



TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

**JOONAS LATUKKA
ELEKTROKARDIOGRAMMIN RR- JA QT-VÄLIEN AIKASARJA-
ANALYYSI SYDÄNTUTKIMUKSESSA**

Kandidaatintyö

Tarkastaja: Prof. Esa Räsänen

TIIVISTELMÄ

JOONAS LATUKKA: Elektrokardiogrammin RR- ja QT-välien aikasarja-analyysi sydäntutkimuksessa

Tampereen teknillinen yliopisto

Kandidaatintyö, 21 sivua, 4 liitesivua

Lokakuu 2018

Teknis-luonnontieteellinen koulutusohjelma

Pääaine: Teknillinen fysiikka

Tarkastaja: Prof. Esa Räsänen

Avainsanat: elektrokardiogrammi, RR, QT, sydän, DFA, Poincarén kartta

Sydämen toiminta on tärkeää tuntea erilaisten sairauksien tunnistamiseksi ja välttämiseksi. Sydäimestä saadaan merkittävästi tietoa mittaamalla sen sähköistä signaalia eli sydänsähkökäyrää. Tästä signaalista voidaan määrittää erilaisia aikavälejä. Perättäisten aikavälien muodostamiin aikasarjoihin voidaan soveltaa aikasarja-analyysissä käytettyjä menetelmiä. Menetelmien perusteella voidaan havaita, mikäli sydän toimii jollain normaalista poikkeavalla tavalla.

Tässä työssä tutustutaan sydämen sähkökäyrästä muodostettaviin aikasarjoihin ja tutkitaan niiden sisältämiä korrelaatioita. Lisäksi perehdytään trendit poistavaan fluktuaatioanalyysiin ja Poincarén karttoihin, joilla saadaan tietoa aikasarjojen sisältämistä korrelaatioista. Työssä suoritettavan tutkimuksen kohteena on PhysioNet-sivustolta ladattu sydändata. Aineisto sisältää 18 sydämen toiminnaltaan terveiltä henkilöiltä mitattua noin vuorokauden pituista sydänsähkökäyrää. Sähkökäyrästä on laskettu RR- ja QT-aikavälit, jotka muodostavat menetelmissä käytettävät aikasarjat.

Tutkimuksen tuloksena on saatu selville, että sydämen lyöntitiheyden lisäksi myös lyönnin kestoa kuvaava aikaväli sisältää pitkän kantaman korrelaatioita. Korrelaatioiden voimakkuus on lyhyellä aikavälillä heikohko, joka ilmenee korrelaation voimakkuutta kuvaavan skaalauseksponentin matalasta arvosta. Skaalauseksponentin arvo kuitenkin kasvaa likimain arvoon 1,0 aikavälin kasvaessa, mikä tarkoittaa sarjassa esiintyvän pitkän aikavälin korrelaatioiden olevan voimakkaita. Tulos on yhdenmukainen muiden saman aihealueen tutkimusten kanssa.

SISÄLLYS

1. Johdanto	1
2. Aikasarjat sydämessä	2
2.1 Fraktaalit ja korreloituneet aikasarjat	2
2.2 Sydämen aikasarjaominaisuudet	4
2.3 Sydänsolut	6
3. Menetelmät	7
3.1 Trendit poistava flukтуаatioanalyysi (DFA)	7
3.2 Alfa-beeta-suodatin	8
3.3 Poincarén kartta	9
4. Tutkimusaineisto	10
4.1 Physionet	10
4.2 RR- ja QT-aikavälien laskeminen	10
4.3 Datan suodatus	11
5. Tulokset	13
5.1 DFA	13
5.2 Alfa-beeta-suodatin	15
5.3 Poincarén kartta	16
6. Yhteenveto	18
Lähteet	20
Liite A. RR- ja QT-arvojen laskemiseen käytetty Python-koodi	22

1. JOHDANTO

Sydän ja sen syke on ihmisen elämän kannalta ehdottoman tärkeässä asemassa. Sydämen vajaatoiminta heikentää elämänlaatua merkittävästi ja kasvattaa kuolemanriskiä. Siksi onkin tärkeää ymmärtää syvällisesti kuinka sydän toimii ja mihin sen toiminta perustuu.

