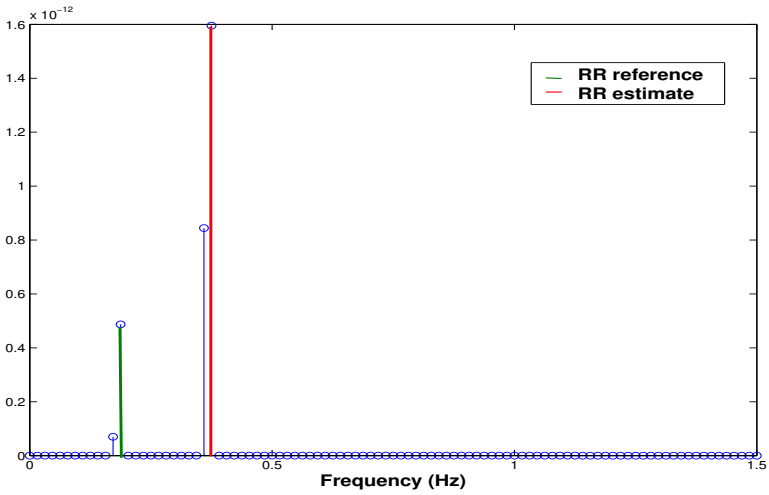


Figure 5.7 – Spectrum of the PPG signal presented in the latter figure 5.6. Spurious peaks in this spectrum have higher amplitude than the peak corresponding to the real RR (0.156 Hz). This causes erroneous estimation of RR.

A deeper analysis of the spectrum of some PPG segments where CS-Med and CS-PPG failed to give an accurate RR reveals that peaks corresponding to RR are not always dominant even when there are no artifacts (figure 5.7). In some cases, harmonic peaks appear in the spectra with greater amplitude than the real RR peak. In other

cases, unidentifiable spurious peaks are observed. These peaks have amplitudes in the same range as the RR peak and their corresponding frequencies are lower than the frequency of the reference RR. These peaks may correspond to Mayer waves as found in [61] and section 4.5.2. A new metric is then necessary to assess the spectral properties of the respiratory signals. The design of such metric is the topic of the next section.

5.3 SRQI impact on respiratory rate estimation

5.3.1 SRQI definition

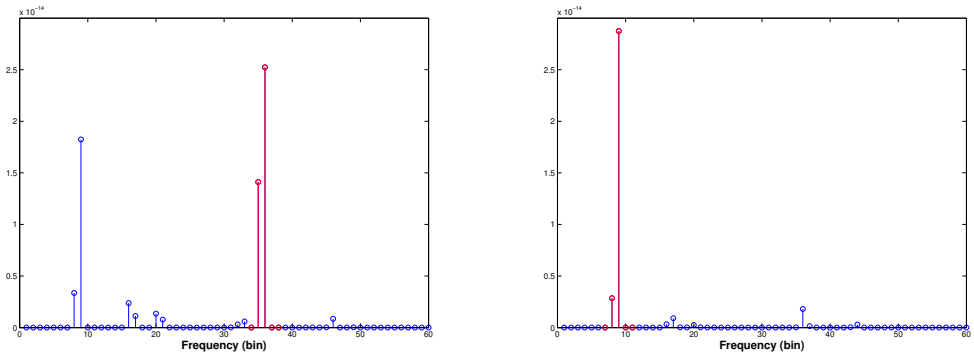
Artifact detection fails to improve RR estimation because it considers only the pulsatile aspect of the PPG signal and not its modulations. On the other hand, the spectral magnitude analysis of the respiratory signal is a promising tool to assess the pertinence of the RR estimate. In fact, an index is needed to evaluate if the maximum peak in the respiratory frequency band $(0.15 - 1Hz)$ is sufficiently dominant to be considered as a relevant respiratory frequency, in comparison with the other peaks in the same interval. In this respect, the Spectral Respiratory Quality Index (SRQI) is defined as the ratio of the power of the dominant peak by the total signal power in the respiratory frequencies band.

The SRQI is thus given by:

$$SRQI = \frac{\sum_{i=i_{max}-2}^{i_{max}+2} S_i}{\sum_{i=i_{RF_{min}}}^{i_{RF_{max}}} S_i}$$

where: i corresponds to the frequency bins index; S_i is the ith value of the power spectrum of a given signal s; $i_{RF_{min}}$ and $i_{RF_{max}}$ are the bins corresponding respectively to the limits of the respiratory frequency band $[0.15, 1]$ Hz; i_{max} is the bin index of the maximum of the spectrum. As a spectral peak can have many frequency components, the dominant peak is defined, in this context, by the maximum of S_i and its two adjacent bins from either side in the respiratory frequency band.

As defined, when only one dominant peak appears in the spectrum, the amplitudes of the other peaks become comparatively negligible with respect to the dominant peak amplitude. Thus, $\sum_{i=i_{max}-2}^{i_{max}+2} S_i$ tends to $\sum_{i=i_{RF_{min}}}^{i_{RF_{max}}} S_i$. So, SRQI tends to 1. In this case, this peak is more likely to correspond to the respiratory frequency. To the contrary,



(a) SRQI=0.7 ***(b)*** SRQI=0.99

Figure 5.8 – Examples of two different cases illustrating the variation of the SRQI according to the number of peaks present in the spectrum. In red, the bins corresponding to the dominant peak. The RR reference value in this example is the 9th bin, which corresponds to 0.14 Hz

the more spurious frequency peaks there are, the smaller the SRQI value is. In this case, the estimated RR is not plausible. Two examples are given in figure 5.8.

5.3.2 Analysis methodology

In this section, the SRQI metric is assessed when used with the proposed algorithms CS-PPG and CS-Med and also when it is applied to Karlen methods Fusion and Smart Fusion.

The CS-PPG method uses the consensus spectrum of the raw PPG signal. Therefore, an SRQI is calculated for every resulting consensus spectrum. A strategy similar to that followed for artifact detection is applied: for a given window, if the SRQI is inferior to some threshold value $\mathcal{Q}$, the RR estimation in the corresponding window is considered to be not accurate enough and is thus rejected. The RMSE values are then calculated for each record for different values of $\mathcal{Q}$ between 0 and 1.

For CS-Med, RR is estimated from modulation signals. An SRQI is attributed to every consensus spectrum derived from the AM, BW and FM signals. Only RR estimates with corresponding SRQI higher than $\mathcal{Q}$ for the three signals are kept to calculate the final RR estimate. The final RR estimate is then given by the median of $RR_i|i = \{AM, FM, BW\}$ that were kept. If all of the three AM, BW and FM signals fail to provide an SRQI above $\mathcal{Q}$, no final RR estimate is given.

$$RR = median(RR_{AM} \text{ (if } SRQI_{AM} > \mathcal{Q}), (RR_{BW} \text{ (if } SRQI_{BW} > \mathcal{Q}), (RR_{FM} \text{ (if } SRQI_{FM} > \mathcal{Q}))$$

The same analysis is also applied for scrutinizing the impact of SRQI on Karlen methods Fusion and Smart Fusion. So, an SRQI is attributed to every power spectrum of each signal AM, BW and FM and then the final RR is given by the same formula as above.

5.3.3 Results on Capnobase

The window length for RR estimation is fixed to $64s$ and the number of windows for consensus is fixed to 4 windows. No artifact detection module is used for the CS-PPG and CS-Med methods. The threshold $\mathcal{Q}$ on SRQI takes values from 0 to 1 with a step of 0.05. For $\mathcal{Q} = 0$, SRQI has no impact and all the RR estimates are accepted. It is worth noting that the value $\mathcal{Q} = 1$ cannot be achieved in reality because of the small fluctuations always present in the spectrum. So, the largest value is fixed to 0.99.

