

Rilevazione remota della frequenza cardiaca e di arousal tramite video

Francesco Scarlata



Fig. 1: Esempio di uso di un ecg



Fig. 2: Un pulsossimetro da polso con pletismografo

Abstract—In questo progetto è stato sviluppato un sistema per il rilevamento della frequenza cardiaca su un soggetto sia ripreso da video che da webcam. L' algoritmo è basato sull'articolo [1] per il calcolo del segnale hrv.

Viene poi fatto il confronto dei risultati con altri metodi già sviluppati nel progetto [2] per la rilevazione della frequenza cardiaca. Per confrontare i risultati, vengono utilizzati dei video in cui il vero battito cardiaco è già stato monitorato. Il confronto viene fatto su tutta la serie come la distanza media della serie calcolata con quella conosciuta.

Infine, viene implementato il metodo per determinare una feature di arousal derivabile dallo spettro di potenza del segnale, descritto in [1]. Qui viene calcolata una versione temporale dinamica prendendo delle finestre temporali successive.

1 INTRODUZIONE

La frequenza cardiaca ("heart rate" o HR) é definita come il numero di occorrenze di battiti cardiaci in uno specifico periodo di tempo, espresso in battiti per minuto. Anche se le occorrenze di battiti cardiaci possono essere innescati da impulsi elettrici generati all'interno del nodo sinotriale (SA), la vera frequenza dell'attività del cuore é per la maggior parte modulata dal sistema nervoso automatico (ANS). Questa regolazione neurale causa variabilità nel HR sia nello stato attivo che a riposo. Inoltre, la variabilità dipende non solo dallo stato psicologico dell'individuo, ma anche dall'età e dalla salute (e malattie).

Al giorno d'oggi, le tecniche standard per la misurazione dell' HR richiedono ai pazienti di indossare sul proprio corpo apparecchiature fisiche. Questa é una cosa che può essere invasiva oltre che avere un costo elevato. Tradizionalmente, il battito cardiaco può essere acquisito tramite strumenti più o meno invasivi, come l'elettrocardiogramma (ECG, figura 1) in cui al soggetto devono essere applicati degli elettrodi sulla pelle (con possibile irritazione della pelle del paziente), un pulsossimetro (figura 2) che viene applicato alle dita del paziente oppure degli smartwatch e altri dispositivi comodi da usare ma che compensano la non invasività con dei costi molto alti. L'invasività tuttavia in alcuni campi può dare delle problematiche legate alla validità e accuratezza della misurazione. Basti pensare alle misure di valutazione fisiologiche per la game user research (nel campo chiamate "biometrics") per il playtesting di videogames [3]. In questi casi valutare il battito cardiaco accurato può risultare difficile a causa della misurazione invasiva.

La variabilità della frequenza cardiaca ("Heart Rate Variability" o HRV) é definita come la variazione temporale tra sequenze di intervalli di battiti cardiaci. Questa é un'altra feature che viene utilizzata nel campo quantitativo della game user research, e che viene sempre più usata sia in campo medico che psicologico poiché viene utilizzato per determinare stress, riconoscere emozioni e in generale per riconoscere cambiamenti nell'ANS. L'HRV infatti é stato riconosciuto come marcatore riflettente l'attività delle componenti dell'ANS sul nodo seno-atriale del cuore. Anche questo parametro dipende da diversi fattori, come il sesso, l'età, la condizione sia fisica che psicologica della persona, interferenze causate dall'assunzione di droghe etc.

Sempre nell'articolo [1] si parla di una feature per

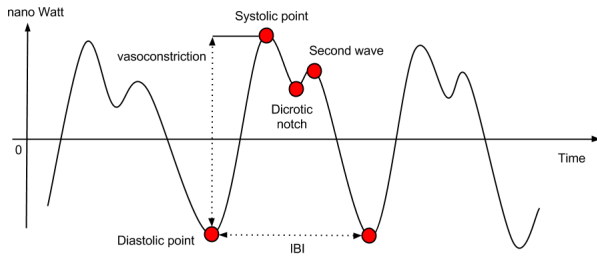


Fig. 3: segnale PPG ideale senza rumore

monitorare l'arousal che prese le very low frequency, low frequency e high frequency, potrebbero consentire di dare qualche informazione in più sullo stato emotivo. Tuttavia si tiene a premettere che questa feature non ha un valore di scala assoluto per ogni persona, ma è una feature che viene calcolata su almeno due sedute: la prima come sessione neutrale sul soggetto, che consente di avere un valore su cui normalizzare, e la successiva in cui si calcola il valore normalizzandolo con quello determinato nella sessione neutrale. Si potrebbe quindi monitorare l'arousal di un paziente nel caso in cui si conoscesse il suo valore durante lo stato neutrale.

Questo progetto si può quindi suddividere in tre parti:

- 1) Lo sviluppo di un sistema per il rilevamento dell'HR di un soggetto sia da video che ripreso da webcam tramite determinazione del segnale rPPG e successivo calcolo del segnale rHRV, il cui sistema si basa sull'algoritmo dell'articolo [1].
- 2) Sviluppo di un sistema per il confronto tra metodi per il rilevamento dell'HR. In questo sistema sono stati confrontati il metodo HRV descritto e i metodi "green" e "pca" sviluppati nel progetto [2] ed adattati per il rilevamento da video.
- 3) Sviluppo del sistema per rilevare la feature dell'arousal descritto nell'articolo [1] e il calcolo di questa feature con delle finestre temporali all'interno dello stesso video per determinare la variazione di arousal nel soggetto.

I contributi di questo progetto sono i seguenti:

- Valutazione dei risultati ottenuti dall'approccio hrv. Gli hr rilevati sono poi stati confrontati con i valori rilevati durante la collezione del dataset utilizzato.
- Confronto dei risultati ottenuti dai diversi approcci.
- Valutazione della feature di arousal su finestre successive.