Sydämen toiminnasta saadaan kattavasti tietoa tutkimalla sydänsähkökäyrää. Sähkökäyrän signaali ohjaa sydämen lyöntitiheyttä ja lyönnin aikaisia prosesseja, eli se on merkittävässä osassa säätelemässä sydämen ominaisuuksia. Elektrokardiogrammia tutkimalla onkin aiemmin todettu, että esimerkiksi sydämen lyöntitiheys on korreloitunut aikasarja, jolloin sydämen lyöntihistoria vaikuttaa myös tuleviin lyön-teihin.

Tämän kandidaatintyön tavoitteena on soveltaa sydämen sykkimisestä saatavaan tietoon erilaisia aikasarja-analyysin menetelmiä, joiden tuloksista voi päätellä sydämessä vaikuttavia lainalaisuuksia. Työssä käytettävät menetelmät ovat trendit poistava fluktuaatioanalyysi, alfa-beeta-suodatin ja Poincarén kartat. Menetelmiä käytetään työssä sydämen RR- ja QT-aikaväleistä koostuvien aikasarjojen tutkimiseen. Tutkimuksessa käytetty data on ladattu PhysioNet-verkkosivustolta. Ladattua dataa on tarvinnut käsitellä ja suodattaa monin tavoin ennen sen käyttämistä analyysissä. Laskennan tuloksena on osoitettu kummankin aikasarjan sisältävän vahvoja sisäisiä korrelaatioita.

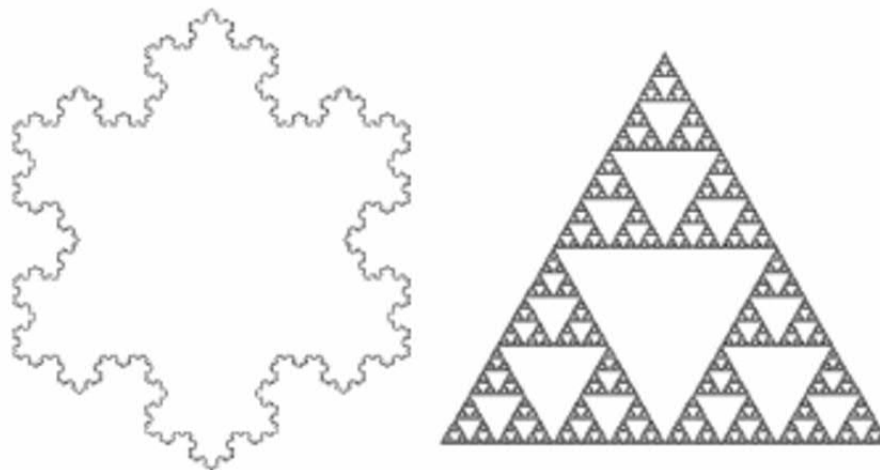
Työn aluksi luvussa 2 esitellään työn taustalla oleva teoria. Luvussa käydään läpi yleisesti fraktaalisuuden ja korreloituneiden aikasarjojen ominaisuuksia sekä esitellään lyhyesti yksittäisten sydänsolujen käyttäminen vastaavissa tutkimuksissa. Seuraavaksi perehdytään työssä käytettäviin aikasarja-analyysin menetelmiin ja niiden vahvuuksiin ja heikkouksiin. Luvussa 4 selostetaan, kuinka tutkimusaineisto on hankittu ja mitä ohjelmia ja algoritmeja on tarvittu sen muokkaamiseksi sopivaan muotoon. Viimeiseksi ennen yhteenvetoa esitellään laskennasta saadut kuvaajat ja tulokset sekä pohditaan niiden merkitystä ja luotettavuutta.

2. AIKASARJAT SYDÄMESSÄ

Työn taustalla vaikuttaa monimutkainen teoria fraktaalien ja korreloituneiden aikasarjojen ominaisuuksista. Tässä työssä teoriaa sovelletaan sydämen aikasarjaominaisuuksien tutkimuksessa. Lisäksi tutustutaan sydämen ja yksittäisten sydänsolujen yhtenevyyksiin ja eroavuuksiin.

