14 / 6 / 2025

ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ

ΘΕΜΑ 12

ΕΚΤΙΜΗΣΗ ΚΑΡΔΙΑΚΗΣ ΣΥΧΝΟΤΗΤΑΣ ΑΠΟ ΒΙΝΤΕΟ (ΠΡΟΣΩΠΟ Ή ΔΑΧΤΥΛΟ)

ΕΞΑΜΗΝΙΑΙΑ ΕΡΓΑΣΙΑ

2024-2025

Ηλίας Ξανθόπουλος 58545

Πίνακας Περιεχομένων

Το παρόν report όπως και το Jupyter Notebook της υλοποίησης είναι ανεβασμένα στο GitHub στο repository: https://github.com/iliaxant/HR_Extraction_from_Video

1. Περιγραφή του προβλήματος

Το πρόβλημα που επιλύεται στην παρούσα εργασία είναι της εκτίμησης του καρδιακού ρυθμού (Heart Rate- HR) από βίντεο προσώπου ή δαχτύλου. Για την λύση αξιοποιείται η φωτοπληθυσμογραφία (photoplethysmography-PPG), μία οπτική μέθοδος μέτρησης των παλμών όγκου αίματος στους μικραγγειακούς ιστούς. Οι παλμοί αυτοί προέρχονται από την συστολή της καρδιάς, οπότε μετρώντας τον χρόνο μεταξύ διαδοχικών αυξήσεων του όγκου αίματος, μπορεί να εξαχθεί η καρδιακή συχνότητα. Η συμβατική μέθοδος PPG γίνεται μέσω οξύμετρου στο δάχτυλο, αλλά η συγκεκριμένη εφαρμογή έχει ως σκοπό την εκτέλεση φωτοπληθυσμογραφίας χωρίς καμία σωματική επαφή, μόνο με βίντεο του υποκειμένου. Ξεκινώντας από μία καταγραφή που περιλαμβάνει το πρόσωπο ή το δάχτυλο του υποκειμένου, μπορεί να γίνει ανάλυση των αόρατων στο ανθρώπινο μάτι χρωματικών αλλαγών του δέρματος που προκαλούνται από τους προαναφερόμενους παλμούς. Στην συγκεκριμένη υλοποίηση, επιλέγεται η εκτίμηση του HR από πρόσωπο, αλλά η ίδια μέθοδος μπορεί με λίγες μόνο αλλαγές να εφαρμοστεί και σε δάχτυλο (το οποίο μπορεί να βρίσκεται είτε σε επαφή με την κάμερα, είτε όχι).

2. Επισκόπηση διαφορετικών προσεγγίσεων

Η εξαγωγή του PPG σήματος είναι εφικτό να γίνει με διάφορες μεθόδους. Η πιο βασική, η οποία χρησιμοποιείται και στην παρούσα εφαρμογή, είναι η **παρακολούθηση με την πάροδο του χρόνου της μέσης έντασης των pixel του δέρματος σε ένα χρωματικό κανάλι**. Αν και οι διακυμάνσεις της έντασης είναι πολύ μικρές, καταφέρνουν να καταγράφονται από την κάμερα και επομένως μπορεί ευκολά από το σήμα που προκύπτει να εξαχθεί η καρδιακή συχνότητα.

Λύσεις παρόμοιες με την προηγούμενη αλλά με πολύ μεγαλύτερο robustness στον θόρυβο (π.χ. αλλαγές στον φωτισμό) είναι οι **μέθοδοι βασισμένη στην χρωματικότητα (Chrominance)**. Αυτές οι τεχνικές αντί να εργάζονται πάνω μόνο σε ένα χρωματικό κανάλι της εικόνας, συνδυάζουν και τα τρία λαμβάνοντας υπόψη μόνο την χρωματικότητα και αγνοώντας την φωτεινότητα (Luminance). Χαρακτηριστικές υλοποιήσεις που αξιοποιούν αυτή την τεχνική είναι η **CHROM** (De Haan & Jeanne, 2013) [1], η οποία συνδυάζει τα RGB σήματα σε δύο ορθογωνικά σήματα χρωματικότητας μέσω δύο σταθερών διανυσμάτων προβολής, και η **Plane-Orthogonal-to-Skin** ή αλλιώς **POS** (Wang et al., 2016) [2], η οποία αποτελεί βελτίωση της CHROM, αφού προσαρμόζει δυναμικά αυτά τα διανύσματα προβολής.

Ένας άλλος τρόπος επίλυσης του προβλήματος είναι ο **Blind Source Seperation (BSS)**, δηλαδή η εξαγωγή από παρατηρήσεις (βίντεο) των ανεξάρτητων πηγών που τις συνθέτουν. Κλασσική μέθοδος BBS είναι η **Ανάλυσης Ανεξάρτητων Συνιστωσών (Independent Component Analysis-ICA)**. Εφαρμόζοντας λοιπόν ICA στο βίντεο, υπολογίζονται τα σήματα όλων των παραγόντων που το επηρεάζουν, συμπεριλαμβανομένου και της καρδίας που μεταβάλλει τον όγκο αίματος και άρα το χρώμα του δέρματος.

Αν και αποτελεσματικές, οι προηγούμενες προσεγγίσεις δεν βοηθούν στην παρατήρηση των χρωματικών μεταβολών σε επίπεδο εικόνας, λόγω της πολύ μικρής τους έντασης. Ωστόσο υπάρχει μια τεχνική η οποία βοηθάει στην οπτικοποίηση των αλλαγών του χρώματος του δέρματος και αυτή είναι η **Ενίσχυση Κινήσεων (Motional Amplification)**. Αυτή η μέθοδος ενισχύει όλες τις μεταβολές που υπάρχουν στο βίντεο, είτε οφείλονται σε κίνηση, είτε αναφέρονται στο χρώμα. Έτσι, οι προηγουμένως ανεπαίσθητες αλλαγές στο χρώμα του δέρματος ενισχύονται σε τέτοιο βαθμό ώστε να είναι πλέον ορατές από το ανθρώπινο μάτι. Μάλιστα αυτή η ενίσχυση είναι τόσο μεγάλη, ώστε μετά την εφαρμογή της το δέρμα να “αναβοσβήνει” όπως φαίνεται στην εικόνα 1.

Εικόνα 1: (1η σειρά) Frames του βίντεο πριν την Ενίσχυση Κίνησης
(2η σειρά) Frames του βίντεο μετά την Ενίσχυση Κίνησης

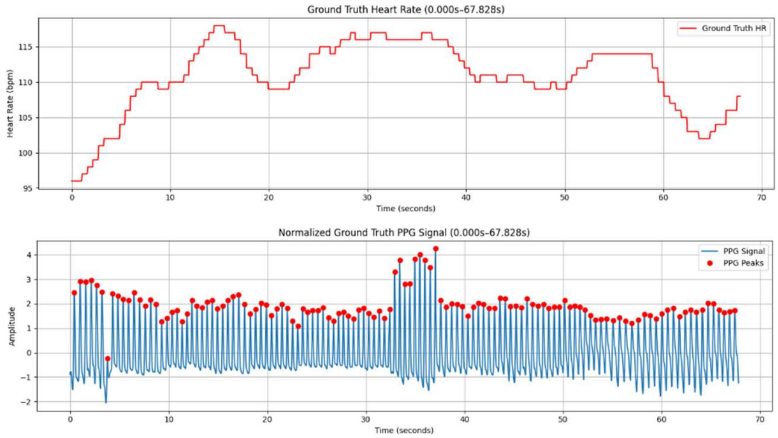
3. Επιλογή Δεδομένων

Τα δεδομένα που επεξεργάζεται η εφαρμογή είναι εκείνα του Dataset UBFC-rPPG [3]. Αυτή η βάση δεδομένων περιέχει βίντεο της τάξεως του ενός λεπτού το καθένα από πρόσωπα ανθρώπων σε απόσταση 1m περίπου από την κάμερα. Τα βίντεο λήφθηκαν από μια χαμηλού κόστους κάμερα (Logitech C920 HD Pro) με συχνότητα 30fps (29.951fps για την ακρίβεια) και ανάλυση 640×480. Το κάθε βίντεο συνοδεύεται από το αντίστοιχο Ground Truth (GT) το οποίο περιλαμβάνει τις ενδείξεις PPG σήματος και HR ενός οξύμετρου (CMS50E pulse Oximeter). Οι μετρήσεις αυτές είναι ευθυγραμμισμένες με το βίντεο έτσι ώστε κάθε frame να αντιστοιχίζεται σε μία ένδειξη του οξύμετρου. Συνεπώς και το GT έχει την ίδια συχνότητα δειγματοληψίας με το βίντεο.