2 ANALISI DELLO STATO DELL'ARTE

Per la misurazione dell'heart rate variability, lo strumento standard è l'elettrocardiogramma, tuttavia diversi studi sono stati fatti in maniera soddisfacente usando la fotoplethysmografia (PPG), e l'articolo [1] si basa appunto su questi studi. Il PPG è una tecnica usata per individuare la variazione di volume di sangue microvascolare nei tessuti ed è una tecnologia che richiede semplicemente una sorgente di luce e un rilevatore fotoelettrico. Tuttavia, il

problema di questa tecnica è che l'onda del segnale del battito deve essere molto precisa per stimare l'HRV, altrimenti può portare all'individuazione di falsi picchi che a sua volta portano ad una stima dell'HRV molto scarsa.

In figura 3 si mostra il segnale PPG ideale senza rumore. Per rPPG (remote PPG) s'intende il segnale del PPG rilevato in maniera remota, cioè senza contatto.

Ci sono altri metodi che invece non utilizzano il segnale PPG per il calcolo dell'hr, come i metodi ICA e i metodi PCA, tuttavia per calcolare l'hrv è necessario il segnale PPG o l'elettrocardiogramma, quindi quei metodi danno una soluzione solo parziale.

Nell'articolo [1] viene quindi sfruttata l'idea di utilizzare dei dati video per stimare i segnali fisiologici, quantificando le piccole variazioni di luce riflessa dalla pelle dovuta alle variazioni di perfusione. In sintesi, l'algoritmo citato procede nel seguente modo:

- 1) Viene localizzato il volto del soggetto su video (fase di face detection)
- 2) Viene localizzata la pelle all'interno del volto. Questa sarà la regione di interesse o ROI. (fase di skin detection)
- 3) La regione di interesse viene separata nei tre canali RGB
- 4) Per ogni canale, si trova la media spaziale sulla regione
- 5) Per ogni canale concatena le medie trovate (si creano così delle tracce temporali)
- 6) Le tracce vengono pre-processate tramite zero mean, viene fatto un detrend usando un smoothness priors approach (spa) [4] e filtrate con un filtro passabanda di Butterworth con frequenze 0.7 e 3.5 Hz.
- 7) Viene estratto il segnale rPPG usando il metodo "CHROM" ("chrominance based-method", un metodo basato sulla cromaticanza) tramite la combinazione lineare delle tracce RGB create nei passi precedenti. Siano quindi $y^c(t)$ le tracce temporali ottenute dopo il pre-processing dove $c \in \{R, G, B\}$ è il colore del canale. Formeremo quindi i seguenti due vettori ortogonali tra loro:

$$\begin{aligned} X(t) &= 3y^R(t) - 2y^G(t) \\ Y(t) &= 1.5y^R(t) + y^G(t) - 1.5y^B(t) \end{aligned} \quad (1)$$

- 8) il segnale rPPG S viene calcolato come $S(t) = X(t) - \alpha Y(t)$ dove $\alpha = \sigma(X)/\sigma(Y)$ e $\sigma(\cdot)$ è la deviazione standard.
- 9) Vengono individuati i picchi del segnale rPPG con un algoritmo apposito descritto nell'articolo [5] (con una modifica che verrà descritta nella sezione 4.3) per calcolare il segnale rHRV.
- 10) resampling del segnale rHRV e calcolo del periodogramma
- 11) calcolo dei valori di "very low frequency"(vlf), "low frequency"(lf) e "high frequency" (hf) dal periodogramma
- 12) calcolo della feature di arousal come $A = \frac{vlf+lf}{hf}$

3 MODELLO TEORICO

3.1 Filtro di Butterworth

Durante l'operazione di pre-processing é stato utilizzato il filtro passabanda di Butterworth. Un filtro di Butterworth é un tipo di filtro di elaborazione del segnale progettato per avere una risposta in frequenza il piú piatta possibile nella banda passante. Lo scopo di questo filtro é quello di fare passare le frequenze volute e di rigettare (o attenuare) quelle non volute, molto utile quando si fa fronte a del rumore e problematiche di questo tipo. Gli elementi di input per questo tipo di filtro sono i seguenti:

- il segnale che si vuole filtrare
- le cutoff frequency, cioè le soglie delle frequenze da rigettare. Per esempio, nel caso del filtro low-pass significa che il filtro farà passare solo le frequenze al di sotto di quella cutoff frequency.
- ordine del filtro. Più l'ordine é alto e piú l'attenuazione porterá a zero velocemente.

In figura 4 si può notare un esempio di filtro passabanda di Butterworth con diversi valori per l'ordine del filtro.

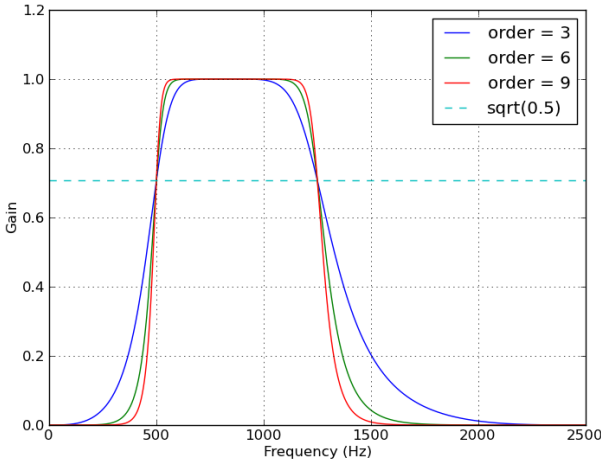


Fig. 4: Esempio Filtro passabanda di Butterworth su diversi ordini

3.2 Richiami matematici

3.2.1 Distanza tra due punti in una dimensione

Poiché bisognava fare un confronto tra serie di dati rispetto a una serie comune, é stato necessario utilizzare il concetto di distanza.