2.1 Fraktaalit ja korreloituneet aikasarjat

Työssä käytettävät menetelmät sekä mittausdata perustuu fraktaaleihin. Fraktaali on itsesimilaarinen kuvio, joka muistuttaa itseään useissa eri mittakaavoissa. Fraktaaleja esiintyy sekä matemaattisina rakennelmina että luonnon ilmentyminä. Tyypillisiä matemaattisia esimerkkejä ovat kuvassa 2.1 näkyvät Kochin lumihiutale ja Sierpińskin kolmio. Luonnossa fraktaaleja voi havaita esimerkiksi maan pinnanmuodoissa, merten rantaviivassa tai pilvissä [1].

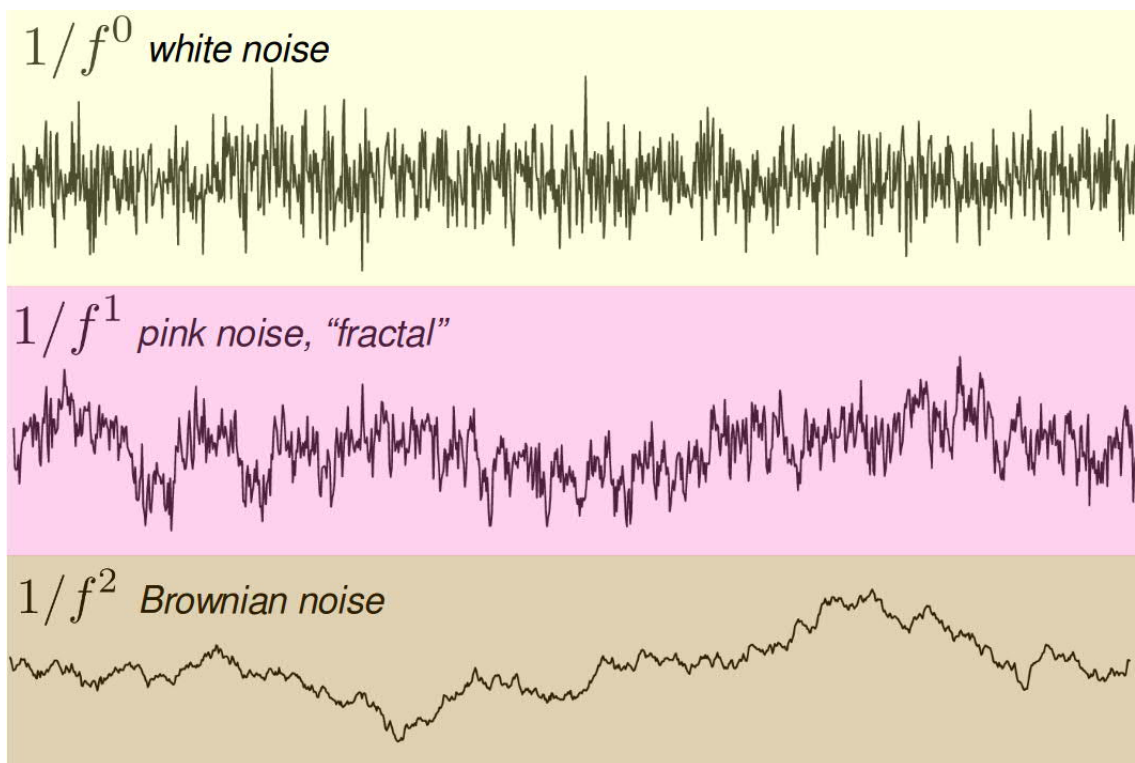


Kuva 2.1 Kochin lumihiutale ja Sierpińskin kolmio [2].

Matemaattinen fraktaali on pilkottavissa äärettömän monta kertaa pienempiin osiin. Luonnossa esiintyvissä sarjoissa tämä ei ole mahdollista, vaan fyysiset tekijät kuten mittaustarkkuus estävät itsesimilaarisuuden jatkumisen. Tämä ei kuitenkaan estä matemaattisista fraktaaleista johdetun teorian käyttämistä reaali maailman ilmiöiden tutkimisessa.

Fraktaalit eivät välttämättä ole spatiaalisia, vaan fraktaalisuuden piirteitä voidaan löytää myös erilaisista aikasarjoista. Aikasarjalla tarkoitetaan sarjaa $\{x_t\}_{t=1,\dots,N}$, jonka datapisteet x_t ovat tasaisin aikavälein mitattuja tutkittavan suureen arvoja. Fraktaalisuus sarjoissa näkyy useimmiten pitkän kantaman riippuvuuksina, jolloin sarjan uudet arvot määräytyvät historian perusteella. Tästä seuraa se, että sarjalle pelkkä keskiarvo tai varianssi ei kerro riittävästi sarjan ominaisuuksista [3].