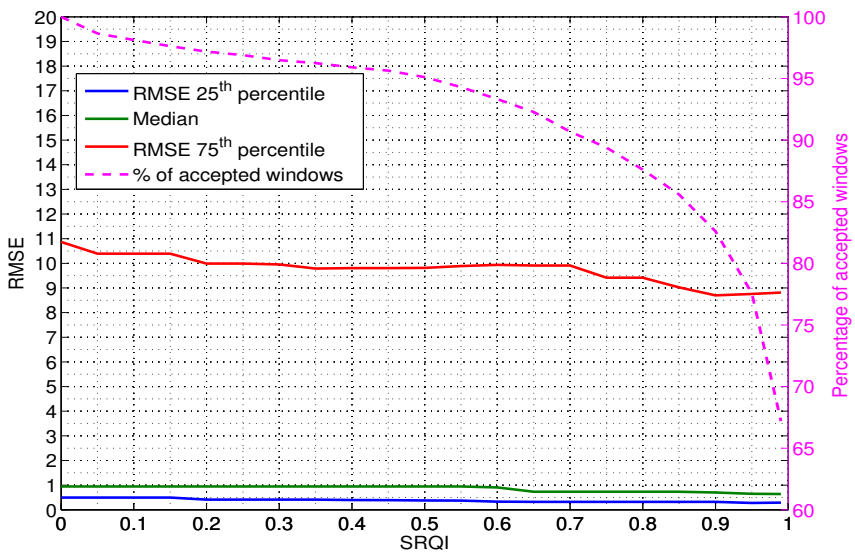


Figure 5.9 – RMSE of RR estimates from the CS-PPG method applied to the Capnobase data, versus the percentage of windows accepted as $\mathcal{Q}$ increases.

The results are given in figure 5.9 and 5.10.

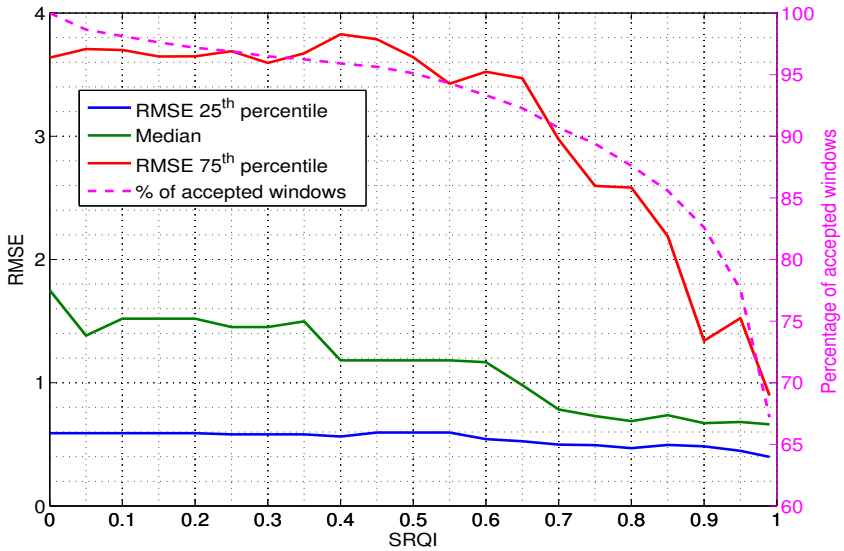


Figure 5.10 – RMSE of RR estimates from the CS-Med method applied to the Capnobase data, versus the percentage of windows accepted as $\mathcal{Q}$ increases.

For each value of $\mathcal{Q}$, the median, the 25th percentile and the 75th percentile of the RMSE are given. The percentage of the accepted RR values is given by the number of accepted RR values divided by the total number of RR estimates.

Little improvement is observed for the CS-PPG method after discarding RR estimates whose SRQI is less than 0.7. But, the values of the RMSE 75th percentile are still too high and do not guarantee a good accuracy of RR estimates.

However, for the CS-Med method (figure 5.10), similar variations are observed for the acceptance rate curve and the RMSE 75th percentile curve. In fact, as the windows of poor spectral quality are discarded, the accuracy of the RR estimation is improved and the inter quantile distance is reduced. The best performance measurements are found for $\mathcal{Q} = 0.99$ where the median of RMSE is 0.66 breath/min and the 25th and the 75th percentile are 0.39 and 0.89 breath/min, respectively. The consensus spectrum have the advantages of getting high resolution peaks. In fact, the regular peak, which frequency is the same for consecutive windows, has a high amplitude compared to other peaks. The amplitude of the other peaks becomes negligible. For this reason, although the SRQI threshold is high, we get an important percentage of accepted windows, which is about 67.2%.

The SRQI criterion is then tested with reference methods of the literature: the Fusion and Smart Fusion of Karlen & *al.*. The online Capnobase package contains the results of Karlen methods for a window of $64s$ with a shift of $3s$. RR estimates from

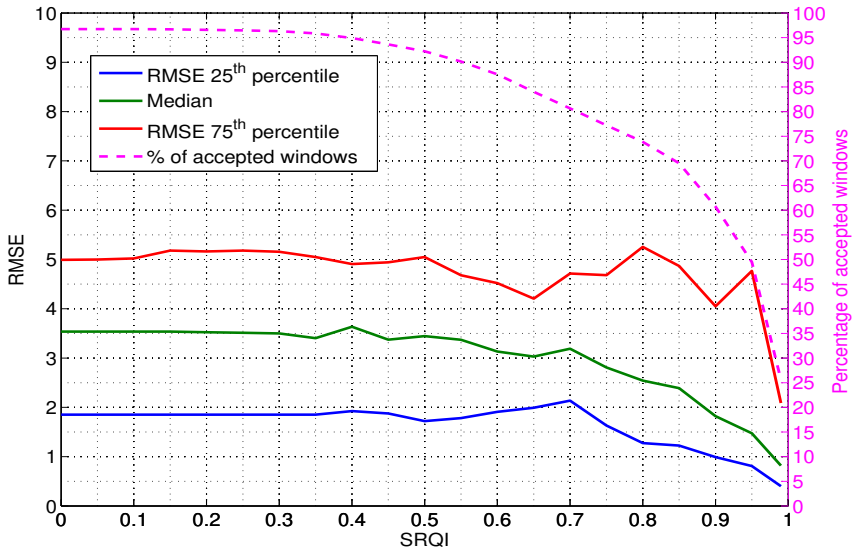


Figure 5.11 – RMSE of RR estimates from the Fusion method applied to the Capnobase data, versus the percentage of windows accepted as $\mathcal{Q}$ increases.