La definizione di distanza euclidiana tra i due punti x e y é la lunghezza del segmento che li connette \overline{xy} . In coordinate cartesiane se $x = (x_1, x_2, \dots, x_n)$ e $y = (y_1, y_2, \dots, y_n)$ sono due punti nello spazio euclideo, allora la distanza (d) da x a y o da y a x é data dalla formula pitagorica:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

In una dimensione, la distanza si calcola fissando due punti su una linea e scegliendone uno come origine. La distanza

tra qualsiasi due punti in una linea reale é quindi il valore assoluto della differenza numerica delle loro coordinate. Quindi, se x e y sono due punti sulla retta reale, la distanza tra loro é data come:

$$d(x, y) = \sqrt{(q - p)^2} = |q - p| \quad (3)$$

Per due serie di dati X e Y , essendo composti da elementi dipendenti si é preferito optare per l'approccio della distanza media (\bar{d}) tra valori corrispondenti, arrivando alla seguente formula:

$$\overline{d(x, y)} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (4)$$

3.2.2 Correlazione tra dati

Poiché parleremo di correlazione a livello qualitativo tra la serie hr e i valori di arousal, definiamo brevemente cos'è la correlazione e l'indice di correlazione.

In statistica per correlazione si intende una relazione tra due variabili statistiche tale che a ciascun valore della prima variabile corrisponda con una "certa regolarità" un valore della seconda.

Il grado di correlazione fra due variabili viene espresso mediante i cosiddetti indici di correlazione. Questi assumono valori compresi tra -1 (quando le variabili considerate sono inversamente correlate) e +1 (quando vi sia correlazione assoluta cioè quando alla variazione di una variabile corrisponde una variazione rigidamente dipendente dall'altra), ovviamente un indice di correlazione pari a zero indica un'assenza di correlazione. I coefficienti di correlazione sono derivati dagli indici di correlazione tenendo presenti le grandezze degli scostamenti dalla media.

In particolare, il coefficiente di correlazione lineare di Pearson-Bravais é calcolato come rapporto tra la covarianza delle due variabili ed il prodotto delle loro deviazioni standard, arrivando alla seguente formula:

$$-1 \leq p_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2} \sqrt{\sum_{i=1}^n (y_i - \mu_y)^2}} \leq +1 \quad (5)$$

3.3 Riflessione sui vettori X e Y del metodo CHROM

Per informazioni sul procedimento per arrivare ai vettori usati X e Y si guardi l'articolo [6], tuttavia qui si vorrebbe far notare i vantaggi della selezione di questi vettori X e Y .

X e Y sono stati scelti per rendere il segnale indipendente dal colore dell'illuminazione, questo significa che questo metodo é migliore in casi di luce non-bianca oltre ad essere indipendente dall'intensità dell'illuminazione.

Quindi, questo metodo potrebbe essere usato in ambienti generali e non solo quelli di laboratorio o con una luminosità già conosciuta.

3.4 Differenze e similitudini tra PCA e CHROM

Poiché facciamo un confronto di risultati tra il metodo hrv/CHROM e il metodo con PCA, si vuole parlare brevemente sulle differenze algoritmiche tra i due metodi.

In breve, il PCA (Principal Component Analysis) é una trasformazione che identifica patterns nei dati, ed esprime i

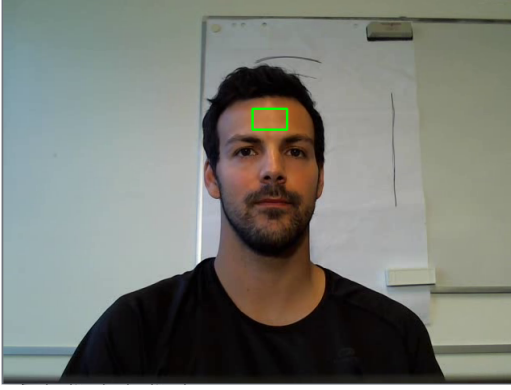


Fig. 5: frame di un video usato del dataset. La zona verde é l'area di interesse per il metodo hrv

dati in modo che evidenzi le somiglianze e differenze. L'idea di base nel PCA é di trovare le componenti s_i con $i = 1, \dots, n$ che possano spiegare la massima quantità di varianza possibile da N componenti trasformate linearmente.

Per maggiori informazioni sul metodo PCA, fare riferimento alla relazione del progetto [2] e relative citazioni. Si fa inoltre notare che i passi in comune (determinazione del ROI, successiva determinazione del battito per minuto, etc) non vengono qui considerati.

Per il metodo PCA principalmente i passi da fare sono i seguenti:

- 1) calcolo della matrice di covarianza sui dati in input
- 2) determinazione degli autovettori e autovalori
- 3) ordinamento degli autovalori affinché $\lambda_i \geq \lambda_{i+1}$

Per il metodo CHROM invece i passi sono i seguenti:

- 1) Pre-process con detrending (spa) e filtro passabanda di Butterworth
- 2) calcolo dei vettori X e Y come definito nell'equazione (1)
- 3) calcolo del segnale rPPG come $S = X - \sigma Y$
- 4) determinazione dei picchi sull'rPPG e calcolo del segnale rHRV

4 SIMULAZIONE E ESPERIMENTI

4.1 Dataset

É stato utilizzato come dataset un sottoinsieme di quello usato nell'articolo [7] chiamato "COHFACE". In questo dataset, i soggetti sono ripresi con quattro diverse condizioni di luminosità e durante le riprese é stato monitorato il valore dell'heart rate, valore che verrà utilizzato come ground truth per il confronto tra i metodi. Il frame rate di questi video é costante a 20 frame per secondo.

Si vuole fare notare questo particolare sul formato del video poiché nell'introduzione dell'articolo [1] viene detto che l'hrv richiede una stima accurata del segnale rPPG e che questo può portare ad una scarsa stima dell'HRV se sono individuati falsi picchi. Gli autori di quell'articolo hanno ovviato con un dataset con video a 50 fps, un framerate che una semplice camera non raggiunge.