Erilaiset riippuvuudet aikasarjassa vaikuttavat merkittävästi sarjasta piirretyn kuvaajan muotoon. Tätä on havainnollistettu kuvassa 2.2, jossa on kolme erilaista aikasarjaa. Aikasarjoissa esiintyvä spektritiheys $1/f^n$, $n = 0, 1, 2$ kertoo sarjan Fouriermuunnoksen skaalautuvuudesta, eli kuinka sarjassa esiintyvien taajuuksien amplitudi muuttuu taajuuden kasvaessa. Kuvan 2.2 ensimmäinen valkoista kohinaa sisältävä aikasarja ei sisällä minkäänlaisia riippuvuussuhteita, vaan sarjan arvot ovat täysin satunnaisesti jakautuneita. Tästä seuraa kuvaajan piikikkyyys sekä hetkittäisen keskiarvon pysyminen likimain vakiona koko sarjan alueella. Tällaisessa sarjassa spektritiheys on jakautunut tasaisesti koko taajuusalueelle.



Kuva 2.2 Esimerkki valkoisesta, vaaleanpunaisesta ja Brownin kohinasta (E.Räsänen).

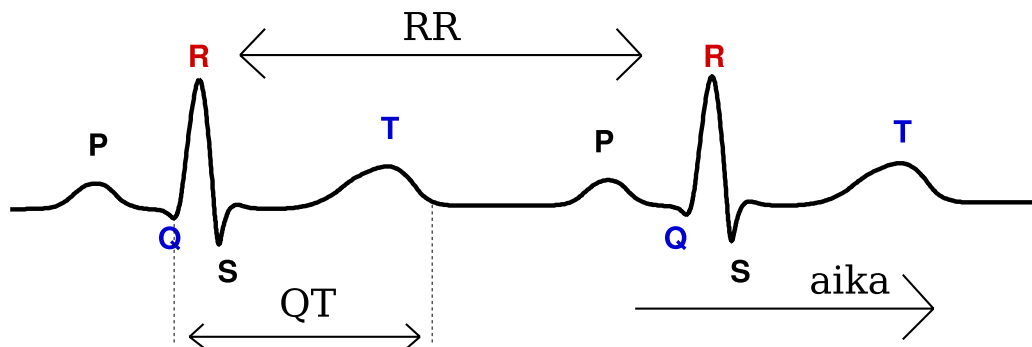
Kuvassa 2.2 alimmaisena olevan Brownin kohinaa esittävän aikasarjan datapisteet ovat sen sijaan vahvasti riippuvaisia edellisistä arvoista. Sarja on muodostettu siten, että sarjassa seuraavana oleva arvo on edellinen arvo lisättynä satunnaisella muutoksella. Käytännössä se on siis integraali valkoisesta kohinasta. Tästä seuraa sarjan

tasaisuus ja lokaalisti muuttuva keskiarvo.

Näiden kahden sarjan välillä on esimerkki vaaleanpunaisesta eli niin sanotusta $1/f$ -kohinasta, jossa on ominaisuuksia kummastakin aiemmin esitellystä sarjasta. Sarjassa on nähtävissä sekä valkoisen kohinan piikikkyys että Brownin kohinan mukainen lokaalin keskiarvon muuttuminen. Tämä saa aikaan kuvaajassa näkyvän itsesimilaarisuuden, eli sarjan data on fraktaalista.

2.2 Sydämen aikasarjaominaisuudet

Sydämen toimintaa voidaan tutkia mittaamalla sen sähköistä toimintaa. Operaatiota kutsutaan elektrokardiogrammiksi (EKG). Normaalin sydämen lyönnin aikana sähkökäyrästä on tunnistettavissa viisi eri aaltoa, jotka ovat P-, Q-, R-, S- ja T-aallot. Nämä on esitetty kuvassa 2.3. Aaltojen esiintymiskohdista aikajanalla voidaan laskea RR-, QT-, PR- ja ST-aikavälit. Näistä kahta ensimmäistä käytetään tämän työn tutkimuskohteena.