Η βάση δεδομένων χωρίζεται σε δύο μέρη· το Dataset 1 και το Dataset 2. Το πρώτο είναι ένα απλοποιημένο σετ δεδομένων που περιλαμβάνει 8 βίντεο ανθρώπων που τους ζητήθηκε να παραμείνουν ακίνητοι (αν και εν τέλει μόνο ένα υποκείμενο δεν κουνήθηκε καθόλου), με τις συνθήκες φωτισμού να αλλάζουν σε ορισμένες περιπτώσεις και ίσως με λίγες κινήσεις στο background. Το δεύτερο περιλαμβάνει 42 βίντεο και αποτελεί ένα πιο ρεαλιστικό σετ δεδομένων. Είναι πιο ρεαλιστικό, γιατί δεν δόθηκε στα υποκείμενα καμία οδηγία να παραμείνουν ακίνητοι και κατά την διάρκεια της λήψης έπρεπε αλληλοεπιδρώντας με ένα υπολογιστή να παίξουν ένα μαθηματικό παιχνίδι με χρονικό όριο. Σε αντίθεση με το πρώτο, αυτό το dataset έχει greenscreen στο παρασκήνιο και οι συνθήκες φωτισμού δεν μεταβάλλονται.

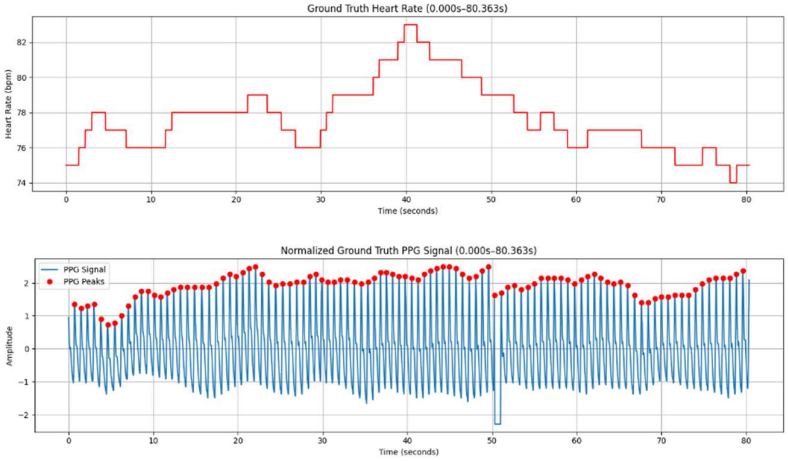
Από όλα τα δείγματα, επιλέχθηκε το subject 45 του Dataset 2 για την βασική εκτέλεση του αλγορίθμου, επειδή παρουσιάζει την λιγότερη κίνηση και δεν εμποδίζεται καθόλου το μέτωπο του. Συμπληρωματικά χρησιμοποιείται και το υποκείμενο 11 του Dataset 1 για την ανάδειξη αδυναμιών του αλγορίθμου αλλά αυτό αναλύεται στο *κεφάλαιο 6*. Στις παρακάτω *εικόνες 2-5* προβάλλονται τα Ground Truth διαγράμματα και ένα sample frame των υποκειμένων που χρησιμοποιούνται.

Εικόνα 2: Δείγμα frame του subject 45 του UBFC-rPPG Dataset 2



Εικόνα 3: Τα Ground Truth μεγέθη του subject 45 του UBFC-rPPG Dataset 2.
(α) Το HR με την πάροδο του χρόνου, (β) Το PPG σήμα με την πάροδο του χρόνου

Εικόνα 4: Δείγμα frame του subject 11 του UBFC-rPPG Dataset 1



Εικόνα 5: Τα Ground Truth μεγέθη του subject 11 του UBFC-rPPG Dataset 1.
(α) Το HR με την πάροδο του χρόνου, (β) Το PPG σήμα με την πάροδο του χρόνου

4. Περιγραφή αλγορίθμου και αποτελέσματα εκτέλεσης

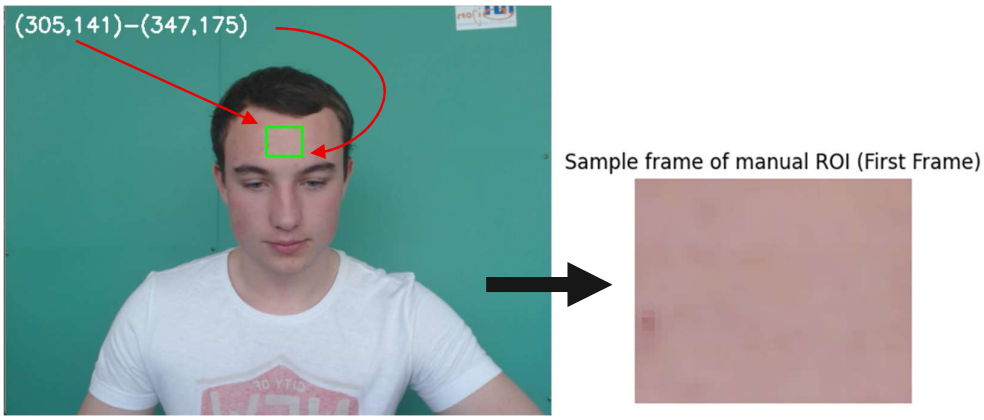
Η επίλυση της του προβλήματος εκτίμησης HR από βίντεο ανάγεται σε 5 βήματα:

1) Ορισμός της ROI
2) Υπολογισμός μέσης έντασης καναλιού (Spatial Averaging)
3) Κανονικοποίηση μέσης έντασης pixel
4) Φιλτράρισμα για απομόνωση ανθρώπινων καρδιακών συχνοτήτων
5) Εφαρμογή STFT

4.α Ορισμός της ROI

Όπως φαίνεται και παραπάνω στις *εικόνες 2* και *4*, τα frames δεν περιλαμβάνουν μόνο το πρόσωπο του υποκειμένου, αλλά το background και ένα μέρος του σώματος του. Η εφαρμογή του αλγορίθμου απευθείας σε όλο το frame θα οδηγήσει σε εσφαλμένα αποτελέσματα, επομένως κρίνεται αναγκαία η εξαγωγή από την κάθε εικόνα της περιοχής στην οποία υπάρχουν μόνο τα αναγκαία για τις μετρήσεις pixel. Συνεπώς, το πρώτο βήμα του αλγορίθμου είναι η εύρεση της περιοχής ενδιαφέροντος (Region of Interest-ROI), η οποία για τα βίντεο προσώπου είναι είτε το μέτωπο, είτε το μάγουλο. Για την συγκεκριμένη εφαρμογή αυθαίρετα επιλέγεται το μέτωπο, αλλά τα αποτελέσματα είναι τα ίδια και για το μάγουλο.

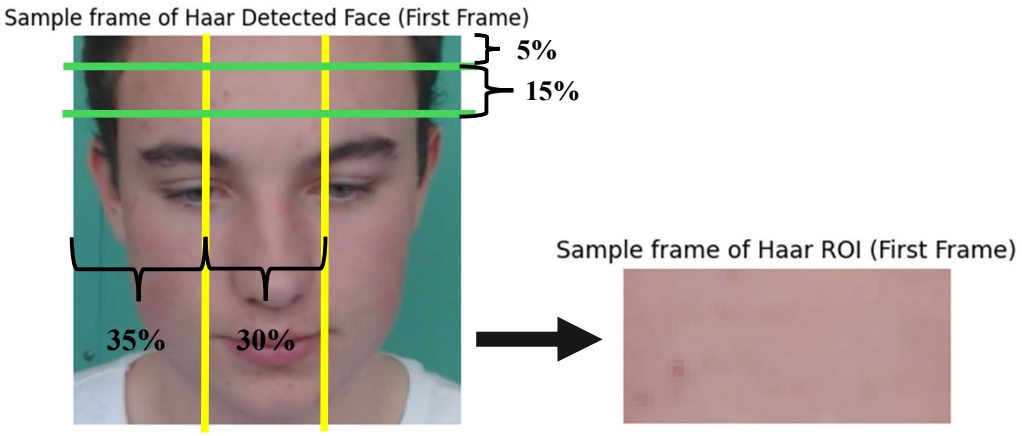
Ο ορισμός της ROI μπορεί να γίνει με δύο τρόπους· μπορεί να οριστεί είτε χειροκίνητα, είτε να προκύψει από ανίχνευση προσώπου. Στην πρώτη περίπτωση, απλώς εντοπίζονται οι θέσεις των pixel που αντιστοιχούν στο κομμάτι του μετώπου που μας ενδιαφέρει και χειροκίνητα επιλέγονται ως ROI για όλα τα frames του βίντεο (*εικόνα 6*). Αυτή η τεχνική προφανώς είναι πολύ απλή και γρήγορη, αλλά έχει τον περιορισμό ότι το υποκείμενο πρέπει να παραμένει όσο πιο ακίνητο γίνεται. Εάν κινηθεί έστω και στιγμιαία, τότε για αυτά τα λίγα frames η περιοχή εντός του "πλαισίου" θα είναι διαφορετική από εκείνη των προηγούμενων στιγμιοτύπων και άρα θα προκύψουν εσφαλμένα αποτελέσματα για εκείνο το χρονικό διάστημα.