Come dicevamo, il dataset [7] ha video con framerate di 20 fps. Ogni frame ha 3 componenti colore (RGB) e la

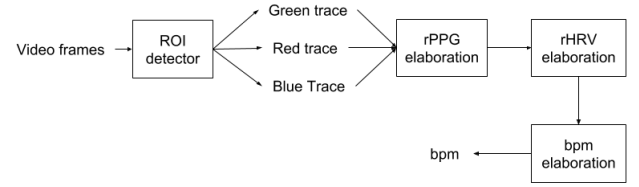


Fig. 6: Architettura dell'algoritmo (metodo hrv) per il calcolo della frequenza cardiaca

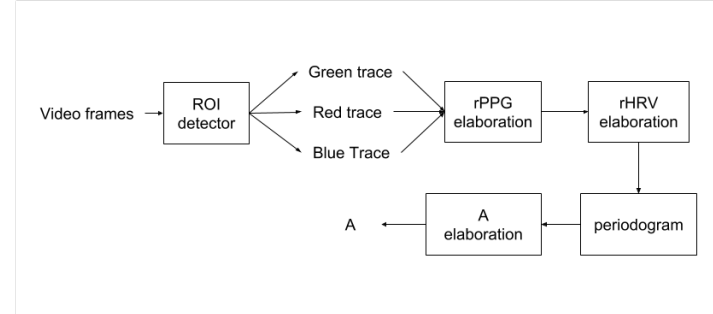


Fig. 7: Architettura dell'algoritmo per il calcolo della feature di arousal

risoluzione dei frame é di 640x480 pixel. In figura 5 si può vedere un frame di un video del dataset con la ragione di interesse evidenziata. Per gli esperimenti si é utilizzato un subset di 20 video diviso sui 5 soggetti registrati.

4.2 Architettura del sistema

In figura 6, viene mostrata la struttura dei passi dell'algoritmo per calcolare la frequenza cardiaca con il metodo hrv e verrà poi descritto nel dettaglio nella Sezione 4.3. Per l'architettura dei metodi pca e green usati per il confronto, vedere la relazione relativa [2].

In figura 7 invece viene mostrata la struttura dei passi dell'algoritmo per calcolare la feature di arousal. Per ottimizzare però, la feature verrà determinata dopo il calcolo del segnale rHRV durante l'approssimazione della frequenza cardiaca.

4.3 Dettagli implementativi

4.3.1 Calcolo delle frequenza cardiaca

In questo lavoro, é stata innanzitutto calcolata la frequenza cardiaca tramite il metodo CHROM e riadattati per video i metodi green e pca sviluppati nel progetto [2] per webcam. I tre metodi iniziano in maniera molto simile:

- 1) Detection del ROI su cui processare:
 - a) metodo green: il ROI é una porzione rettangolare della testa statica durante l'esecuzione
 - b) metodo pca: il ROI é la regione più ampia del viso nel frame
 - c) metodo hrv: il ROI é una porzione rettangolare della testa (centrata sulla fronte) individuata ad ogni frame
- 2) La ROI viene suddivisa nelle tre componenti RGB

3) Calcolo delle tracce dei canali RGB:

- Nel primo metodo, ad ogni frame calcola la media dei pixel del ROI solo sulla componente Green.
- Nel secondo e terzo metodo, ad ogni frame viene calcolata la media dei pixel del ROI su tutte le componenti RGB.

Ogni traccia conterrà quindi la sequenza delle medie della regione sulla corrispondente componente colore per ogni frame del video.

Il processing sulle tracce é diverso tra i diversi metodi, quindi si rimanda alla relazione [2] nella sezione 4.3 per i metodi green e pca. Parleremo invece dei passi del metodo hrv.

Ecco i passi del metodo CHROM successivi:

- 1) Date le tracce con le medie viene fatto il pre-process in cui su ogni traccia viene calcolata la zero mean, viene fatto un detrend con spa (smoothness prior approach, vedi l'articolo [4]) e di seguito viene utilizzato un filtro passabanda di Butterworth con cutoff Frequency di 0.7 e 3.5 Hz.
- 2) Vengono calcolati i vettori X e Y come combinazioni lineari delle tre tracce. I vettori X e Y sono mostrati nell'equazione (1)
- 3) Viene calcolato il segnale rPPG S come $S = X - \alpha Y$ dove $\alpha = \sigma(X)/\sigma(Y)$ e $\sigma(\cdot)$ é la deviazione standard.
- 4) viene fatta una peak detection sul segnale S. Come algoritmo per il peak detection viene utilizzato quello dell'articolo relativo [5]. Si vuole però fare notare che invece del valore del coefficiente utilizzato nell'articolo (in cui il coefficiente é 0.7) é stato usato il valore 0.625 per minimizzare l'errore con l'attuale framerate dei video.
- 5) vengono calcolati gli intervalli tra i picchi individuando quindi il segnale rHRV.
- 6) per calcolare il battito viene utilizzato come approccio quello della media degli intervalli tra battiti (anche detti "inter-beat interval" o ibi) arrivando quindi alla seguente formula:

$$bpm = \frac{\overline{ibi}}{60} \quad (6)$$

dove per \overline{ibi} si intende appunto la media della serie di ibi.

Si può discutere sulla scelta dell'approccio utilizzato a causa di problematiche in caso di pazienti affetti da aritmie, tuttavia l'approccio con la ricerca del picco più alto sulle frequenze del segnale rPPG sembrava dare delle approssimazioni più grossolane e per questo é stato scelto l'approccio descritto.