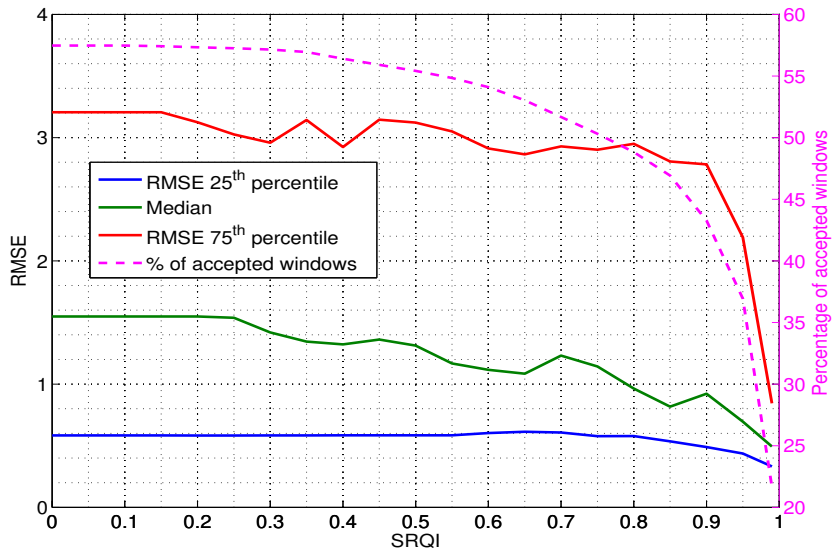


Figure 5.12 – RMSE of RR estimates from the Smart Fusion method applied to the Capnobase data, versus the percentage of windows accepted as $\mathcal{Q}$ increases.

AM, FM and BW signals are given for each clean window. Results from artifactual segments detected by the IMS method described in [31], are automatically discarded for both the Fusion and the Smart Fusion methods. In addition to this preliminary selection, for the Smart Fusion method, only RR estimates that meet the Smart Fusion criterion [31] are kept . For this reason, before applying the SRQI criterion, the initial acceptance rate for the Fusion method is 96.7% and is about 57% for the Smart Fusion method.

Now, when using the SRQI criterion, the performance measurements of the Fusion method are improved from a RMSE of 3.53(1.85-4.99) breath/min to a RMSE of 0.81 (0.39-2.08) breath/min. However, the accepted rate is deteriorated and is about 25.5% (figure 5.11).

The same observations are found with the Smart Fusion method when coupled to the SRQI criterion. A great amelioration is noticed for $\mathcal{Q} = 0.99$. In fact, the RMSE decreases from 1.54 (0.58-3.2) breath/min to 0.33 (0.49-0.84) breath/min but the accepted rate drops to 21.65%.

5.3.4 Results on Reastoc

The same tests are carried out with the Reastoc records. For the CS-PPG method (figure 5.13), no change in the RMSE values has been observed from $\mathcal{Q} = 0$ to $\mathcal{Q} = 0.75$. But, from $\mathcal{Q} = 0.8$ to $\mathcal{Q} = 0.95$, the RMSE values increase. In fact, the SRQI criterion rejected good RR estimates and not erroneous ones, thus generating a higher RMSE. This trend was corrected for $\mathcal{Q} = 0.95$ where RMSE=0.39(0.19-1.41) breath/min and the accepted rate is about 75.6%.

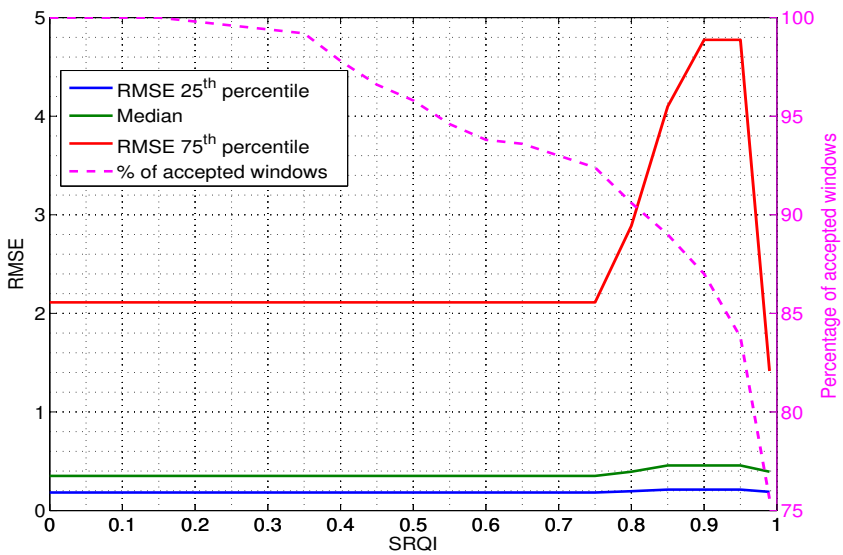


Figure 5.13 – RMSE of RR estimates from the CS-PPG method applied to the Reastoc data, versus the percentage of windows accepted as $\mathcal{Q}$ increases.

However, for the CS-Med method (figure 5.14), the SRQI has better impact on the RMSE results. The same behavior as for the Capnobase data is observed. The RMSE 75th percentile decreases as $\mathcal{Q}$ increases, which shows that the SRQI criterion is rejecting non relevant RR estimates. The best performance is observed for $\mathcal{Q}$=0.99

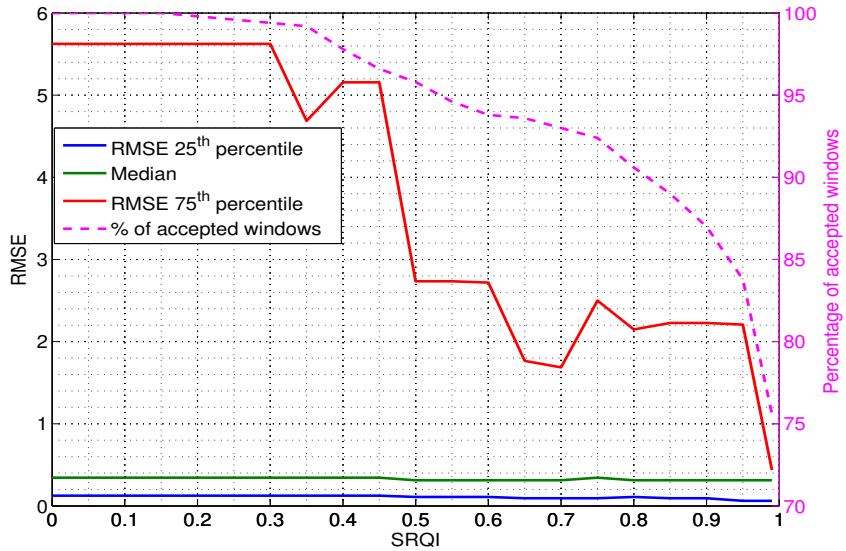


Figure 5.14 – RMSE of RR estimates from the CS-Med method applied to the Reastoc data, versus the percentage of windows accepted as $\mathcal{Q}$ increases.

with RMSE=0.31 (0.06-0.43) breath/min. Nerveless, it has to be noted that this error rate has same magnitude order as that found with Capnobase.

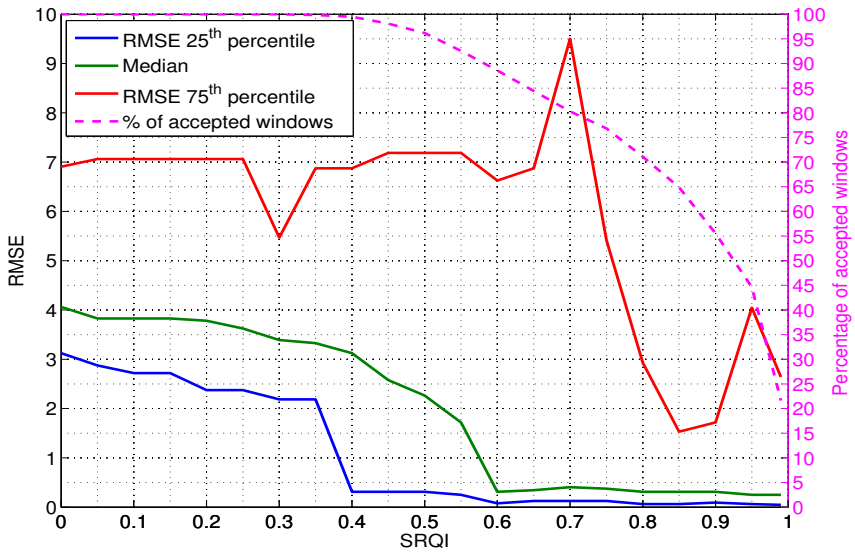


Figure 5.15 – RMSE of RR estimates from the Fusion method applied to the Reastoc data versus the percentage of windows accepted as $\mathcal{Q}$ increases.