Εικόνα 6: Η χειροκίνητα ορισμένη ROI για όλα τα frames του subject 45 και η ROI του πρώτου frame.

Από την άλλη το πρόβλημα αυτό εξαφανίζεται με τον δεύτερο τρόπο, διότι η ανίχνευση προσώπου σε κάθε frame εξασφαλίζει ότι ακόμα αν υπάρχει (εντός ορίων) κίνηση, τα pixels της ROI θα αντιστοιχούν στην ίδια περιοχή με εκείνη των προηγούμενων. Προϋπόθεση βέβαια είναι να μην κάνει το πρόσωπο καμία είδους περιστροφική κίνηση, γιατί, αν και θα ανιχνευτεί το πρόσωπο, η ROI θα είναι διαφορετική από της άλλες. Η μέθοδος face detection που επιλέγεται στην παρούσα εφαρμογή είναι ο αλγόριθμος Viola και Jones [4] και βασίζεται πάνω στα χαρακτηριστικά Haar.

Πρέπει να σημειωθεί ότι δεν χρησιμοποιείται απευθείας το ανιχνευμένο πρόσωπο ως περιοχή ενδιαφέροντος, αλλά ένα ποσοστό αυτού, ώστε να περιέχει μόνο το μέτωπο. Το ποσοστό που επιλέγεται φαίνεται παρακάτω στην *εικόνα 7* και είναι το ίδιο και για τα δύο υποκείμενα.



Εικόνα 7: Το ποσοστό του ανιχνευμένου προσώπου που ορίζεται ως ROI και το αποτέλεσμα στο πρώτο frame

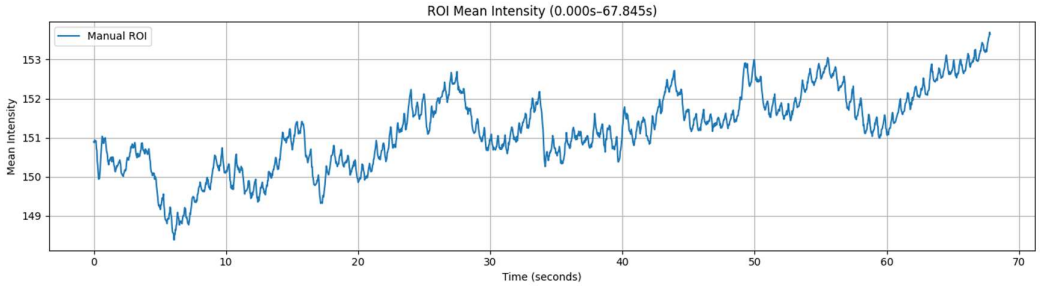
4.β Υπολογισμός μέσης έντασης καναλιού (Spatial Averaging)

Επόμενο βήμα μετά τον ορισμό της ROI είναι η εφαρμογή Spatial Averaging, το οποίο δεν είναι τίποτα άλλο από ό,τι ο υπολογισμός μέσα στην ROI της μέσης έντασης ενός καναλιού για κάθε frame του βίντεο. Δηλαδή, αν μέσα στην περιοχή ενδιαφέροντος υπάρχουν N pixels, τότε για το frame n υπολογίζεται το

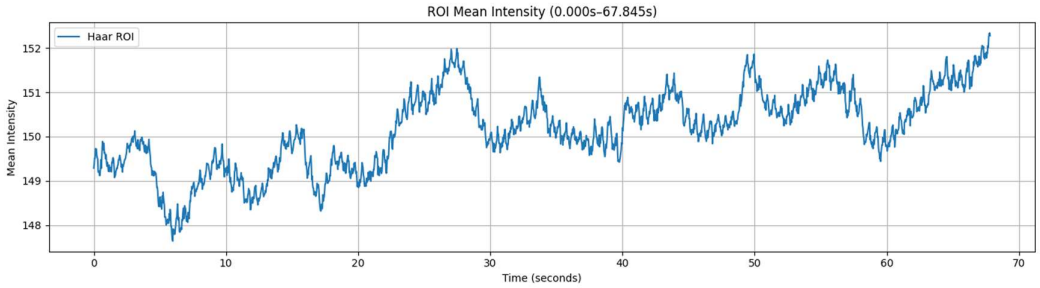
$$I_{ROI}(n) = \frac{\sum_{x,y \in ROI} I_{x,y}(n)}{N}$$

όπου $I_{x,y}$ είναι η ένταση του pixel στην x-οστή γραμμή και y-οστή στήλη του βίντεο που ανήκει στην ROI. Δεδομένου ότι η ανάλυση γίνεται στο χρωματικό χώρο RGB, το spatial averaging εκτελείται πάνω στο πράσινο κανάλι (G), το οποίο είναι το καλύτερο από τρία για πληθυσμογραφία. Η καταλληλότητα αυτού το καναλιού, σύμφωνα με το [5], οφείλεται στο ότι το πράσινο φως καταφέρνει να διεισδύσει μέσα στο δέρμα βαθύτερα από το μπλε και ταυτόχρονα απορροφάται από την αιμογλοβίνη καλύτερα από το κόκκινο.

Εφαρμόζοντας, με βάση τα παραπάνω, το Spatial Averaging, προκύπτει ένα σήμα της μορφής των *εικόνων 8* και *9*.



Εικόνα 8: Το σήμα που προκύπτει μετά από Spatial Averaging του G καναλιού της χειροκίνητα ορισμένης ROI



Εικόνα 9: Το σήμα που προκύπτει μετά από Spatial Averaging του G καναλιού της ορισμένης από face detection ROI

Παρατηρούμε ότι στο μεγαλύτερο μέρος τους τα σήματα που αντιστοιχούν στην manual και στην Facial Detection ROI μοιάζουν μεταξύ τους. Μια όμως εμφανής διαφορά είναι ότι το δεύτερο φαίνεται να είναι λίγο θορυβώδες σε σχέση με το πρώτο. Αυτό πιθανόν να οφείλεται στις μικρές μεταβολές φωτισμού, οι οποίες είναι φυσική απόρροια της συγκεκριμένης μεθόδου καθορισμού της περιοχής ενδιαφέροντος. Πιο συγκεκριμένα, μπορεί να ανιχνεύεται κάθε φορά το πρόσωπο του υποκειμένου με τον ίδιο τρόπο και ανεξάρτητα από την κίνηση του και έτσι η ROI να αντιστοιχεί πάντα στην ίδια περιοχή μεταξύ των frames, αλλά η πηγή φωτισμού δεν μετακινείται, πόσο μάλλον με τον ίδιο τρόπο που μεταφέρεται το πρόσωπο. Επομένως, ακόμη και αν το υποκείμενο κουνιέται ελάχιστα, η ROI θα φωτίζεται με διαφορετικό τρόπο σε σχέση με πριν και έτσι θα υπάρχει μια μικρή απόκλιση στην υπολογισμένη μέση ένταση, η οποία συνολικά σε όλο το σήμα παρουσιάζεται ως θόρυβος.

Από την άλλη, μπορεί το αποτέλεσμα της χειροκίνητης ROI να μην είναι τόσο ακριβές λόγω των μικρών κινήσεων, αλλά περιέχει πολύ λιγότερο θόρυβο επειδή η πηγή φωτισμού παραμένει σταθερή. Άρα, εφόσον οι μετακινήσεις είναι αρκετά μικρές και το κεφάλι του υποκείμενου δεν περιστρέφεται με οποιοδήποτε τρόπο, το πρόσωπο μέσα στην ROI θα φωτίζεται με τον ίδιο τρόπο μεταξύ των frames.

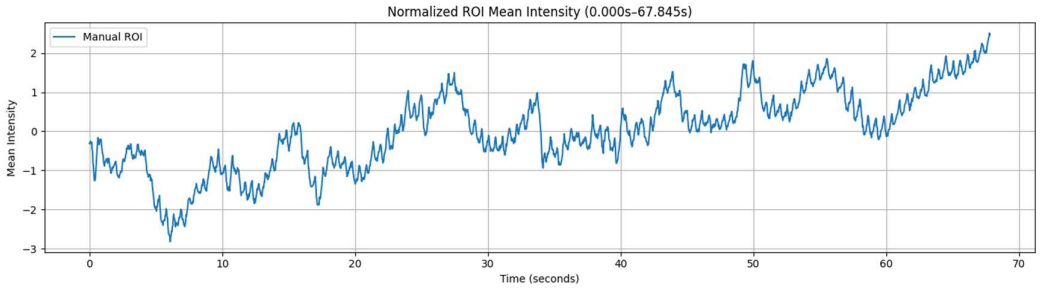
Συμπεραίνουμε λοιπόν ότι σε αυτό το στάδιο υπάρχει ένα trade-off αποτελεσματικότητας-θορύβου ανάμεσα στις μεθόδους ορισμού της ROI· manual για πιο καθαρό σήμα αλλά ανακρίβεια σε συγκεκριμένες περιπτώσεις, face detection για ορθότητα αλλά περισσότερο θόρυβο.