Questo metodo é stato poi modificato aggiungendo una "finestra". Per finestra si intende che il calcolo della frequenza cardiaca non viene calcolato su tutti i frame passati ma solo quelli di un determinato numero di frame precedenti, eliminando quelli antecedenti. Di norma, il calcolo della frequenza cardiaca viene effettuato sulle tracce che contengono i valori di tutti i frame precedenti. Questo

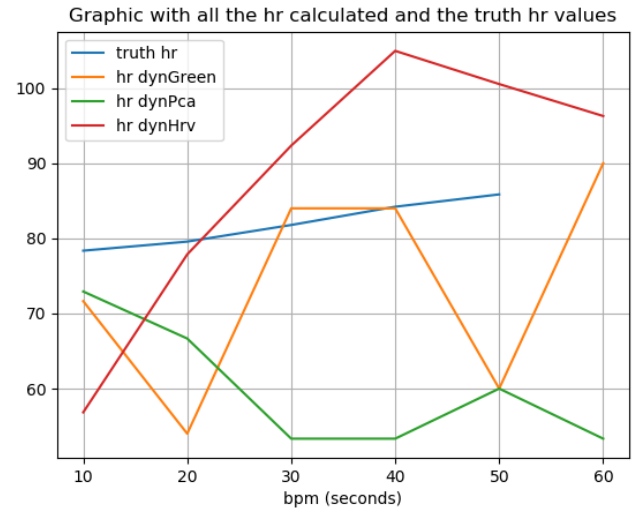


Fig. 8: Esempio di valori hr calcolati con una finestra di 10 secondi. Le distanze medie sono: G=13, P=21, H=12

significa che dopo 60 secondi di un video a 20 fps, avremo le tracce di dimensione $x = 60 * 20 = 1200$. Se invece volessimo come finestra 2 secondi, quello che facciamo é semplicemente il calcolo della frequenza cardiaca ogni 2 secondi e dopo il calcolo si azzerano le tracce. Utilizzando lo stesso fps, ogni x secondi di video, avremo tracce di dimensione $x*20$. Questo é utile nel caso in cui il battito non é stabile durante tutto il video o di video molto lunghi ma richiedenti un'approssimazione in real-time, tuttavia l'uso di finestre troppo piccole comporta valori con errore solitamente più alti rispetto al calcolo senza finestra, come si può osservare in figura 8. Per differenziare i due approcci, é stato scelto di chiamare gli approcci con la finestra "dinamici" (o "dyn").

Il metodo hrv é stato inoltre sviluppato per il calcolo della frequenza cardiaca in real-time tramite webcam, tuttavia non si parlerá di questo perché il metodo é identico con l'unica differenza di come vengono raccolti i frame, cosa che a livello algoritmico o di accuratezza non é rilevante con il funzionamento in sé.

4.3.2 Confronto tra i metodi utilizzati

Prima di fare il confronto, si deve prima rendere la serie ground truth della stessa dimensione delle serie calcolate, quindi facciamo un'approssimazione tramite media dei valori. Nel caso avessimo fatto il calcolo della frequenza cardiaca ogni secondo, vorremmo che il ground truth avesse pure la stessa dimensione. Per fare questo approssimiamo la serie di ground truth in maniera da avere la media dei valori dell'intervallo richiesto. Per esempio, se nel ground truth sono stati monitorati i valori x e y al tempo 0.4 e 0.7, e vogliamo il valore per l'intervallo 0-1, approssimeremo il valore per il secondo 1 come $\frac{x+y}{2}$. In questa maniera per ogni valore calcolato dei metodi, avremo un valore di confronto approssimato nel ground truth. Tramite questa approssimazione, é possibile comparare sia i valori trovati dai metodi "dinamici" sia quelli "classici" calcolati senza finestra. Poiché però i due metodi hanno dimensioni di-

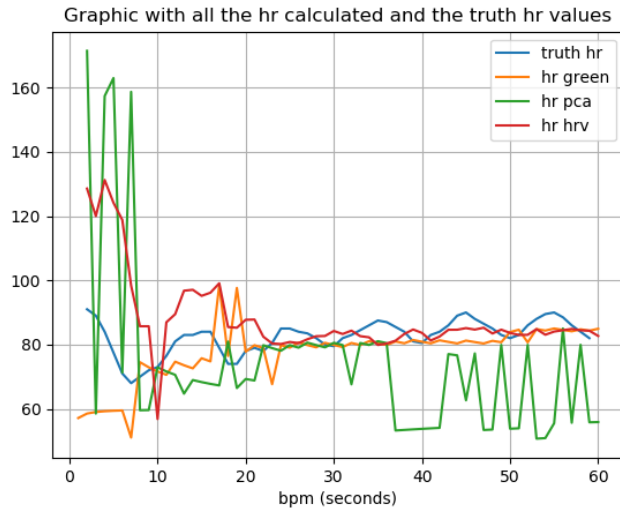


Fig. 9: Esempio di valori hr calcolati senza finestra. Le distanze medie sono: G=6, P=19, H=9.

verse, i confronti si faranno solo tra metodi dello stesso tipo, quindi tra metodo hrv e green classici e non tra hrv classico e hrv dinamico per esempio. Per confrontare le serie di frequenze cardiache calcolate tramite i metodi, sono stati svolti i seguenti passi per ogni serie calcolata:

- 1) Per ogni valore della serie si determina la distanza tra il valore calcolato (dal metodo) e il valore approssimato nel ground truth. (si veda l'equazione 3 per la distanza)
- 2) Dalla serie di distanze determinate, si calcola la distanza media.

In questa maniera tutti i metodi hanno un valore da confrontare. È abbastanza banale notare che il confronto tra i metodi quindi verrà fatto tra queste medie invece che farlo singolarmente tra valori della serie, come è anche banale che il metodo che si comporta meglio è quello con la distanza media più bassa rispetto al ground truth. Nelle figure 8 e 9, abbiamo un esempio di valori trovati e il confronto grafico tra le serie. Le due immagini sono state prese dal calcolo sullo stesso video con i due diversi approcci (con finestra e non).