Let us see now if the SRQI can improve results obtained from the Fusion and Smart Fusion methods. In fact, in figure 5.15, for the Fusion method, we observe that the RMSE results are ameliorated as $\mathcal{Q}$ increases but are not stable. In fact, for $\mathcal{Q}>0.9$, the SRQI has the tendency to discard good RR estimates.

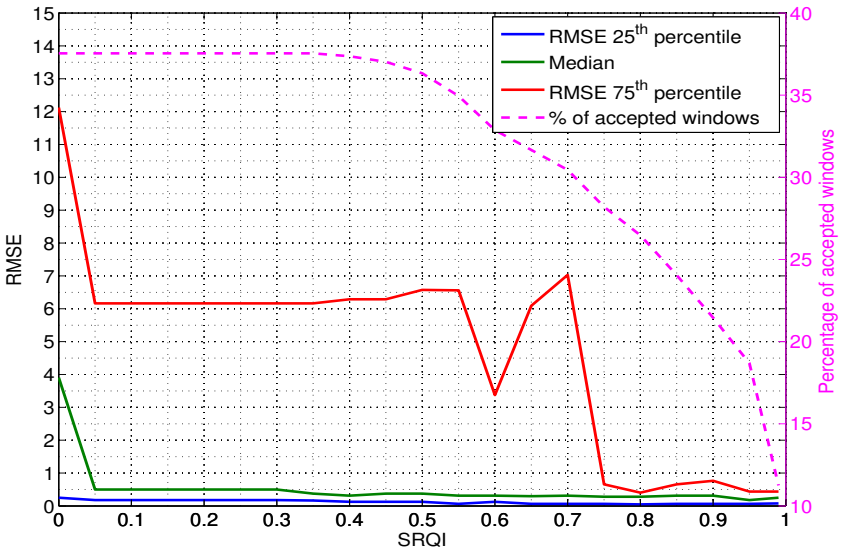


Figure 5.16 – RMSE of RR estimates from the Smart Fusion method applied to the Reastoc data, versus the percentage of windows accepted as $\mathcal{Q}$ increases.

For the Smart Fusion method (figure 5.16), a significant improvement is observed for thresholds $\mathcal{Q}$>0.75. However, the accepted rate is too low. In fact, the selective condition of the Smart Fusion [31] in addition to the SRQI are too strict. Thus, performance is improved at the expense of the percentage of accepted RR estimates.

5.3.4.1 Results according to age range and ventilation mode

In section 4.5.1, for Capnobase data, results were impacted by the patient's age range and the respiration ventilatory mode. Further tests are carried out in this section to assess the SRQI effect according to age and ventilatory mode. The RMSE results are reported in Table 5.1 when using the CS-Med method. Based on the previous tests, we fix the SRQI threshold to 0.99.

Results are significantly improved for all age ranges. But, this improvement was followed by some degradation of the acceptance rate, especially for persons over 30 years, for whom the acceptance rate is about 55.64%.

According to the ventilation mode, results are improved as well. However, there is no significant difference between performance measurements for patients under controlled ventilation and patients under spontaneous respiration.

In section 4.5.2, a high gap was found in the CS-Med performance between subjects < 60 and subjects $\geqslant 60$ years.

Table 5.1 – RMSE performances on Capnobase data when SRQI threshold is fixed to 0.99.

Results according to age range		
Age range	RMSE (breath/min)	Acceptance rate
<12	0.6803 (0.6641 - 0.7824)	67.30%
12-30	0.5078 (0.2958 - 1.7936)	81.22%
>30	0.5082 (0.3940 - 0.7602)	55.64%
Results according to ventilation mode		
Ventilation mode	RMSE (br/min)	Acceptance rate
spontaneous	0.6665 (0.4852 - 0.7492)	68.70%
controlled	0.6488 (0.3589 - 1.7936)	65.46%

For tests involving the SRQI criteria, a great amelioration is observed especially for subjects $\geqslant 60$. The RMSE passes from 3.43 (0.29-8.76) breath/min to 0.28 (0.12-4.31) breath/min. But, in return, the acceptance rate becomes only 58.9%.

Table 5.2 – RMSE performances on Reastoc data when the SRQI threshold is fixed to 0.99.

Results according to age range		
Age range	RMSE (breath/min)	Acceptance rate
<60	0.3125 (0.0469 - 0.4375)	88.61 %
>=60	0.2812 (0.1250 - 4.3125)	58.9 %

5.4 Discussion

5.4.0.1 Artifact detection versus SRQI

The artifact detection module improves slightly the performance of the RR estimation. However, this improvement is not sufficient. In fact, having good RR estimates is conditioned by many other factors besides artifacts. These factors cannot be detected by observing the temporal components of the PPG signals only. Spectral analysis of PPG signals reveals that the respiratory frequency is in some cases hidden by other phenomena, even if the PPG signal is clean of noise as shown in figure 5.6.

The SRQI permits to classify spectra into two categories: The first one involves the good spectra that have one dominant peak only in the respiratory frequency range. The second category contains suspicious spectra that contain more than one dominant

peak. For the latter category, the frequency peak corresponding to the actuall RR could not be distinguished from harmonics or from other peaks resulting from Mayer waves.

Artifact distortion can be observed in the PPG spectrum. In fact, random peaks resulting from artifacts appear in the spectrum. If the magnitudes of the artifact peaks are negligible in comparison to RR peaks, the SRQI will be close to 1. In this case, good RR estimation can be assured, although the signal distortion results from artifacts. To the contrary, if the magnitudes of the artifact peaks are in the same range as that of the RR peak, the SRQI will have a low value and thus good RR estimates cannot be reliable. The SRQI reveals to be an appropriate tool for detecting relevant segments for RR estimation, even in case of signal distortions resulting from artifacts. Consequently, in the context of RR estimation, when using the SRQI criteria, the artifact detection step turns out to have no significant contribution. Therefore, a considerable processing time can be saved by involving SRQI criteria only.

By observing the results of the SRQI impact for each method and each database, the best performance is given by the CS-Med method for both the Capnobase and Reastoc databases.

When comparing the CS-Med to the CS-PPG, the CS-Med gives better accuracy than the CS-PPG. In fact, the respiration information can be present in one of the three respiratory signals or more. The selection, via the SRQI criterion, of the better respiratory signals among the AM, BW and FM signals permits to estimate the RR from pertinent signals only. In addition, the information about respiration activity is more likely to be distinguishable from the AM, BW and FM signals than from raw PPG signals. Besides, no conclusion could be drawn up about the accuracy prevalence of one signal among the three available respiratory signals. A remarkable performance change is noticed especially for the Reastoc database. In fact, without the use of the SRQI, the CS-PPG outperforms the CS-Med. However, with the integration of the SRQI, the balance shifted in favor of the CS-Med method.