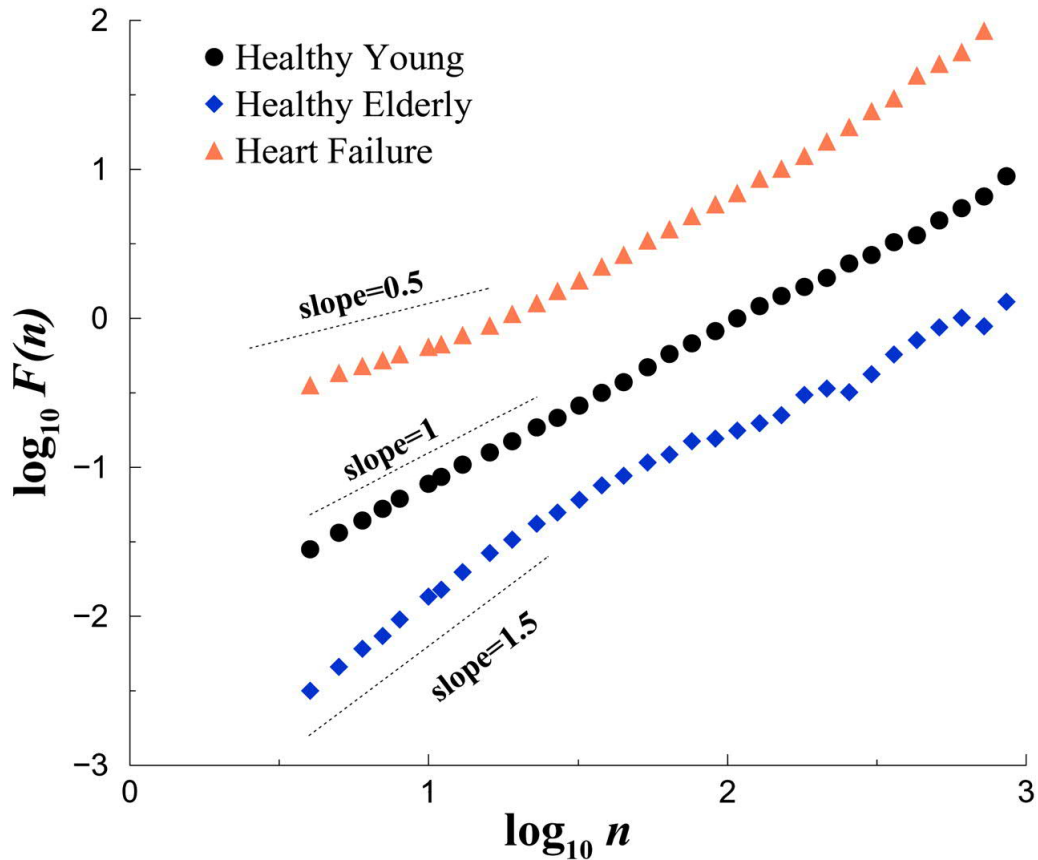


Kuva 2.3 Elektrokardiogrammikäyrän osat

RR-aikavälillä tarkoitetaan siis sydämen kahden lyönnin R-piikkien välistä aikaa. Siten yleisemmin mitattu sydämen syketiheys on kääntäen verrannollinen RR-aikavälin arvoon. Sydän mukautuu kehon rasitukseen RR-aikaväliä muuntelemalla. Kasvaneen hapenkulutuksen aikana sydän sykkii tiheämmin, jolloin RR-aikaväli pienenee. Aikavälin arvo voi siksi muuttua huomattavasti mittauksen aikana, jos mitattavan henkilön rasitusaste muuttuu. Vaihtelu on huomattavaa myös tasaisen rasituksen alaisuudessa, vaikkakin vaihtelu yleisesti pienenee rasituksessa [4].

QT-aikavälillä tarkoitetaan aikaa QRS-kompleksin alusta T-aallon loppuun. Aikaväli kuvaa siten yksittäisen sydämen lyönnin kestoa. QT-aikavälin vaihtelu on terveellä

potilaalla huomattavasti pienempää kuin RR-aikaväleillä. RR- ja QT-aikasarjat ovat kuitenkin yhteydessä toisiinsa ja niiden väliltä on löydetty korrelaatioita [5, 6].