4.3.3 Calcolo valore di arousal dallo spettro di potenza dell'rHRV

Nell'articolo [1], il segnale rHRV viene utilizzato per il calcolo di una feature basata sull'arousal che secondo gli autori è la prova di correlazione tra stati emotivi e il segnale rHRV. Prima di discutere sull'efficacia mostriamo i passi per ricavare questo valore:

- 1) determinato il segnale rHRV, si fa resampling del segnale stesso
- 2) Si calcola il periodogramma sul segnale rHRV
- 3) Si calcolano i valori VLF (very low frequency), LF (low frequency) e HF (high frequency) sul periodogramma trovato
- 4) Si determina la feature di arousal come

$$A = \frac{VLF + LF}{HF} \quad (7)$$

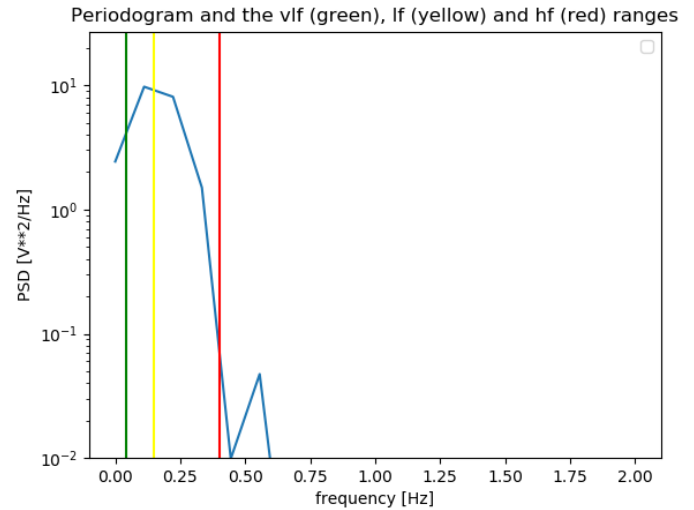


Fig. 10: periodogramma di una finestra da 10 secondi calcolato su un video del dataset

In figura 10 si mostra un periodogramma ricavato con una finestra di 10 secondi su un video del dataset. Le linee verticali indicano i range delle frequenze per renderle identificabili graficamente. Il range del VLF è tra 0 e 0.04, LF tra 0.04 e 0.15 ed HF tra 0.15 e 0.4 Hz.

Nell'articolo, questa feature viene calcolata a fine video (cioè con la traccia completa) in sessione neutrale e in sessione con alto arousal. Calcolati questi due valori, utilizzano il valore determinato nella sessione neutrale per normalizzare il valore della sessione con arousal. Questo significa che per poter calcolare una differenza significativa tra feature, bisogna avere un riferimento ad una sessione neutrale del soggetto, quindi non è una rilevazione generalizzabile in maniera efficace su persone su cui non si hanno dati. Per questo progetto viene determinato la feature per ogni finestra, e viene mostrata graficamente la serie di valori normalizzati sul primo valore determinato nel video e la rispettiva frequenza cardiaca calcolata, quindi abbiamo calcolato la feature per finestre temporali successive, tuttavia si fa notare nuovamente che non è molto significativo se non si hanno altre informazioni oltre alla frequenza cardiaca rilevata e il video stesso.

4.3.4 Librerie utilizzate

Per avere un progetto più replicabile e riproducibile, si è optato per inserire tutte le dipendenze (e relativa versione usata durante lo sviluppo) nel file README.md del repository. Ecco le librerie necessarie per utilizzare questo progetto:

- OpenCv: per il rilevamento delle regioni di interesse, il mostrare i video con i ROI durante l'utilizzo e per il sistema con metodo hrv anche su webcam.
- numPy: per le varie operazioni sui dati, trasformate di Fourier eccetera
- matplotlib: per visualizzare i grafici delle frequenze cardiache, periodogrammi eccetera
- mdp dipendenza necessaria del progetto arrota per il processo PCA

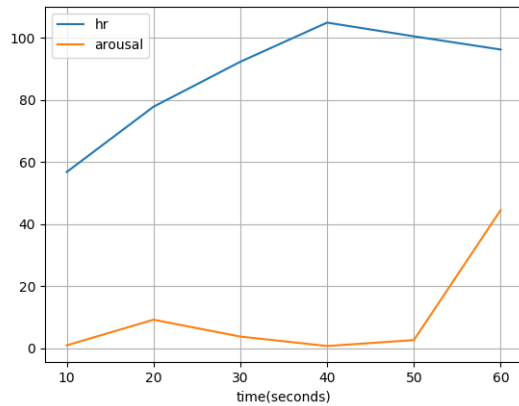


Fig. 11: Esempio di valori della feature di arousal calcolati

- `sciPy`: utilizzato per l'interpolazione di array, della trasformazione in segnale e del filtro di Butterworth
- `pandas`: utilizzato nel peak detection per il calcolo del "rolling mean", cioè il filtro smoothing "moving average" (utilizzato per il peak detection nell'articolo [5]).

5 RISULTATI OTTENUTI

Si vuole fare nuovamente presente che i test sono stati effettuati sul un dataset con video a 20 fps mentre il dataset utilizzato nell'articolo [1] utilizza video a 50 fps. Come diceva nell'introduzione l'articolo, una stima grossolana del segnale rPPG comporta una stima con errori sul segnale rHRV. Si vuole sottolineare questo prima di procedere con i risultati ottenuti.

5.1 Calcolo del battito cardiaco con il metodo CHROM

Il calcolo della frequenza cardiaca, come si faceva notare, deriva dall'approssimazione del segnale rPPG, con conseguente approssimazione del segnale rHRV poiché la determinazione dei picchi corretti dipende molto dal valore dei punti trovati nella serie. Per rendersi conto di quanto la differenza sia importante si guardino le figure 3 e 12 in cui la prima mostra un segnale ppg ideale pulito, il secondo invece è il segnale rPPG determinato su un video a 20 fps del dataset. In figura 12 sono inoltre stati inseriti i picchi determinati durante l'algoritmo, in maniera da rendere visibile i punti trovati. Come si può osservare in figura 9, il valore di hr varia molto all'inizio della serie o comunque quando la finestra su cui calcola è molto stretta, ma proseguendo si arriva a errori di calcolo sempre più piccoli, almeno nei casi del dataset (in cui il battito è più o meno stabile). La distanza media per video con questo metodo verrà discussa nella prossima sezione (5.2), però si fa notare che le distanze medie sono comprese tra i 9 e 30 senza finestra e tra i 10 e i 40 con la finestra da 10 secondi, un risultato non molto accurato.