When comparing the CS-Med to Karlen's methods Fusion and Smart Fusion, the CS-Med method provides the best compromise between performance and acceptance rate. The use of the SRQI criterion on the Consensus spectrum is more efficient than when it is applied to the FFT spectrum. In fact, the resolution of peaks resulting from the consensus method is better than that obtained by FFT. The magnitudes of the peaks in the neighborhood of a peak of interest in the consensus spectrum is very high in comparison to those of other peaks. However, for the FFT spectrum, peaks due to brief phenomena (artifacts, noise, filtering effects) can have magnitudes around those

of peaks of interest. Consequently, the ratio of energy defined by the SRQI is more likely to tend to 1 when applied to the consensus spectrum than when applied to the FFT spectrum. This observation explains the fact that the Fusion and Smart Fusion methods have very low acceptance rate compared to the CS-Med method.

5.4.0.2 Comparison to existing methods

As described previously, in the literature, few studies involved the description of explicit quality indexes for respiratory signals. Proposed algorithms contain usually a preprocessing module for detecting artifacts and thus qualifying temporal and pulsatile components of PPG signals. Otherwise, some criteria are used to select the more accurate RR estimates when fusing values from the three respiratory signals. For example, in Karlen & *al.* [31], the criterion selecting the best estimates is the standard deviation of the three respiratory rates RR_{AM}, RR_{FM} and RR_{BW}. If the standard deviation is more than 4, all the RR estimates in the corresponding window are rejected. This criterion assumes implicitly the homogeneity of the three RR values. All the three RR values should be in the same range. However, it does not take into consideration the quality of the respiratory modulations, as the SRQI actually does. Thanks to the SRQI criterion, only one good modulation signal can be sufficient for estimating the final RR. Indeed, with this selection, the acceptance rate of the Smart Fusion is about 55,63% for Capnobase data and 38% for Reastoc data which is considered too low compared to the method of CS-Med coupled to the SRQI. In addition, the performance stability of the Smart Fusion method remains questionable when changing the database. In contrast, the same performance range is established for both the Capnobase and the Reastoc databases when using the CS-Med equipped with the SRQI.

Another recent study by Drew & *al.* [62] was interested in establishing respiratory modulation quality from AM, FM and BW signals for both PPG and ECG signals. For every modulation signal, three Respiratory Quality Index (RQI) metrics are calculated for each window. The methods used are based on FFT, auto-regression and auto correlation. The three RQIs metrics are then fused using a linear regression model. After obtaining a single value for each window and each signal modulation, the six resulting RQIs (AM, FM and BW from both ECG and PPG) are thresholded. The RQIs higher than the threshold are then kept. A weighted average of the RR estimates is then calculated using the retained RQIs as weights.

The method by Drew& *al.* has a good performance with a good acceptance rate. In fact, for the Capnobase data, the mean absolute error (MAE) was 0.71 $\pm$ 0.89

breath/min, while eliminating only 1.3 % of the windows through the proposed RQI criterion. For the BIDMC dataset, which was also used in the study [58] discussed in section 4.6.3, the MAE was about 3.12 ± 4.39 breath/min and the reject rate was about 23.2%. It is true that such algorithm gives the best performance until today for estimating RR. However, this method requires simultaneous ECG signal in addition to the PPG signal. This condition cannot be affordable in all patients monitoring contexts, especially for remote home monitoring. Other disadvantage of this method is the heavy processing cost. In fact, for each window, in all, 18 (6x3) RQI values must be calculated. The linear regression model for fusing RQI values was trained on a proportion of Capnobase data, then applied to MIMIC data, which may explain the difference of performance between the two databases. Consequently, the model for fusing RQI can be not representative for all age classes and subject conditions.

The SRQI criterion with the CS-Med method presents many advantages compared to the latter method. The main one is the simplicity of the method with very few parameters. This is a key element for future implementation in real-time environments. The second advantage is the universality of the method. In fact, similar performance is found for both the Capnobase and the Reastoc databases, for all age ranges and for different ventilatory modes. The third advantage is the robustness of this method with respect to artifacts. Indeed, the SRQI can detect and reject cases where artifactual segments affect RR estimation. Therefore, there is no need for additional processing module dedicated to artifact detections.

However, the acceptance rate of the CS-Med method is still relatively moderate compared to the results found by Drew& *al.*. However, it is worth noticing that the use of the ECG signal compensates the lack of respiratory information from the PPG signal, which may explain the acceptance rate difference between our method and that of Drew& *al.*. Nevertheless, further experiments should be carried out to ameliorate this acceptance rate by developing tools that differentiate RR peak from harmonics and Mayer's wave peaks. Large scale dataset composed of subjects from intensive care, ambulatory and at home health care should be tested to validate the stability of the proposed method.

5.5 Conclusion

In this chapter, we have presented an analysis of the impact of artifact detection for improving RR estimation. The positive impact of such detection is somewhat relative and not that dramatic. It follows that other tools are necessary to improve RR

estimation from PP signal modulations. In this respect, we have assessed a new metric for qualifying respiratory frequency spectral peaks. The SRQI criterion combined to the RR estimation methods exhibits good alternative for selecting good RR estimations and improving the performance of the estimation algorithms.

The best performance was observed when combining the SRQI to the CS-Med method for both the Capnobase and the Reastoc datasets. The accuracy of the RR estimation has been considerably improved. In summary, the key steps for an effective respiratory rate estimation system are given by figure 5.17.

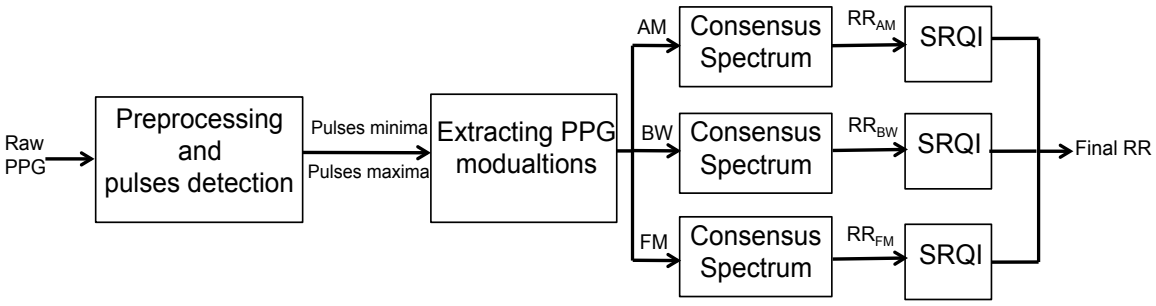


Figure 5.17 – Main steps for a respiratory rate estimation system based on the CS-Med method and SRQI criterion.

Otherwise, the acceptance rate of this method is still moderate especially for aged people. Further studies should be investigated to ameliorate this rate. Nevertheless, additional tests are required to validate the reproducibility of this algorithm's performance in others contexts.

CHAPTER 6 Conclusion and perspectives

6.1 Summary and main contributions 101
6.2 Perspectives and future works 104

In this thesis, we addressed the problem of developing an effective system for robust estimations of the respiratory rate from PPG signal. The main conclusions and original contributions are summarized in this chapter. Possible perspectives are then proposed.