4.γ Κανονικοποίηση μέσης έντασης pixel

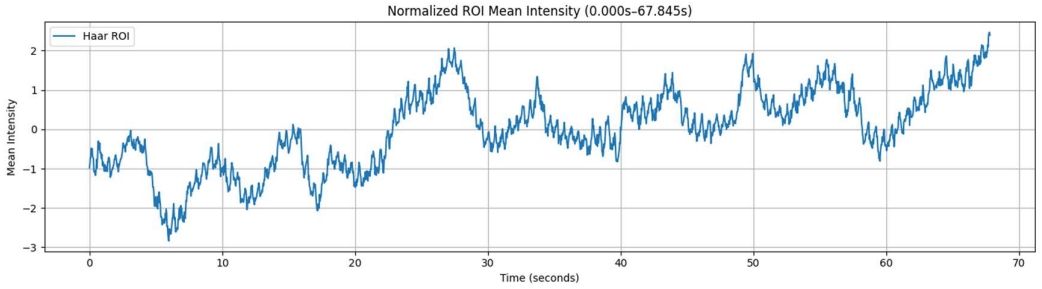
Όπως φαίνεται στον άξονα y των γραφικών παραστάσεων των *εικόνων 8* και *9*, στο σήμα υπάρχει ένας DC παράγοντας ο οποίος πρέπει να αφαιρεθεί. Εφαρμόζεται λοιπόν κανονικοποίηση σύμφωνα με το τύπο:

$$\widetilde{I_{ROI}(n)} = \frac{I_{ROI}(n) - \overline{I_{ROI}}}{\sigma}$$

όπου $\overline{I_{ROI}}$ και σ είναι ο μέσος όρος και η τυπική απόκλιση του σήματος αντίστοιχα πριν την κανονικοποίηση. Μετά την μετατροπή αυτή, η μέση ένταση του καναλιού θα έχει μηδενικό μέσο όρο (άρα καθόλου DC συνιστώσα) και μοναδιαία τυπική απόκλιση, όπως ακριβώς στις *εικόνες 10* και *11*.



Εικόνα 10: Η κανονικοποιήμενη μέση ένταση G καναλιού της χειροκίνητα ορισμένης ROI



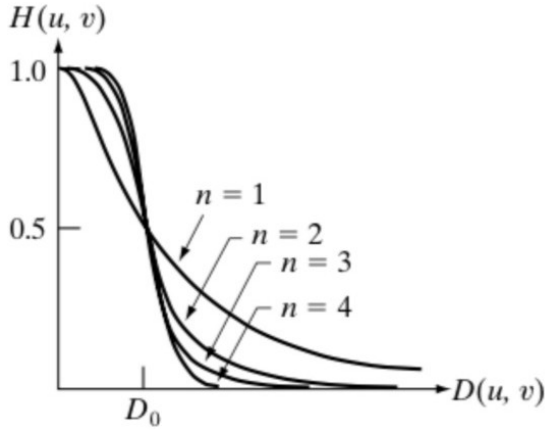
Εικόνα 11: Η κανονικοποιήμενη μέση ένταση G καναλιού της facial detection ορισμένης ROI

4.δ Φιλτράρισμα για απομόνωση ανθρώπινων καρδιακών συχνοτήτων

Τα σήματα των *εικόνων 10* και *11* φαίνεται να μην έχουν καμία ομοιότητα με το PPG σήμα του ground truth (*εικόνα 3.β*). Αυτό συμβαίνει γιατί αποτελούν συνδυασμό και άλλων σημάτων διαφορετικών συχνοτήτων πέρα αυτών που αναζητούνται. Για την εξαγωγή, λοιπόν της τελική κυματομορφής φωτοπληθυσμογραφίας είναι απαραίτητο να εφαρμοστεί ένα ζωνοπερατό φίλτρο το οποίο περιορίζει το σήμα στις ανθρώπινες καρδιακές συχνότητες που ψάχνουμε.

Για το φιλτράρισμα λοιπόν επιλέγεται ένα ζωνοπερατό φίλτρο Butterworth $1^{ης}$ τάξης με κατώτερη και ανώτερη συχνότητα αποκοπής 0.75Hz και 3.5Hz αντίστοιχα. Οι συχνότητες αποκοπής ορίζονται έτσι ώστε να απομονώνονται οι καρδιακές συχνότητες

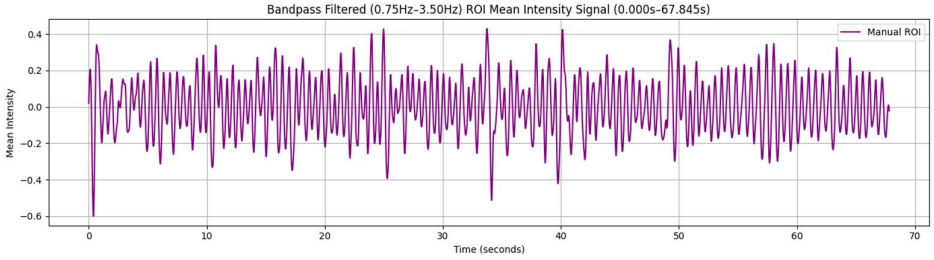
από 45bpm εώς και 210bpm, μια ζώνη αρκετά μεγάλη ώστε να καλύπτει με άνεση όλα τα πιθανά HR. Επιπλέον επιλέχθηκε η χαμηλότερη δυνατή τάξη του φίλτρου, άρα και η λιγότερο απότομη ζώνη μετάβασης (*εικόνα 12*), ώστε να μην υπάρχει το φαινόμενο του κλυδωνισμού.



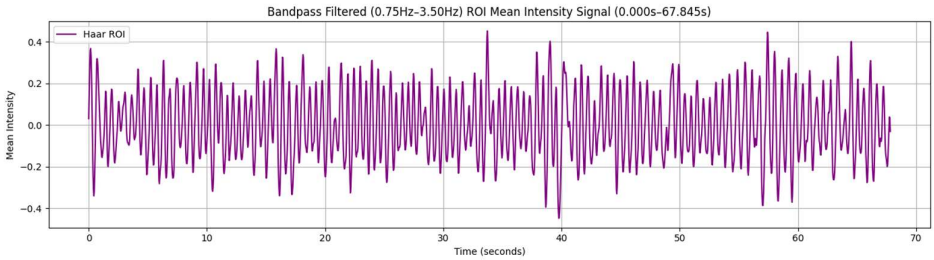
Εικόνα 12: Πλάτος μονοδιάστατου LP φίλτρου Butterworth ανάλογα με την τάξη του n

Πρέπει να σημειωθεί ότι η διάταξη φιλτραρίσματος αποτελείται από 2 από το προαναφερόμενο IIR φίλτρα σειριακά συνδεδεμένα, με τους συντελεστές του ενός να είναι ίδιοι με του άλλου, απλώς με την αντίστροφη σειρά. Αυτό γίνεται για την επίτευξη γραμμικής φάσης και επομένως την αποφυγή του phase distortion.

Εισάγοντας λοιπόν το κανονικοποιήμενο σήμα στην παραπάνω διάταξη, προκύπτει ένα σήμα που πλέον ομοιάζει με PPG σήμα (*εικόνες 13* και *14*), πάνω στο οποίο μπορούμε να εφαρμόσουμε STFT για να εκτιμήσουμε τον καρδιακό ρυθμό.



Εικόνα 13: Η έξοδος του ζωνοπερατού 1ης τάξης φίλτρου Butterworth με εύρος συχνοτήτων 0.75Hz-3.5Hz και είσοδο το κανονικοποιημένο σήμα (Manual ROI)

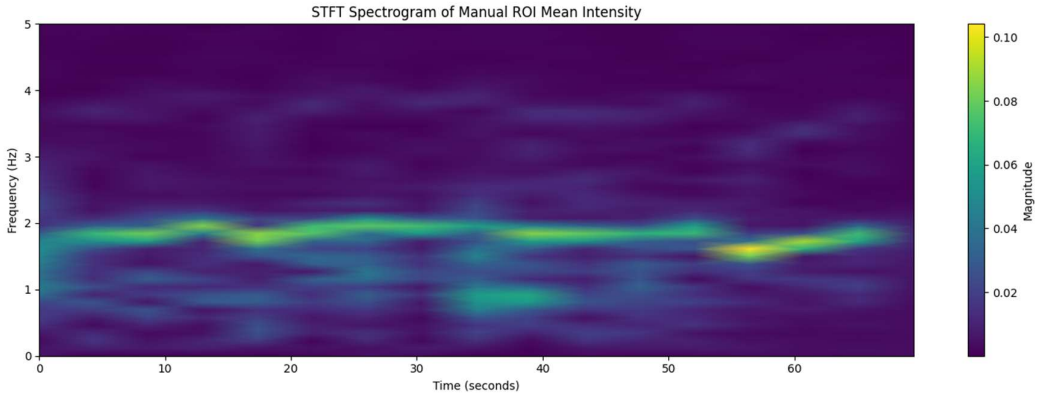


Εικόνα 14: Η έξοδος του ζωνοπερατού 1ης τάξης φίλτρου Butterworth με εύρος συχνοτήτων 0.75Hz-3.5Hz και είσοδο το κανονικοποιημένο σήμα (Face Detection ROI)