Kuva 2.4 Erot nuoren ja vanhan terveen sydämen sekä vajaatoiminnasta kärsivän sydämen välillä ovat nähtävissä DFA-analysissä [7].

Aiemmistä tutkimuksista tiedetään, että terveen sydämen RR-aikavälit ovat korreloituneita ja fraktaalisia. Tällöin sydämen on mahdollista muuttaa hetkellisesti RR-aikavälin keskiarvoa isommaksi tai pienemmäksi ja näin sopeutua ympäristön olosuhteisiin. Kuvassa 2.4 on havainnollistettu fraktaalisuutta käyttämällä trendit poistavaa fluktuaatioanalyysiä (detrended fluctuation analysis, DFA). Menetelmä on selitetty tarkemmin luvussa 3.1. Sykevälivaihteluiden skaalautuvuutta kuvaavat kulmakertoimet ovat 1 ja 1,5 terveille yksilöille sekä 0,5 vajaatoiminnasta kärsivälle sydämelle. Tämä kertoo siitä, kuinka oleellista sydämen toiminnan kannalta on sisältää pitkän aikavälin korrelaatioita, mikä näkyy kuvassa kasvaneena kulmakertoimena. Korrelaatioiden puuttuessa sydämen RR-aikavälit alkavat muistuttaa valkoista kohinaa, mikä näkyy matalampana kulmakertoimena. Tämän ilmiön on todettu olevan yhteydessä sydänsairauksiin kuten vajaatoimintaan [7].

2.3 Sydänsolut

Kokonaisen sydämen sijasta voidaan tutkia myös yksittäisiä sydänsoluja. Kantasoluista on mahdollista kasvattaa uusia sydänsoluja. Ne tuottavat sykkiessään sydämen tavoin muuttuvan sähköisen potentiaalin, joka on mitattavissa. Mitatusta signaalista on tunnistettavissa samoja ominaisuuksia kuin sydämen EKG-signaalista. Ne eivät ole kuitenkaan identtisiä, minkä takia vastaavista suureista käytetään erilaisia nimityksiä. Sydämen RR-aikaväliä vastaava mitta on solussa "interbeat interval" (IBI), joka kertoo siis kahden sykähdyksen välisen ajan. Yhden sykähdyksen kestoa mitta niin sanottu kenttäpotentiaalin kesto eli "field potential duration" (FPD), joka on sydämen QT-aikaväliä vastaava suure.[8]

Yksittäisille soluille tehtävä tutkimus voi tuoda paljon uutta tietoa sydämen toiminnasta. Fraktaalisen aikasarjan tuottamiseksi on systeemissä oltava olemassa kytkentämekanismi eli eräänlainen muisti systeemin historiasta. Tällaisen muistin muodostamiseksi systeemiltä vaaditaan monimutkaisia rakenteita. Soluille tehtävä tutkimus voi selventää, ovatko nämä rakenteet jo sydämen soluissa itsessään vai muodostuvatko ne useiden solujen yhteisvaikutuksesta sydämessä. Lisäksi sydänsoluja käyttämällä voidaan tutkia tehokkaasti esimerkiksi lääkaineiden vaikutusta sydämen toimintaan.

Tähän mennessä tehdyistä tutkimuksista on selvinnyt, että sydänsoluilla on todellakin vastaavia fraktaalisia ominaisuuksia kuin sydämellä varsinkin lyöntitiheyden osalta [9]. Solujen FPD-aikavälin varianssi on huomattavasti suurempi kuin sydämen QT-aikaväleillä, mutta tästä huolimatta FPD-sarjan skaalauseksponentti muuttuu samalla tavoin kuin QT-sarjalla skaalan kasvaessa [8]. Soluilla tehdyillä tutkimuksilla on saatu myös selville, kuinka lääkkeaineet, jotka aiheuttavat QT-aikavälin pidentymistä, vaikuttavat myös solujen FPD-aikaväliin pidentävästi [10]. Tämä mahdollistaa tehokkaamman potilaskohtaisen tutkimuksen tapauksissa, joissa potilas tarvitsee sydämen toimintaan vaikuttavia lääkkeitä.