5.2 Confronto dei metodi

In figura 13 e 14 possiamo vedere le distanze medie dei metodi su tutto il dataset preso in considerazione. Le linee

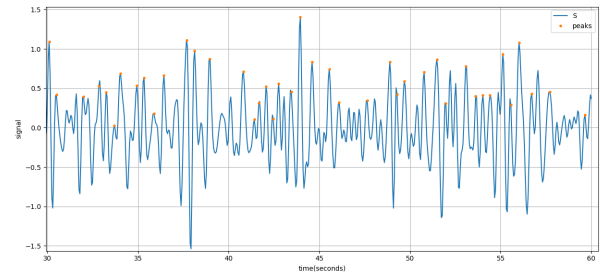


Fig. 12: Esempio di segnale rPPG determinato su un video del dataset

evidenziano i range di valori di errore medio. I range utilizzati sono i seguenti:

- blu: distanza media tra 0 e 3. Serie accurata con un errore irrilevante.
- giallo: distanza media da 3 a 7. Serie accurata con errore accettabile.
- rosso: distanza superiore a 7. Serie non accurata.

Lasciando al lettore il confronto qualitativo sull'accettabilità delle serie calcolate date in figura, si vuole far notare che nel caso dei metodi senza finestra, l'errore medio minimo tra i tre metodi è stato per il 42.11% del metodo 'Green', il 42.11% del metodo 'HRV/CHROM' e il 15.79% del metodo PCA, mostrando che nel minuto i due metodi minimizzano la distanza media. Inoltre, nella maggioranza dei casi in cui il metodo hrv è migliore del metodo 'green' questo accade quando l'illuminazione è bassa e c'è un alto contrasto, dato magari dalla indipendenza dalla luce per come viene costruito il segnale.

L'errore medio nelle serie calcolate con una finestra di 10 secondi ci dice invece qualcosa di molto diverso. In questo caso l'errore medio minimo è stato per ben l'80% merito del metodo Green, mentre solo il 20% del metodo hrv (con invece lo 0%, cioè non ha mai avuto una distanza media minima, il metodo pca). In figura 14 si può vedere che la distanza media del metodo green rientra nel range accettabile per la maggioranza dei casi, anche nei casi di illuminazione scarsa. Questo ci mostra che nei casi con finestra piccola (cioè con un monitoraggio inferiore ai 30 secondi) sembra migliore l'utilizzo del metodo Green. Tuttavia bisogna anche considerare che la distanza media della finestra è calcolata con la media delle frequenze cardiache monitorate nell'intervallo e non confrontati invece con l'ultimo valore dell'intervallo. Questo significa che quando battito è stabile allora il valore è corretto, ma quando ci sono sbalzi (alzamenti o abbassamenti del battito) allora potrebbe non esserlo più.

5.3 Determinazione feature di arousal

Poiché non ci sono valori di arousal di riferimento nel dataset e poiché il dataset stesso era stato concepito per uno scopo diverso, risulta difficile esprimere dei risultati quantitativi sull'accuratezza della feature o dei valori stessi che l'hanno derivata. Potremo quindi dare solo dei risultati qualitativi, ma si invita il lettore interessato ad utilizzare il progetto su un dataset con dei valori di arousal monitorati o noti. Si riguardi la figura 11. La serie di arousal, come

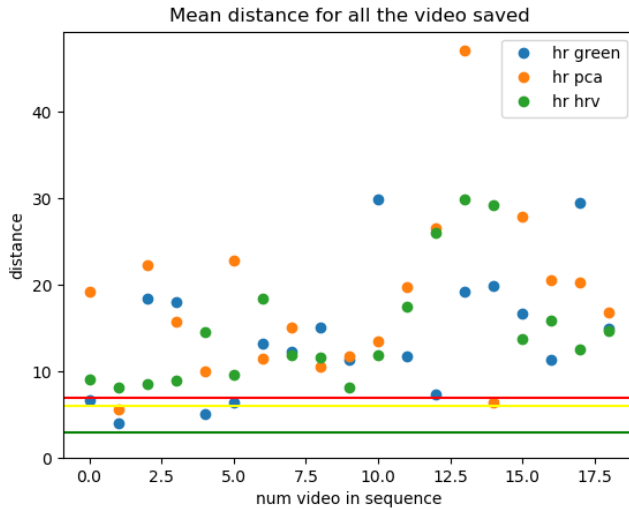


Fig. 13: Distanze medie dei metodi per ogni video del subset del dataset dato

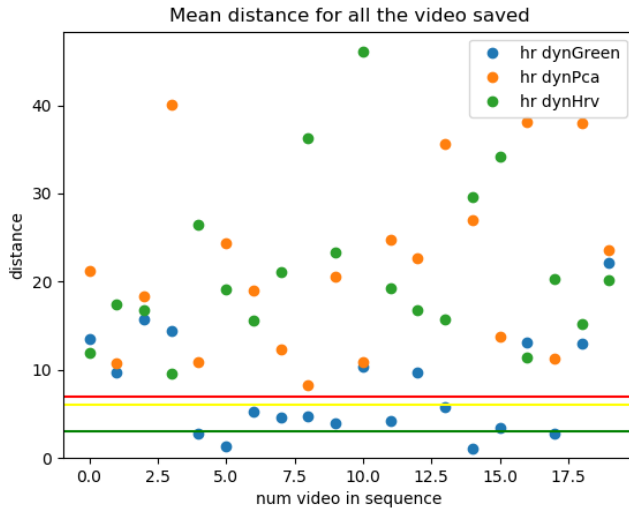


Fig. 14: Distanze medie dei metodi dinamici per ogni video del subset del dataset dato

fatto anche nell'articolo [1], viene normalizzato. Poiché non si avevano informazioni sullo stato di arousal dei soggetti, si è optato di utilizzare come "stato neutrale" il primo valore calcolato nella serie di feature. Qualitativamente si può notare che non c'è correlazione tra il battito cardiaco calcolato, tuttavia questo può essere causato da diversi elementi. La causa più banale potrebbe essere che l'approssimazione è così mediocre che tutto ne risente, dal calcolo del battito cardiaco al calcolo della feature. A prova di questo si mostra in figura 10 il periodogramma calcolato per ricavare la feature sul metodo dinamico. Come si può notare l'approssimazione ha davvero pochi punti e questa potrebbe essere una delle cause. Un'altra causa potrebbe essere che il valore iniziale non era in una sessione neutrale, e questo può rovinare la normalizzazione dei valori successivi. Altra causa potrebbe essere semplicemente il fatto che le due