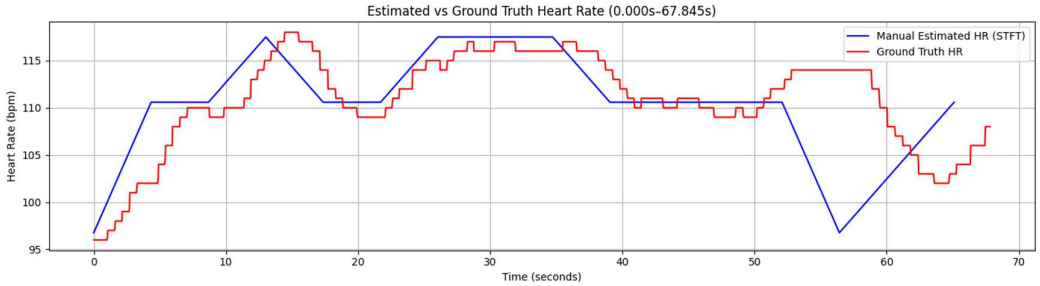
4.ε Εφαρμογή STFT

Σε αυτό το στάδιο έχουμε ήδη εξάγει το PPG σήμα (εικόνες 13 και 14) και μένει μόνο να εντοπίσουμε τις συχνότητες στις οποίες αυτό αντιστοιχεί. Προφανώς το σήμα είναι χρονικά μεταβαλλόμενο ως προς την συχνότητα, αφού το HR αλλάζει με την πάροδο του χρόνου, και επομένως δεν είναι ορθό να εφαρμόσουμε Fast Fourier Transform (FFT) σε όλη την κυματομορφή και να εξάγουμε την κυρίαρχη συχνότητα. Αυτό που πρέπει να εφαρμοστεί στην συγκεκριμένη περίπτωση είναι Short-Time Fourier Transform (STFT), ο οποίος χωρίζει τον χρόνο σε μικρότερα ισομεγέθη και αλληλεπικαλυπτόμενα διαστήματα, στο καθένα από τα οποία εφαρμόζεται ξεχωριστά ο FFT. Η πιο έντονη συχνότητα του κάθε παραθύρου αντιστοιχισμένη σε bpm θεωρείται ως το HR στο αντίστοιχο χρονικό διάστημα, ολοκληρώνοντας έτσι την διαδικασία εκτίμησης του καρδιακού ρυθμού.

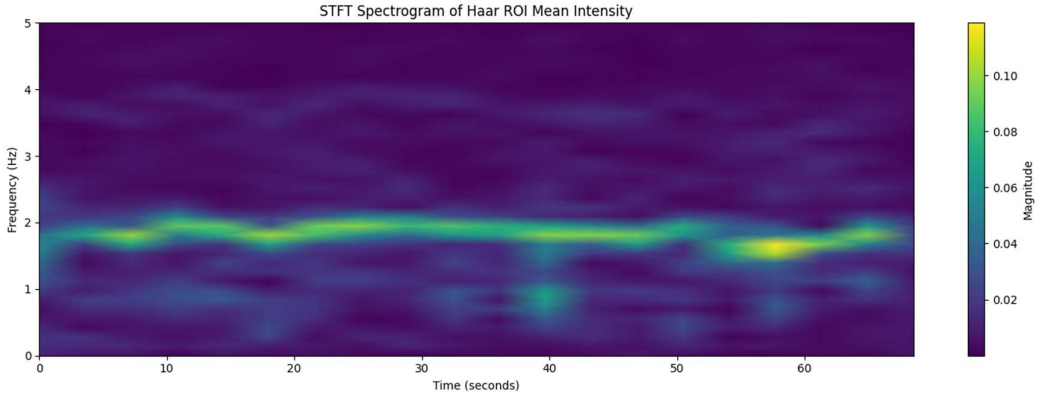
Τόσο για τον STFT του PPG σήματος που αντιστοιχεί στην χειροκίνητα ορισμένη ROI, όσο και για εκείνον που αντιστοιχεί στην ROI από ανίχνευση προσώπου, ορίζεται ποσοστό αλληλοεπικάλυψης παραθύρων 50%, ενώ τα μεγέθη τους επιλέγονται να είναι ίσα με 260 και 215 frames αντίστοιχα. Η εκτέλεση με αυτές τις παραμέτρους οδηγεί στα Spectograms των *εικόνων 15* και *17*. Τα spectograms είναι γραφικές παραστάσεις στις οποίες προβάλλεται το πόσο έντονη είναι η κάθε συχνότητα μέσα στο σήμα μια δεδομένη χρονική στιγμή. Εξάγοντας από αυτά τα διαγράμματα την μέγιστη σε εμφάνιση συχνότητα της κάθε χρονικής στιγμή, προκύπτουν οι καμπύλες των εκτιμημένων HR οι οποίες μπορούν και να συγκριθούν με το αντίστοιχο Ground Truth, όπως φαίνεται στις *εικόνες 16* και *18*.



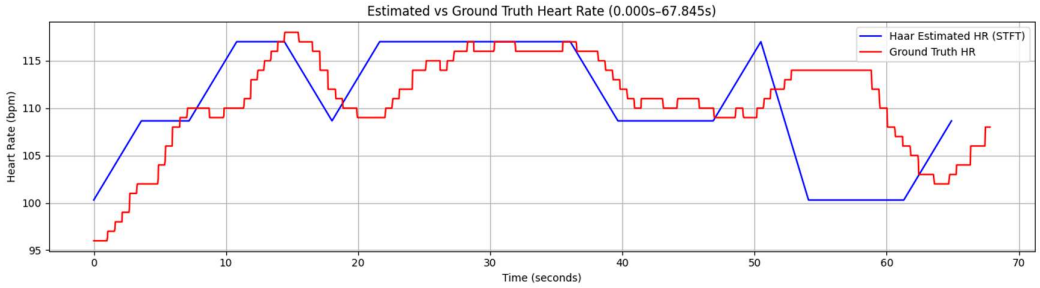
Εικόνα 15: Το spectrogram που προκύπτει από STFT του PPG σήματος (Manual ROI)



Εικόνα 16: Η εκτιμώμενη HR (manual ROI) σε με το HR του ground truth



Εικόνα 17: Το spectrogram που προκύπτει από STFT του PPG σήματος (Facial Detection ROI)



Εικόνα 18: Η εκτιμώμενη HR (Facial Detection ROI) σε σύγκριση με το HR του ground truth

5. Σχολιασμός Αποτελεσμάτων

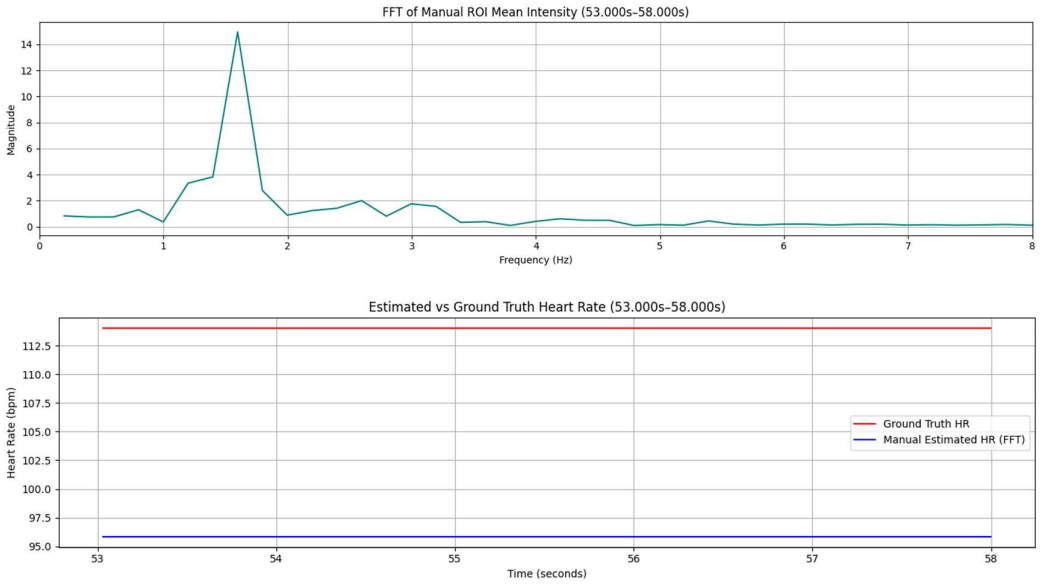
Αρχικά, και οι δύο μέθοδοι οδήγησαν σε αρκετά καλές εκτιμήσεις. Μπορεί τα τελικά σήματα να μην ακολουθούν τις τοπικές και σύντομες αυξομειώσεις του ground truth, αλλά καταφέρνουν με επιτυχία να ακολουθήσουν την γενικότερη πορεία του. Γενικά υπάρχουν μόνο δύο αποκλίσεις που μπορεί δημιουργούν απορία σχετικά με την αποτελεσματικότητα της αλγορίθμου.