6.1 Summary and main contributions

As stated in Chapter 1, the main objective of this thesis was to develop effective non-invasive methods for estimating accurate respiratory rate from only PPG signal. The choice of this signal is justified by the fact that PPG signal is widely used in clinical routine and also in other applications such as remote monitoring. In chapter 2, an overview of the PPG signal features was presented. In addition to the principal use of PPG signals for measuring blood oxygen saturation, vital indicators such as HR and RR can be extracted from the PPG signal. The main advantages of this signal are: non-invasiveness, simplicity of use, cost-effectiveness. However, some drawbacks limit the use of the PPG signal for monitoring HR and RR. The two main limits of PPG are: firstly, the sensitivity of the PPG signal to artifacts resulting from subject's motion and acquisition environments; secondly, the variability of the signal depending on age, gender and subject's health state. Therefore, mathematical modeling of the PPG signal cannot always be representative of the real PPG characteristics and variability. To get out from these limitations, new signal processing approaches were proposed. In chapter 3, methods for detecting artifacts from the literature were discussed. As a result of the PPG signal variability, there are no statistical models to differentiate clean PPG from artifactual PPG. So filtering methods fail to return clean valuable PPG signals. The

classification methods have good potential but the training issues encountered every time datasets may change limit their use for operational application in presence of unknown new PPG signals. We propose a novel method based on pulse morphology comparison by template matching and a decision test by Random Distortion Testing. This method requires little parameterization and has the advantage of fitting the model to each considered signal. A first step, with fixed RDT threshold for every record, was carried out by testing the proposed method on short records. The algorithm performs well compared to the state-of-the-art method proposed by Sukor & *al.* The method was then generalized by proposing an adaptive threshold for long records. In this case, the RDT threshold is updated progressively as new clean pulses are detected. So, little fluctuations resulting from physiological activity are taken into account through the new updated threshold. The accuracy achieved by adaptive RDT when applied to Capnobase data and Reastoc data were respectively, 91% and 92%. The originality of the proposed approach lies in its universality and reproducibility. In fact, we have proposed a standalone system that can be adapted for every record and could update its parameters when substantially changes occur.

After this preliminarily step, we have discussed in chapter 4 the feasibility of a system for robust estimation of the respiratory rate. Literature presents a plethora of methods for estimating RR, generally in one of the following contexts: healthy subjects, ambulatory patients or patients in general care floor. However, the assessment of these methods, whatever the base is, is not always discussed, especially for the methods which require a predefinition parameter like the order for AR methods or the level of decomposition for wavelet transforms. We have proposed some spectral methods for estimating RR from raw PPG signals and also from respiratory induced signals extracted from PPG signals. We have looked for accurate tools that allow us to have the best spectral peak resolution corresponding to the respiratory rate. Best performance was founded with the so-called CS-PPG method, which is used with raw PPG signals, and so-called CS-Med method, which is used with respiratory induced signals. In fact, we have employed the Consensus Spectrum (CS) by multiplying consecutive FFT spectra in order to reveal the regular peak corresponding to the respiratory frequency. The paramount advantage of the CS-Med and CS-PPG methods is that few parameters are needed regardless of the data characteristics. In addition, the proposed methods are relatively simple. Consequently, the computational requirements of the algorithms are not costly and are appropriate for implementation in real-time setting. However, the estimation of RR by these methods depend on several factors, such as age and ventilatory mode. We have experimentally demonstrated in chapter 5 that even for PPG

signals without artifacts, the algorithms may fail to estimate an accurate RR. Actually, having good RR estimates does not depend necessarily on the pulsatile PPG components, but rather on the quality of the PPG modulation signals. Therefore, the Spectral Respiratory Quality Index is introduced in chapter 5 to qualify the resolution of the spectral respiratory peak. A great improvement in performance is observed especially when combining the SRQI criterion to the CS-Med method. We have thus proposed a whole system for an effective estimation of RR. This system involves extracting PPG modulations, calculating RR from the three modulation signals by consensus spectrum, estimating the final RR by calculating the median value of the three RR with respect to the SRQI criterion and finally communicating the final RR estimate to the clinician.

Another contribution of this thesis is the experimental testing carried out for different patients' states. The first database that we have used to assess our algorithms for estimating RR was Capnobase. It is a good online data set to benchmark studies on RR estimation. We have used this dataset to compare our results to those of Karlen's methods and other methods from the literature. We have found that our proposed system outperforms the Karlen methods in terms of accuracy and also in terms of acceptance rate. However, the Capnobase data cannot be considered as an absolute reference for assessing results, because of its lack of representativity of different age classes. We therefore have explored the versatility of our approach when facing another dataset by using the Reastoc dataset. Patients from this datasets are in intensive care. They are relatively aged compared to Capnobase data and under mechanical ventilation. We have shown that the CS-Med coupled to the SRQI criterion has also good performance for the Reastoc data.

In summary, monitoring vital signs from PPG signal is a good alternative to traditional measurement systems. Artifact detection from PPG is important to get pulsatile information from PPG such as HR measurement and to study HR variability. However, the artifact detection is not necessarily useful when it comes to estimating RR. The CS-MED method combined to the SRQI criterion gives accurate RR estimates without recourse to artifact detection. Thanks to its simplicity, reliability and minimum resource requirement, the proposed approach can be integrated in real-time clinical workflows for different monitoring applications, whether in hospitals or in remote home care. It represents also an alternative to classification methods, which require large training data.

6.2 Perspectives and future works

Possible research lines to enlarge the work in this thesis are presented. Future work with large-scale tests is necessary to validate the algorithms in order to generalize this technique of estimating RR for different patients and contexts. In addition, the presented results are validated in this thesis on offline data. Further tests should evaluate the potential of the proposed system in real-time clinical environments. The major clinical interest for estimating RR is the monitoring of the respiratory activity in order to detect abnormalities on patients' health state. Additional computing tools should be developed to detect these events. Then, continuous monitoring tests should be carried out for specific patient anomalies to assess the potential of the proposed method for detecting alarming cases.

Despite the accuracy of the proposed method for estimating RR, the acceptance rate may be judged to be insufficient for real monitoring application. Other lines of investigations should thus be examined to improve this rate. In fact, knowledge from the patient's data concerning his/her age and health state should be analyzed and introduced in the RR estimation methods. Until today, the proposed algorithms for monitoring PPG signals generally consider only features about the signals, but do not use parameters characterizing the patient status in the processing system. Future approach should therefore involve subjects' characteristics and not signals only to discriminate novel medical states.

In the present work, we have proposed some methods for detecting artifacts and estimating features from PPG signal. However, these methods can be generalized to other physiological signals with similar characteristics, especially if they involve recursive patterns with some signal modulations. This is the case, for example, with ECG signals. Therefore, it would be desirable to carry out further studies and experiments to assess the universality of the proposed methods for other physiological signals.

Bibliography

[1] M. Feissel. The pulse oxymetry plethysmographic curve: an old signal with a great future? principles and clinical applications. *Reanimation*, 16:124–131, 2017.

[2] J Abdul Sukor, S J Redmond, and N H Lovell. Signal quality measures for pulse oximetry through waveform morphology analysis. *Physiological Measurement*, 32(3):369, 2011.

[3] Bhusana Premanode, Jumlong Vongprasert, and Christofer Toumazou. Noise reduction for nonlinear nonstationary time series data using averaging intrinsic mode function. *Algorithms*, 6(3):407–429, 2013.

[4] Leroy Hood and Mauricio Flores. A personal view on systems medicine and the emergence of proactive p4 medicine: predictive, preventive, personalized and participatory. *New Biotechnology*, 29(6):613–624, 2012.