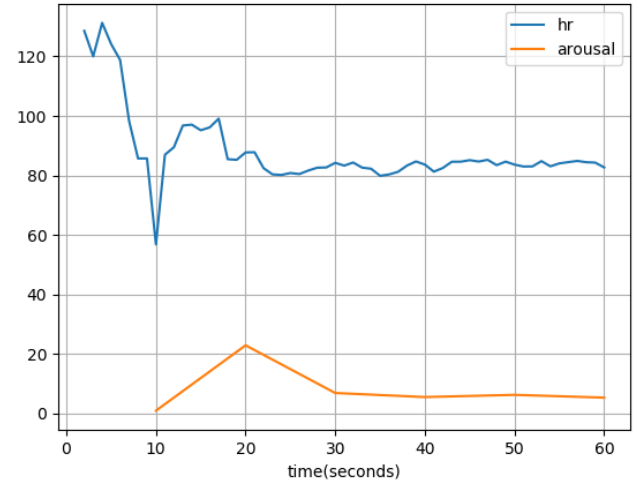


Fig. 15: Confronto tra la feature calcolata ogni 10 secondi e la serie di hr calcolata con metodo hrv senza finestre su un video del dataset

misure non sono correlate, anche se potrebbe sembrare poco plausibile o improbabile.

Si guardi ora invece la figura 15. La feature di arousal è stata calcolata sempre ogni 10 secondi, ma su tutta la serie invece che solo sulla finestra dei 10 secondi come accade nei metodi dinamici. Qui invece una lieve correlazione tra la frequenza cardiaca sembra esserci, cosa che confermerebbe il problema delle finestre molto piccole.

Si consideri infine che lo scopo di questa feature è solo l'affermare che dato uno stato neutrale del soggetto, se il valore della feature è più alto del valore della sessione neutrale allora l'arousal è aumentata (e conseguentemente lo stato emotivo è variato). In più se usiamo il valore della sessione neutrale per normalizzare, si ha che se il valore normalizzato è maggiore di 1 allora l'arousal è aumentata.

6 COMMENTI CONCLUSIVI

Come si può vedere dai risultati nel capitolo 5, il calcolo con il metodo HRV senza finestra lascia abbastanza a desiderare, ma con finestre superiori ai 30 secondi sembra dare dei risultati accettabili, come si può anche vedere in figura 9 dopo il 30-esimo secondo.

Il metodo HRV dinamico con finestre di 10 secondi, come si poteva immaginare da quanto appena detto, non dà risultati molto decenti, dove invece sembra andare molto bene il metodo Green analogo. Ovviamente come è stato ripetuto più volte uno dei problemi di questo metodo è che richiede un frame rate molto alto per ridurre gli errori dati da rumore e approssimare bene il segnale rPPG. Quando questo non succede il segnale è rumoroso e i risultati non sono accettabili con finestre piccole. Questo significa che il metodo sembra accettabile se può calcolare la frequenza cardiaca con finestre da almeno 30 secondi.

Un'idea potrebbe essere quella di arrivare a serie di x secondi (con $x \geq 30$) e poi invece di eliminare completamente la finestra si potrebbe eliminare il primo secondo

inserito della serie (cioé tutti i valori del secondo indicato) e inserire il nuovo secondo trovato alla fine. In questo modo si dovrebbero avere dei valori piú accettabili. Il metodo Green ha a sua volta altre problematiche, come il fatto che lo skin detection é statico, che lo rende utile solo in casi molto specifici in cui il paziente non deve o può muoversi. Nel caso del game user research per esempio non ci si può aspettare che il playtester stia immobile perché questo influirebbe sul suo modo di giocare e renderebbe l'utilizzo dell'hr in maniera remota utile quanto un monitoraggio invasivo (e anche meno, avendo piú rumore).

Per quanto riguarda la feature di arousal, poiché il feature non é generale per tutte le persone o persino per tutti i momenti (si pensi che se una persona é in tensione durante la sessione neutrale il resto delle valutazioni non saranno valide) sulla stessa persona, si pensa che come feature é interessante, ma richiede non solo un alto frame rate, ma anche delle conoscenze iniziali che si potrebbero non avere e questo rende il suo utilizzo molto meno utile di quello che ci si potrebbe aspettare inizialmente. Dati gli elementi registrati durante una sessione di playtesting (solitamente una registrazione video sul viso, una sul controller e una sul gameplay), sarebbe piú semplice derivare lo stato emotivo tramite face recognition e cross-validare queste informazioni con le variazioni della frequenza cardiaca e l'hrv.

REFERENCES

- [1] R. M. K. N. R. G. e. a. Yannick Benezeth, Peixi Li, "Remote heart rate variability for emotional state monitoring," 2018.
- [2] L. Arrotta, "Rilevazione contactless del battito cardiaco di un soggetto ripreso da webcam," 2018.
- [3] L. E. Nacke, *Games User Research and Physiological Game Evaluation*. Cham: Springer International Publishing, 2015, ch. 4, pp. 63–86.
- [4] P. O. R.-a. Mika P. Tarvainen and P. A. Karjalainen, "An advanced detrending method with application to hrv analysis," *IEEE Transactions On Biomedical Engineering*, 2002.
- [5] L. R. G. R. Srinivas Kuntamalla, "An efficient and automatic systolic peak detection algorithm for photoplethysmographic signals," *International Journal of Computer Applications*, vol. 97, 2014.
- [6] V. J. Gerard de Haan, "Robust pulse-rate from chrominance-based rppg," *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, 2013.
- [7] S. M. G. Heusch, A. Anjos, "A reproducible study on remote heart rate measurement," 2017. [Online]. Available: <https://pdfs.semanticscholar.org/48ac/7cb27b3bb620ea0de8cadea8f7c1c5653885.pdf>