3. MENETELMÄT

Sydändatan kompleksisuuden takia tutkimuksessa on käytettävä edistyneitä laskennallisia menetelmiä. Menetelmät pyrkivät selvittämään aikasarjoissa esiintyvien korrelaatioiden voimakkuuksia eri aikaväleillä. Seuraavaksi esitellään tässä tutkimuksessa käytetyt menetelmät.

3.1 Trendit poistava flukтуаatioanalyysi (DFA)

Yksi tutkimuksessa käytetyistä menetelmistä on niin kutsuttu trendit poistava flukтуаatioanalyysi (detrended fluctuation analysis, DFA). Menetelmän etuna on pitkän aikavälin korrelaatioiden karakterisointi epästationaarisissa aikasarjoissa. Menetelmässä pyritään selvittämään aikasarjassa vaikuttava skaalausekspONENTTI α , jonka arvo kertoo korrelaation voimakkuuden erisuuruuksilla aikaväleillä.

DFA voidaan tiivistää neljään eri vaiheeseen. Olkoon sarjan pituus N ja siinä esiintyvä data $\{x_t\}_{t=1,\dots,N}$. Ensimmäiseksi muodostetaan aikasarjan profiili laskemalla integroitu summa aikasarjan arvoista vähennettynä sarjan keskiarvolla:

$$y(k) = \sum_{t=1}^k x_t - \langle x \rangle. \quad (3.1)$$

Toisessa vaiheessa profiili jaetaan yhtä suuriin ikkunoihin, joiden leveys on s . Tällöin ikkunoiden lukumäärä on N/s . Jokaisessa ikkunassa sarjaan sovitetaan n astetta oleva polynomi y_{tr} käyttäen pienimmän neliösumman menetelmää. Sovitetta käytetään poistamaan datasta paikalliset n astetta olevat trendit. Tässä työssä $n = 1$, jolloin menetelmä poistaa datassa esiintyvät lineaariset trendit. Koska koko profiili ei jakaannu tasaisesti eri ikkunoille, on muodostettava ikkunat myös päinvastaisessa järjestyksessä, jolloin ikkunat kattavat tasaisesti koko profiilin [8].

Seuraavaksi lasketaan keskimääräinen varianssi eri ikkunoiden sarjoista trendien poiston jälkeen. Tuloksesta otetaan neliöjuuri, joka määrittää sarjan fluktuaation F kullakin ikkunan koolla:

$$F(s) = \sqrt{\frac{1}{2N/s} \sum_{m=1}^{2N/s} \left[\frac{1}{s} \sum_{i=1}^s [y_m(i) - y_{tr,m}(i)]^2 \right]}. \quad (3.2)$$

Ikkunan leveys on mahdollista määrittää välillä $s = 4 \dots N/4$.

Fluktuaatio noudattaa potenssilakia, jolloin $F(s) \sim s^\alpha$. Tällöin skaalauseksponentti saadaan selvitettyä piirtämällä fluktuaatiosta ikkunan koon suhteen log-log-kuvaaja. Kuvaajan kulmakerroin kertoo α :n suuruuden. Eksponentin arvoilla alle 0,5 aikasarja on antikorreloitunut, kun taas yli 0,5:n arvolla sarja on korreloitunut. Arvolla $\alpha = 0,5$ sarja kuvaa valkoista kohinaa, ja arvolla $\alpha = 1,5$ kyseessä on Brownin kohina. Arvo $\alpha = 1$ vastaa vaaleanpunaista $1/f$ -kohinaa, jolloin aikasarjan data on fraktaalista [8].

DFA-menetelmän heikkoutena on trendien poistamisesta johtuva fluktuaation heikentyminen. Kaikista pienimmillä ikkunakoilla ikkunaan sovitettu polynomi ei poista pelkästään todellista trendiä, vaan se mukautuu myös satunnaisvaihtelun arvoihin, jolloin datasarjan arvojen etäisyys sovitteesta pienenee todellista arvoa matalammaksi. Satunnaisvaihtelun poistamisen vaikutus on kuitenkin melko vähäinen, kun skaalauseksponentti $\alpha > 1$ ja trendien poistossa käytetään lineaarista sovitetta. Jos ikkunakoko on liian suuri, tuloksiin vaikuttaa todellisen fluktuaation sijasta sovitettavan polynomin yhtenevyys todellisen trendin kanssa. Kun sovite kuvaa hyvin todellista trendiä, ikkunan fluktuaatio pienenee, ja huonosti toimivalla sovitteella fluktuaatio taas kasvaa. Samalla ikkunoiden lukumäärä vähenee, mikä aiheuttaa kokonaisfluktuaation varianssin kasvamisen. Tällöin skaalauseksponentin määrittämisen tarkkuus heikkenee [11]. Skaalauseksponentin tarkkuutta voi parantaa erilaisilla suodatusmenetelmillä, esimerkiksi niin sanotuilla Kalman-suodattimilla, joiden tutkimus on edennyt lupaavasti viime aikoina [12].