Η πρώτη αφορά την χρονική απόκλιση του εκτιμώμενου HR. Είναι φανερό ότι και στις δύο αντίστοιχες *εικόνες 16* και *18,* το τελικό σήμα προηγείται εκείνου του Ground Truth. Αυτό ωστόσο οφείλεται στον χρόνο αντίδρασης του οργάνου μέτρησης, δηλαδή του οξύμετρου. Πιο συγκεκριμένα, ενώ ο αλγόριθμος έχει πρόσβαση σε ολόκληρο το σήμα της φωτοπληθυσμογραφίας και εξάγει τη καρδιακή συχνότητα κάθε χρονικής στιγμής σε δεύτερο χρόνο, το οξύμετρο πρέπει να προβάλει τις εκτιμήσεις του σε πραγματικό χρόνο. Αυτό συνεπάγεται σε κάποιες χρονικές καθυστερήσεις αναγκαίες για την εκτέλεση των απαραίτητων υπολογισμών και έτσι σε μία πρόβλεψη που αντιστοιχεί σε μία παλιότερη στιγμή. Επομένως, η χρονική απόκλιση της εκτίμησης του αλγορίθμου με του οξύμετρου, δεν οφείλεται σε καμία αστοχία και μπορεί να αγνοηθεί, εφόσον βέβαια η μετατόπιση του Ground Truth προς τα δεξιά οδηγεί σε αλληλοεπικάλυψη των δύο γραφικών παραστάσεων. Αυτό στο μεγαλύτερο μέρος φαίνεται να ισχύει, με εξαίρεση το χρονικό διάστημα 53s-58s περίπου, το οποίο αποτελεί και την δεύτερη από τις δυο προς εξέταση αστοχίες.

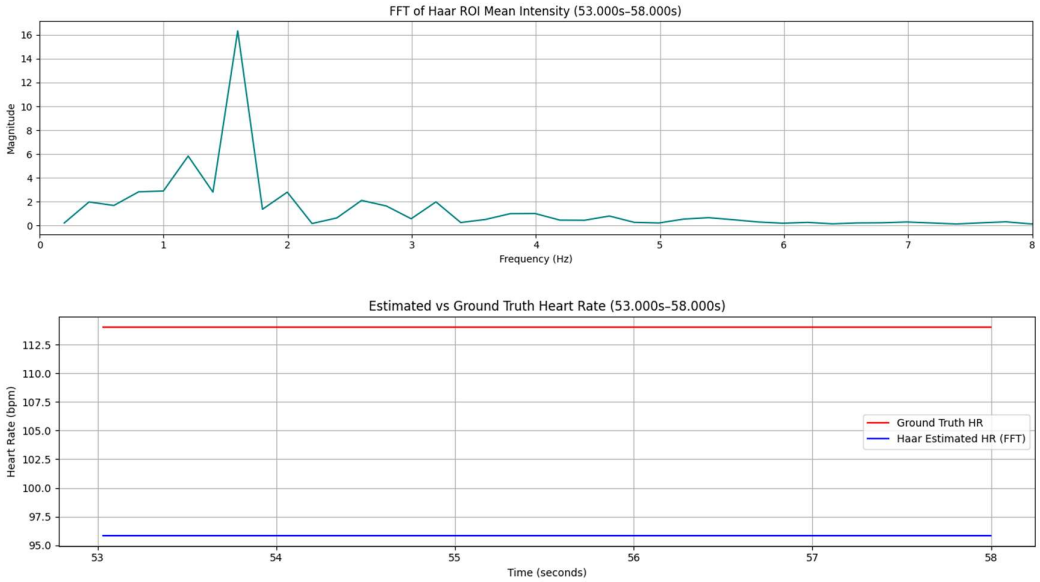
Στο συγκεκριμένο διάστημα οι τιμές του HR που υπολόγισε ο αλγόριθμος διαφέρουν σημαντικά από την καρδιακή συχνότητα του Ground truth, η οποία είναι σταθερή στα 114bpm. Εκεί ο αλγόριθμος εκτιμάει HR της τάξεως των 96bpm με 100bpm και πράγματι, αν εφαρμοστεί FFT σε αυτό το διάστημα (σύμφωνα με το Ground Truth, τότε το σήμα είναι μη χρονικά μεταβαλλόμενο και άρα μπορεί να εφαρμοστεί FFT) προκύπτει και για τις δύο μεθόδους ότι η καρδιακή συχνότητα θα έπρεπε να είναι στα 96bpm περίπου (*εικόνες 19* και *20*). Αυτό μπορεί κάποιος εύλογα να το αποδώσει σε κάποια αστοχία/αδυναμία του αλγορίθμου. Ωστόσο, μια πιο προσεκτική ματιά στο PPG σήμα του Ground Truth οδηγεί σε μια ενδιαφέρουσα παρατήρηση.

Πιο συγκεκριμένα, υπάρχει ο όρος του στιγμιαίου Heart Rate, το οποίο υπολογίζεται για κάθε κορυφή του PPG σήματος μετρώντας την οριζόντια απόστασή από το προηγούμενο peak. Εάν αυτή η χρονική απόκλιση είναι ίση με Δt, τότε το στιγμιαίο HR σε bpm είναι:

$$Instantaneous\ HR = \frac{60}{\Delta t}$$



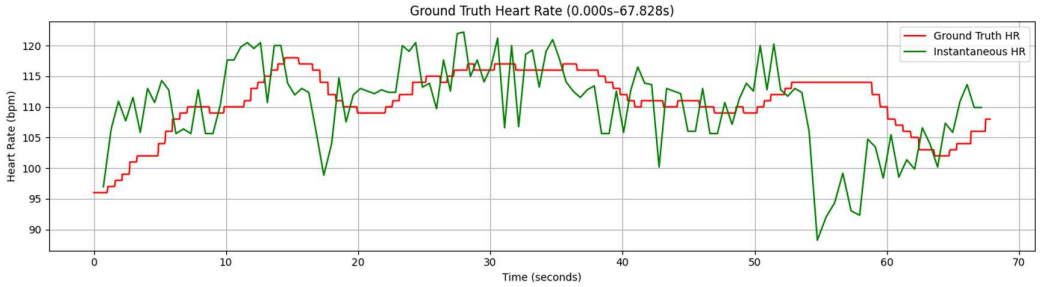
Εικόνα 19: Αποτέλεσμα FFT (53sec-58sec) στο PPG σήμα της Manual ROI και απόκλιση από το HR του Ground Truth



Εικόνα 20: Αποτέλεσμα FFT (53sec-58sec) στο PPG σήμα της Facial Detection ROI και απόκλιση από το HR του Ground Truth

Η εξέταση αυτού του μεγέθους μόνο για ένα ζευγάρι κορυφών δεν είναι τίποτα άλλο από μια ανακριβή εκτίμηση χωρίς κάποια πρακτική αξία. Όμως, εάν υπολογιστούν όλες οι στιγμιαίες καρδιακές συχνότητες ενός PPG και εξεταστούν μαζί ως σύνολο, όπως συμβαίνει και με το PPG του Ground Truth στην *εικόνα 21*, τότε πάλι προκύπτει μία ανακριβής εκτίμηση, αλλά αυτή φαίνεται να ακολουθεί μια συγκεκριμένη πορεία. Με λίγα λόγια, η καμπύλη των IHR περιβάλλει την καμπύλη των HR του Ground Truth,

με εξαίρεση για ακόμη μία φορά το διάστημα 53sec-58sec, στο οποίο πέφτει το IHR σε τιμές ίδιας τάξεως με τις εκτιμήσεις. Επομένως, η ανακάλυψη αυτή σε συνδυασμό με τα αποτελέσματα του FFT είναι αρκετά ώστε να δημιουργήσουν αμφιβολίες περί της ορθότητας είτε του HR είτε του PPG σήματος του Ground Truth στο συγκεκριμένο χρονικό διάστημα.