[5] D. Pastor and Q. T. Nguyen. Random distortion testing and optimality of thresholding tests. *IEEE Transactions on Signal Processing*, 61(16):4161–4171, August 2013.

[6] A. Hertzman and C. Spealman. Observations on the finger volume pulse recorded photoelectrically. *Am. J. Physiol.*, 1937.

[7] JR. Squire. Instrument for measuring quantity of blood and its degree of oxygenation in the web of the hand. *Clin Sci*, 4, 1940.

[8] T.Aoyagi, M. Kishi, K. Yamaguchi, H. Hirai S. Nakajima, H. Takase, and A. Kuse. New pulsed-type earpiece oximeter (author's transl) [in japanese]. *Kokyu To Junkan*, 23, 1975.

[9] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):R1, 2007.

[10] Lena M. Nilsson. Respiration signals from photoplethysmography. *anesthesia and analgesia*, 117(4):859–865, 2013.

[11] Mayank Kumar, Ashok Veeraraghavan, and Ashutosh Sabharwal. Distanceppg: Robust non-contact vital signs monitoring using a camera. *Biomedical Optics Express*, 6(5):1565–1588, 2015.

[12] B. G. Lee, B. L. Lee, and W. Y. Chung. Smartwatch-based driver alertness monitoring with wearable motion and physiological sensor. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6126–6129, Aug 2015.

[13] P. Shi, S. Hu, Y. Zhu, J. Zheng, Y. Qiu, and P. Y. S. Cheang. Insight into the dicrotic notch in photoplethysmographic pulses from the finger tip of young adults. *Journal of Medical Engineering & Technology*, 33(8):628–633, 2009.

[14] Wikipedia. Photoplethysmogram — wikipedia, the free encyclopedia, 2017. [Online; accessed 29-June-2017].

[15] Diego Martin-Martinez, Pablo Casaseca de-la Higuera, Marcos Martin-Fernandez, and Carlos Alberola-Lopez. Stochastic modeling of the ppg signal: A synthesis-by-analysis approach with applications. *IEEE Transactions On Biomedical Engineering*, 60(9), september 2013.

[16] R. Banerjee, A. Ghose, A. Dutta Choudhury, A. Sinha, and A. Pal. Noise cleaning and gaussian modeling of smart phone photoplethysmogram to improve blood pressure estimation. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 967–971, April 2015.

[17] Yun Luo, Wenfeng Li, Wenbi Rao, Xiuwen Fu, Lin Yang, and Yu Zhang. *A New Modeling Method of Photoplethysmography Signal Based on Lognormal Basis*, pages 12–21. Springer International Publishing, Cham, 2016.

[18] Eckhard Limpert, Werner A. Stahel, and Markus Abbt. Log-normal distributions across the sciences: Keys and clues. *Bioscience*, 51(5):341–352, May 2001.

[19] J. Pottecher, G. Bouzou, and A.Van de Louw. Monitorage de la saturation de pouls : intérêts et limites pulse oximetry's monitoring: advantages and limits. *Réanimation*, 12:30–36, 2003.

[20] J. W. Chong, N. Esa, D. D. McManus, and K. H. Chon. Arrhythmia discrimination using a smart phone. *IEEE Journal of Biomedical and Health Informatics*, 19(3):815–824, May 2015.

[21] A. Solosenko, A. Petrenas, and V. Marozas. Photoplethysmography-based method for automatic detection of premature ventricular contractions. *IEEE Transactions on Biomedical Circuits and Systems*, 9(5):662–669, Oct 2015.

[22] E. Gil, P. Laguna, J. P. Martínez, Ó. Barquero-Pérez, A. García-Alberola, and L. Sörnmo. Heart rate turbulence analysis based on photoplethysmography. *IEEE Transactions on Biomedical Engineering*, 60(11):3149–3155, Nov 2013.

[23] Paul S. Addison. Respiratory effort from the photoplethysmogram. *Medical Engineering and Physics*, 41:9–18, 2017.

[24] Nathaniel S. Marshall, Keith K. H. Wong, Peter Y. Liu, Stewart R. J. Cullen, Matthew W. Knuiman, and Ronald R. Grunstein. Sleep apnea as an independent risk factor for all-cause mortality: The busselton health study. *Sleep*, 31(8):1079–1085, 2008.

[25] Ayal Romem, Anat Romem, Dafna Koldobskiy, and Steven M. Scharf. Diagnosis of obstructive sleep apnea using pulse oximeter derived photoplethysmographic signals. *Clin Sleep Med.*, 10(3):285–290, 2014.

[26] C. Karmakar, A. Khandoker, T. Penzel, C. Schöbel, and M. Palaniswami. Detection of respiratory arousals using photoplethysmography (ppg) signal in sleep apnea patients. *IEEE Journal of Biomedical and Health Informatics*, 18(3):1065–1073, 2014.

[27] L. M. Sepúlveda-Cano, E. Gil, P. Laguna, and G. Castellanos-Dominguez. Sleep apnoea detection in children using ppg envelope-based dynamic features. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1483–1486, 2011.

[28] J. Lázaro, E. Gil, J. M. Vergara, and P. Laguna. Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. *IEEE Journal of Biomedical and Health Informatics*, 18(1):240–246, 2014.

[29] C.Weissman M.Shamir. Plethysmographic waveform variation as an indicator to hypovolemia. *Anesthesia & Analgesia*, 97(2):602, 2003.

[30] W. Karlen, M. Turner, E. Cooke, G. Dumont, and J. M. Ansermino. Capnobase: Signal database and tools to collect, share and annotate respiratory signals. In *Annual Meeting of the Society for Technology in Anesthesia (STA)*, West Palm Beach, 2010.

[31] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont. Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering*, 60(7):1946–1953, July 2013.

[32] Pengfei Wei, Ruiwen Guo, Jingmeng Zhang, and Y. T. Zhang. A new wristband wearable sensor using adaptive reduction filter to reduce motion artifact. In *2008 International Conference on Information Technology and Applications in Biomedicine*, pages 278–281, May 2008.

[33] H. H. Asada, P. Shaltis, A. Reisner, Sokwoo Rhee, and R. C. Hutchinson. Mobile monitoring with wearable photoplethysmographic biosensors. *IEEE Engineering in Medicine and Biology Magazine*, 22(3):28–40, May 2003.

[34] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and K. A. Reddy. A novel approach for motion artifact reduction in ppg signals based on as-lms adaptive filter. *IEEE Transactions on Instrumentation and Measurement*, 61(5):1445–1457, May 2012.

[35] Fulai Peng, Zhengbo Zhang, Xiaoming Gou, Hongyun Liu, and Weidong Wang. Motion artifact removal from photoplethysmographic signals by combining temporally constrained independent component analysis and adaptive filter. *BioMedical Engineering OnLine*, 13(1):50, Apr 2014.

[36] D. L. Donoho and I. M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[37] M. Raghuram, K. Venu Madhav, E. Hari Krishna, and K. Ashoka Reddy. Evaluation of wavelets for reduction of motion artifacts in photoplethysmographic signals. In *10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010)*, pages 460–463, May 2010.

[38] Q. T. Nguyen, D. Pastor, F. Lellouche, and E. L'Her. Mechanical ventilation system monitoring: Automatic detection of dynamic hyperinflation and asynchrony. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 5207–5210, July 2013.

[39] N.E. Huang, Z. Shen, S.R. Long, M.L. Wu, H.H. Shih, Q. Zheng, N.C. Yen C.C. Tung, and H.H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond.A*, 454, 1998.