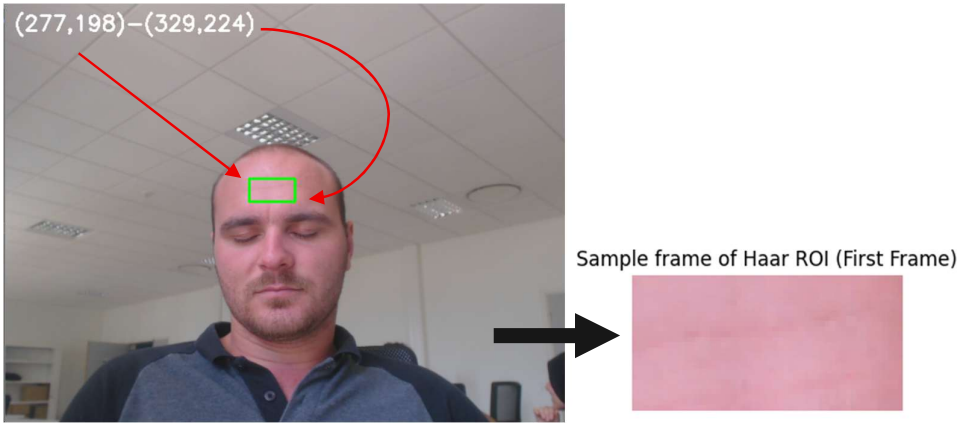
Εικόνα 21: Η καμπύλη των στιγμιαίων HR (πράσινη) σε σύγκριση με το HR του Ground Truth (Κόκκινη)

Έχοντας δικαιολογήσει τις δύο αστοχίες μένει μόνο η σύγκριση των επιδόσεων των δύο μεθόδων ορισμού ROI του αλγορίθμου. Παρατηρώντας τις *εικόνες 16* και *18* είναι φανερό ότι και οι δύο μέθοδοι οδηγούν σε αρκετά καλές εκτιμήσεις, με εκείνη της ROI ορισμένης από το facial detection να αποδίδει λίγο καλύτερα από την άλλη. Εκτός από το ότι η μέθοδος της manual ROI δεν ακολούθησε σωστά όλη την πορεία της Ground truth HR καμπύλης, αφού αστοχεί περίπου στα 50sec, χρειάζεται και μεγαλύτερο μέγεθος παραθύρου (260 frames, ενώ η άλλη χρειάζεται 215) για να οδηγήσει σε ικανοποιητικά αποτελέσματα. Μεγάλο μέγεθος χρονικού παραθύρου, σημαίνει χαμηλότερη ανάλυση (resolution) στο πεδίο του χρόνου (άρα μεγάλο time step) και έτσι η εκτίμηση δυσκολεύεται να ακολουθήσει την πραγματική καμπύλη του HR τις σωστές στιγμές. Επομένως, η μέθοδος που έχει το ίδιο καλή εκτίμηση αλλά για μικρότερο μέγεθος παραθύρου κρίνεται ως η πιο αποτελεσματική.

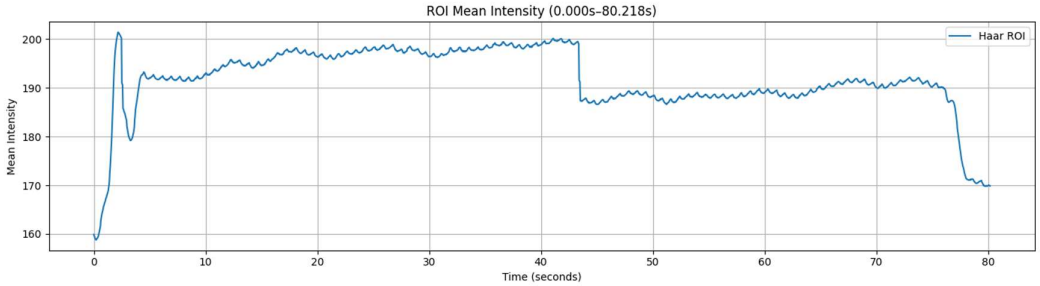
6. Υποθέσεις και περιορισμοί

Αν και ο συγκεκριμένος αλγόριθμος κάνει μία αρκετά ικανοποιητική εκτίμηση για την απλότητα του, έχει κάποιες σημαντικές αδυναμίες. Μια από αυτές και η σημαντικότερη είναι η ευαισθησία του στις συνθήκες φωτισμού. Για να λειτουργήσει σωστά, δουλεύει με την υπόθεση ότι καθόλη την διάρκεια του βίντεο οι πηγές φωτισμού παραμένουν σταθερές, τόσο ως προς το σημείο τους στον χώρο, όσο και ως προς το πλήθος και την ένταση τους. Εάν αλλάξει οτιδήποτε από αυτά, τότε οι εκτιμήσεις στο διάστημα αυτών των αλλαγών μπορεί να αποτυγχάνουν εώς και καταστροφικά.

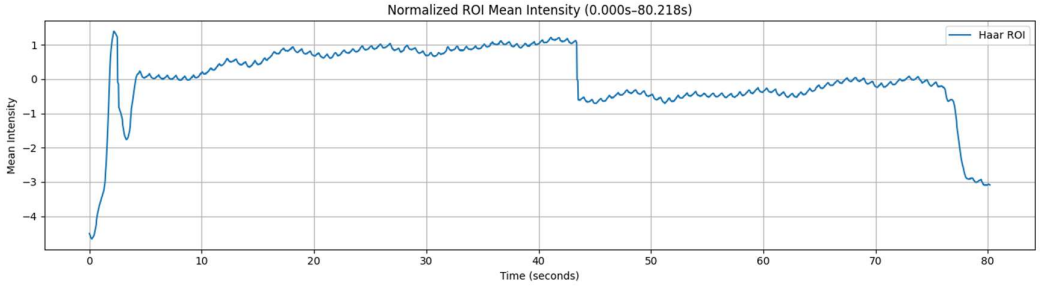
Χαρακτηριστικό παράδειγμα της αδυναμίας αυτής αποτελεί το βίντεο του υποκειμένου 11 στο Dataset 1. Σε αυτή την καταγραφή εντοπίζονται 3 αλλαγές φωτισμού· η μία είναι στην αρχή (0 εώς 5 sec), η άλλη στα 43-44sec περίπου και η τελευταία στο τέλος του βίντεο από τα 76 δευτερόλεπτα περίπου και μετά. Εάν εκτελεστεί ο αλγόριθμος για όλη την διάρκεια του βίντεο και για χειροκίνητα ορισμένη ROI ακριβώς όπως στην *εικόνα 22* (δεν υπάρχει κανένα πρόβλημα να χρησιμοποιηθεί η manual ROI, γιατί το υποκείμενο παραμένει ακίνητο), τότε προκύπτουν οι γραφικές παραστάσεις των *εικόνων 23-27*. Παρατηρούμε ότι οι τρεις αλλαγές φωτισμού είχαν τεράστιο αντίκτυπο στην μέση ένταση καναλιού (*εικόνα 23*), το οποίο δεν μπόρεσε να αφαιρεθεί πλήρως ούτε με το ζωνοπερατό φιλτράρισμα (*εικόνα 25*). Εν τέλη η επίδραση αυτή οδήγησε σε σημαντικά εσφαλμένες εκτιμήσεις στην αρχή (*εικόνα 27*), αλλά όχι στην μέση και στο τέλος. Ως προς την μεσαία μεταβολή φωτεινότητας, όντως αυτή δεν μπόρεσε να υπερισχύσει έναντι των άλλων συχνοτήτων, αλλά η τελική δεν επικράτησε λόγω του time window του STFT. Δηλαδή, το παράθυρο είναι τόσο μεγάλο, ώστε αυτή η συχνότητα να μην μπορεί να υπερισχύσει έναντι των πραγματικών, το οποίο αν και βοηθάει στην συγκεκριμένη περίπτωση, χαλάει σημαντικά το time resolution. Από τα παραπάνω λοιπόν βγαίνει το συμπέρασμα ότι, προκειμένου να χρησιμοποιηθεί αυτός αλγόριθμος, πρέπει να ισχύει ο περιορισμός του να μην γίνεται καμία αλλαγή στη φωτεινότητα του περιβάλλοντος.



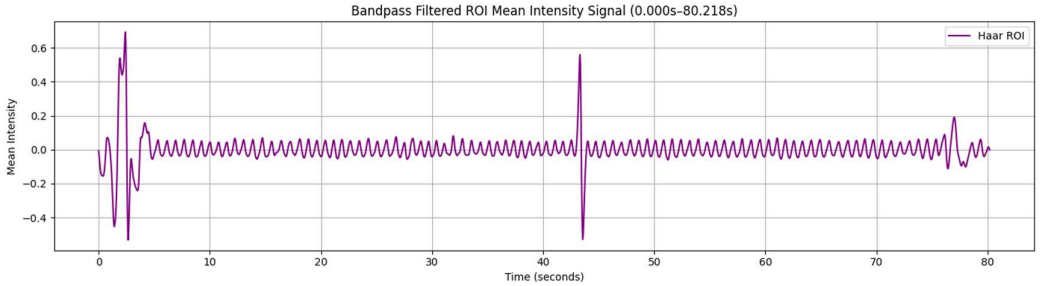
Εικόνα 22: Η χειροκίνητα ορισμένη ROI για όλα τα frames του subject 11 και το αποτέλεσμα στο πρώτο frame



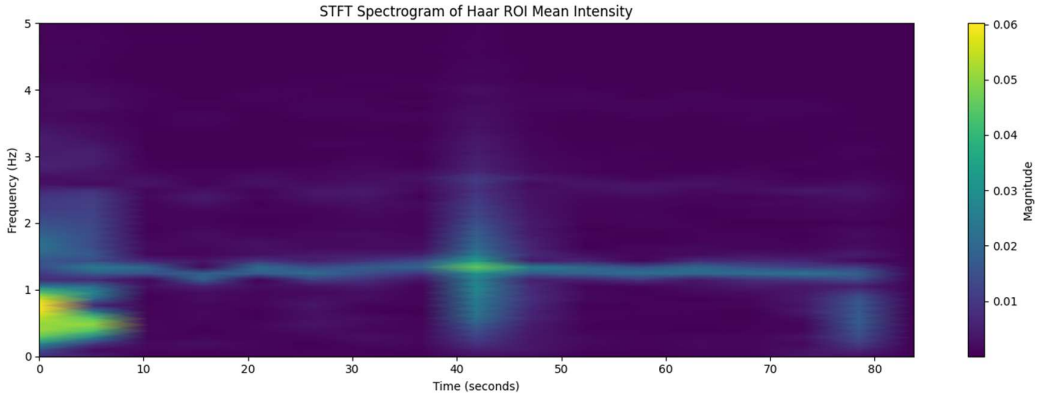
Εικόνα 23: Το σήμα που προκύπτει μετά από Spatial Averaging του G καναλιού της χειροκίνητα ορισμένης ROI. Οι τρεις απότομες μεταβολές οφείλονται στις αλλαγές φωτισμού του αντίστοιχου βίντεο.



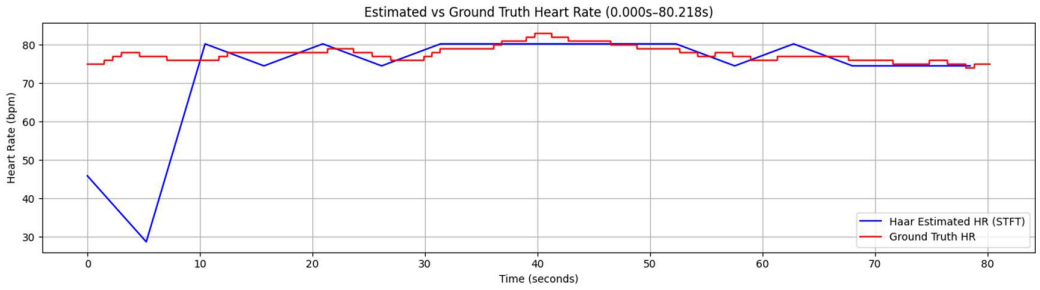
Εικόνα 24: Η κανονικοποιήμενη μέση ένταση G καναλιού της χειροκίνητα ορισμένης ROI. Οι τρεις απότομες μεταβολές οφείλονται στις αλλαγές φωτισμού του αντίστοιχου βίντεο.



Εικόνα 25: Η έξοδος του ζωνοπερατού $1^{ης}$ τάξης φίλτρου Butterworth με εύρος συχνοτήτων 0.75Hz-3.5Hz και είσοδο το κανονικοποιημένο σήμα (Manual ROI). Οι αλλαγές φωτισμού δεν κατάφεραν να απομακρυνθούν.



Εικόνα 26: Το spectrogram που προκύπτει από STFT στο PPG σήμα (Manual ROI). Οι συχνότητες των αλλαγών φωτισμού του βίντεο φαίνεται να κατέχουν σημαντικές θέσεις στα αντίστοιχα διαστήματα.



Εικόνα 27: Η εκτιμώμενη HR (manual ROI) συγκριτικά με το HR του ground truth. Η αλλαγή φωτισμού στην αρχή επηρεάζει σημαντικά την εκτίμηση εκείνο το χρονικό διάστημα.

Ένας άλλος περιορισμός ο οποίος προαναφέρθηκε και στο *κεφάλαιο 4.α* είναι το ότι το υποκείμενο πρέπει να παραμένει όσο πιο ακίνητο γίνεται. Πολύ μικρές κινήσεις δεν έχουν τεράστια επίδραση στο τελικό αποτέλεσμα, ειδικά στην περίπτωση χρήσης face detection, αλλά μεγαλύτερες οδηγούν σε αλλαγές της φωτεινότητας πάνω στο μέτωπο και επομένως ακολουθούν τα ίδια που συμβαίνουν και στην περίπτωση της προηγούμενης αστοχίας. Οι τεχνικές που θα μπορούσαν να αντικαταστήσουν την παρούσα, ώστε να εξαλειφθούν στο μεγαλύτερο μέρος τους οι περιορισμοί, τόσο της σταθερής φωτεινότητας, όσο και της μηδενικής κίνησης, είναι οι τεχνικές CHROM και POS, καθώς διαθέτουν αυξημένη ανοχή στον θόρυβο.

Τέλος, μια ακόμη αδυναμία του αλγορίθμου, οφείλεται στην επιλογή εύρεσης συχνοτήτων με STFT. Πιο συγκεκριμένα, ο καρδιακός ρυθμός είναι ένα σήμα του οποίου η τιμή (δηλαδή η συχνότητα) αλλάζει πολύ συχνά, με πολύ μικρό βήμα και σε σχετικά σύντομο χρονικό διάστημα. Ένα τέτοιο σήμα αντιμετωπίζει μεγάλα προβλήματα ως προς την ανάλυση συχνοτήτων με STFT, λόγω της αρχής αβεβαιότητας. Σύμφωνα με αυτήν, όσο μεγαλύτερο είναι το time resolution Δt, δηλαδή όσο μικρότερο είναι το time window, τόσο μικρότερο είναι το frequency resolution Δf που συνεπάγεται σε μεγαλύτερη απόσταση μεταξύ 2 διαδοχικών διακριτών τιμών της συχνότητας. Αντίστροφα, όσο μεγαλύτερη είναι ανάλυση της συχνότητας και άρα όσο πιο κοντά είναι δύο διαδοχικές τιμές τις συχνότητας, τόσο μικρότερη είναι η ανάλυση του χρόνο (μεγάλο time window/χρονική ασάφεια). Αυτή η σχέση περιγράφεται από τον τύπο:

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi}$$

και δημιουργεί μεγάλα ζητήματα στην ρύθμιση των παραμέτρων του STFT, καθώς μας ενδιαφέρει και η ανάλυση ως προς τον χρόνο και ως προς την συχνότητα. Εν τέλει, όπως φαίνεται και στις *εικόνες 16*, *18* και *27*, δίνεται μεγαλύτερο βάρος στην ανάλυση συχνότητας, έτσι ώστε να προβάλλονται οι σωστές τιμές του Heart Rate. Όμως σε περιπτώσεις καρδιακών σημάτων, όπως του subject 11 που το HR μεταβάλλεται πολύ συχνά αλλά με πολύ μικρή τυπική απόκλιση, το time resolution χειροτερεύει ακόμη περισσότερο, προκειμένου να ομοιάζει έστω και λίγο το εκτιμώμενο HR με το πραγματικό.

Λύση εώς έναν βαθμό σε αυτό το πρόβλημα θα μπορούσε να αποτελεί η ανάλυση συχνοτήτων με Continuous Wavelet Transform (CWT). Αυτός ο μετασχηματισμός προσαρμόζει δυναμικά την frequency και time resolution, ανάλογα με το πόσο μεγάλη είναι η συχνότητα του σήματος εκείνο το χρονικό διάστημα, σε αντίθεση με τον STFT που ορίζει μια σταθερή ανάλυση για όλο το σήμα. Γενικά αυτή η τεχνική είναι πιο ικανή στο να ακολουθεί τις σχετικά γρήγορες αλλαγές συχνότητες, καθιστώντας την πιο αποτελεσματική στην εξέταση των φυσικών/καρδιακών σημάτων.

7. Αναφορές

[1] G. de Haan and V. Jeanne, "Robust Pulse Rate From Chrominance-Based rPPG," in *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878-2886, Oct. 2013, doi: 10.1109/TBME.2013.2266196.

[2] W. Wang, A. C. den Brinker, S. Stuijk and G. de Haan, "Algorithmic Principles of Remote PPG," in *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479-1491, July 2017, doi: 10.1109/TBME.2016.2609282.

[3] S. Bobbia, G. Benezeth, M. Mansouri, and C. Dubois, Feb. 6, 2019, UBFC-rPPG, ver. 1.0, Université de Bourgogne Franche-Comté. [Online]. Available: https://sites.google.com/view/ybenezeth/ubfcrppg

[4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 2001, pp. I-I, doi: 10.1109/CVPR.2001.990517.

[5] Verkruysse W, Svaasand LO, Nelson JS. Remote plethysmographic imaging using ambient light. Opt Express. 2008 Dec 22;16(26):21434-45. doi: 10.1364/oe.16.021434. PMID: 19104573; PMCID: PMC2717852.

[6] Berggrem, A., Berggrem J. (2019) Non-contact measurement of heart rate using a camera, [Master's thesis, Lund University]. lup.lub.lu.se. https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=8972235&fileOId=8972240

[7] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography", Pattern Recognition Letters, 2017.