[40] Qian Wang, Ping Yang, and Yuanting Zhang. Artifact reduction based on empirical mode decomposition (emd) in photoplethysmography for pulse rate detection. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 959–962, Aug 2010.

[41] Jo Woon Chong, Duy K. Dao, S. M. A. Salehizadeh, David D. McManus, Chad E. Darling, Ki H. Chon, and Yitzhak Mendelson. Photoplethysmograph signal reconstruction based on a novel hybrid motion artifact detection–reduction approach. part i: Motion and noise artifact detection. *Annals of Biomedical Engineering*, 42(11):2238–2250, 2014.

[42] Q Li and G D Clifford. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiological Measurement*, 33(9):1491, 2012.

[43] Mohamed Elgendi, Ian Norton, Matt Brearley, Derek Abbott, and Dale Schuurmans. Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions. *PLoS ONE*, 8:1–11, 10 2013.

[44] D. Pastor and F.-X. Socheleau. Random distortion testing with linear measurements. *Signal Processing*, 145:116 – 126, 2018.

[45] Ainara Garde, Walter Karlen, J. Mark Ansermino, and Guy A. Dumont. Estimating respiratory and heart rates from the correntropy spectral density of the photoplethysmogram. *PLoS ONE*, 9(1), 01 2014.

[46] K. H. Chon, S. Dash, and K. Ju. Estimation of respiratory rate from photoplethysmogram data using time frequency spectral estimation. *IEEE Transactions on Biomedical Engineering*, 56(8):2054–2063, Aug 2009.

[47] P.S. Addison, J.N. Watson, M.I. Mestek, and R.S Mecca. Developing an algorithm for pulse oximetry derived respiratory rate : a healthy volunteer study. *Clin. Monit. Comput.*, (26):45–51, jan 2012.

[48] Paul S. Addison, James N. Watson, Michael L. Mestek, James P. Ochs, Alberto A. Uribe, and Sergio D. Bergese. Pulse oximetry-derived respiratory rate in general care floor patients. *Journal of Clinical Monitoring and Computing*, 29(1):113–120, Feb 2015.

[49] P. S. Addison and J. N. Watson. Secondary wavelet feature decoupling (swfd) and its use in detecting patient respiration from the photoplethysmogram. In

Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439), volume 3, pages 2602–2605 Vol.3, Sept 2003.

[50] C. Orphanidou, S. Fleming, S.A. Shah, and L. Tarassenko. Data fusion for estimating respiratory rate from a single-lead ecg. *Biomedical Signal Processing and Control*, 8(1):98 – 105, 2013.

[51] Jesús Lázaro, Eduardo Gil, Raquel Bailón, Ana Mincholé, and Pablo Laguna. Deriving respiration from photoplethysmographic pulse width. *Medical & Biological Engineering & Computing*, 51(1):233–242, Feb 2013.

[52] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton. Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering*, 64(8):1914–1923, Aug 2017.

[53] Peter H Charlton, Timothy Bonnici, Lionel Tarassenko, David A Clifton, Richard Beale, and Peter J Watkinson. An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. *Physiological Measurement*, 37(4):610, 2016.

[54] Irena Cosic. *The Resonant Recognition Model of Macromolecular Bioactivity Theory and Applications*. Switzerland: Birkhauser Verlag, 1997.

[55] Quang-Thang Nguyen. *Contribution to statistical signal processing with applications in biomedical engineering*. PhD thesis, Télécom Bretagne, 2012.

[56] D.Giavarina. Understanding bland altman analysis. *Biochemia Medica.*, 25(2):141–151, 2015.

[57] Gulshan Sharma and James Goodwin. Effect of aging on respiratory system physiology and immunology. *Clin Interv Aging.*, 1(3):253–260, 2006.

[58] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton. Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering*, 64(8):1914–1923, Aug 2017.

[59] C. M. Lee and Y. T. Zhang. Reduction of motion artifacts from photoplethysmographic recordings using a wavelet denoising approach. In *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering, 2003.*, pages 194–195, Oct 2003.

[60] H. Fukushima, H. Kawanaka, M. S. Bhuiyan, and K. Oguri. Estimating heart rate using wrist-type photoplethysmography and acceleration sensor while running. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2901–2904, Aug 2012.

[61] Claude Julien. The enigma of mayer waves: Facts and models. *Cardiovascular Research*, 70:12–21, 2006.

[62] D. Birrenkott, M. A. F. Pimentel, P. J. Watkinson, and D. A. Clifton. A robust fusion model for estimating respiratory rate from photoplethysmography and electrocardiography. *IEEE Transactions on Biomedical Engineering*, pages 1–1, 2017.

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Titre : Estimation robuste de la fréquence respiratoire par traitement et analyse du signal de photoplethysmographie

Mots clés : photopléthysmogramme, fréquence respiratoire, détection des artéfacts, indice de qualité

Résumé : Le photopléthysmogramme (PPG) est un signal optique acquis à partir de l'oxymètre de pouls, dont l'usage principal consiste à mesurer la saturation en oxygène. Avec le développement des technologies portables, il est devenu la technique de base pour la surveillance de l'activité cardio-respiratoire des patients et la détection des anomalies. En dépit de sa simplicité d'utilisation, le déploiement de cette technique reste encore limité pour deux principales raisons: 1. L'extrême sensibilité du signal aux distorsions. 2. La non-reproductibilité entre les sujets et pour les mêmes sujets, en raison de l'âge et des conditions de santé. L'objectif de cette thèse est le développement de méthodes robustes et universelles afin d'avoir une estimation précise de la fréquence respiratoire (FR) indépendamment de la variabilité intra et interindividuelle du PPG. Plusieurs contributions originales en traitement statistiques du signal PPG sont proposées. En premier lieu, une méthode adaptative de détection des artefacts basée sur la comparaison de modèle a été développée. Des tests par la technique Random Distortion Testing sont introduits pour détecter les pulses de PPG avec des artefacts. En deuxième lieu, une analyse de plusieurs méthodes spectrales d'estimation de la FR est proposée. Afin de mettre en évidence la robustesse des méthodes proposées face à la variabilité du signal, plusieurs tests ont été effectués sur deux bases de données avec de différentes tranches d'âge et des différents modes respiratoires. En troisième lieu, un indice de qualité respiratoire spectral (SRQI) est conçu dans le but de communiquer au clinicien que les valeurs fiables de la FR avec un certain degré de confiance.

Title : Effective signal processing methods for robust respiratory rate estimation from photoplethysmography signal

Keywords : photoplethysmography, respiratory rate, artifacts detection, quality index

Abstract : One promising area of research in clinical routine involves using photoplethysmography (PPG) for monitoring respiratory activities. PPG is an optical signal acquired from oximeters, whose principal use consists in measuring oxygen saturation. Despite its simplicity of use, the deployment of this technique is still limited because of the signal sensitivity to distortions and the non-reproducibility between subjects, but also for the same subject, due to age and health conditions. The main aim of this work is to develop robust and universal methods for estimating accurate respiratory rate regardless of the intra- and inter-individual variability that affects PPG features. For this purpose, firstly, an adaptive artefact detection method based on template matching and decision by Random Distortion Testing is introduced for detecting PPG pulses with artefacts. Secondly, an analysis of several spectral methods for Respiratory Rate (RR) estimation on two different databases, with different age ranges and different respiratory modes, is proposed. Thirdly, a Spectral Respiratory Quality Index (SRQI) is attributed to respiratory rate estimates, in order that the clinician may select only RR values with a large confidence scale. Promising results are found for two different databases.