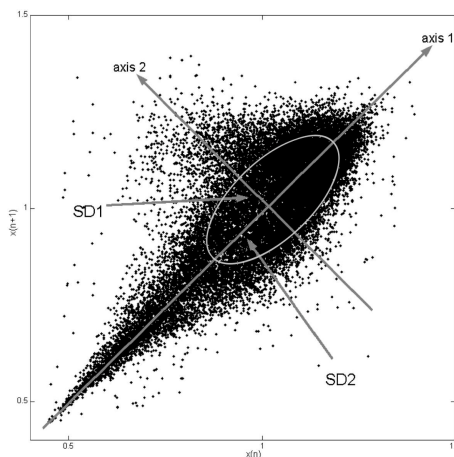
3.2 Alfa-beeta-suodatin

Erityisen pitkän ja kompleksisen aikasarjan ymmärtämiseksi voi olla hyödyllistä selvittää useampia skaalauseksponentteja erisuuruksilla ikkunoiden leveyksillä. Useasti sarjasta lasketaan kaksi eri skaalauseksponenttia, α_1 ja α_2 , joita käytetään selittämään aikasarjan korrelaatioita lyhyillä (α_1) ja pitkällä (α_2) aikaväleillä. Tämän lisäksi on mahdollista laskea jatkuva muutos fluktuaatiolle ikkunan leveyden suhteen. Käyttämällä alfa-beeta-suodatinta DFA:n tuottamiin fluktuaatioiden voimakkuuksiin voidaan muodostaa spektri $\alpha(s)$, jolloin saadaan monipuolisempi kuva systeemissä vaikuttavista korrelaatioista eri skaaloilla s . Menetelmässä muodostetaan

DFA:n flukтуаatioiden logaritmista eri ikkunankoilla $\log F(s)$ jatkuva funktio käyttäen interpolointia. Tämän funktion gradienttia arvioidaan hyödyntäen tasauskerroimia α ja β , joiden avulla voidaan minimoida pienistä oskillaatioista johtuvat virheet. Tuloksena saatu gradientti kertoo skaalauseksponentin arvon jatkuvana funktiona [13].

3.3 Poincarén kartta

Kaikista lyhyimmät eli niin sanotut lag-1-korrelaatiot voidaan visualisoida käyttämällä Poincarén karttaa. Kuvassa 3.1 on esimerkki tavallisesta Poincarén kartasta. Menetelmässä muodostetaan kuvaaja aikasarjasta $\{x_t\}_{t=1,\dots,N}$ kuvaamalla arvot x_t arvojen x_{t+1} suhteen. Mikäli arvot ovat positiivisesti korreloituneet, kuvaaja muodostaa pistejoukon, jonka keskimääräinen arvo kasvaa positiivisen x-akselin suunnassa. Negatiivisen korrelaation tapauksessa pistejoukko on laskeva. Mikäli minkäänlaista korrelaatiota ei ole olemassa, pisteet jakautuvat kuvaajan alueella tasaisesti.



Kuva 3.1 Esimerkki Poincarén kartasta [14].

Korrelaatiolle voidaan laskea myös numeerinen arvo käyttämällä Pearsonin korrelaatiokerrointa. Kerroin r lasketaan kaavalla

$$r = \frac{\text{cov}(x_t, x_{t+1})}{\sigma_{x_t} \sigma_{x_{t+1}}}, \quad (3.3)$$

jossa $\text{cov}(x_t, x_{t+1})$ on sarjojen x_t ja x_{t+1} välinen kovarianssi ja σ_{x_t} ja $\sigma_{x_{t+1}}$ näiden keskihajonta. Korrelaatiokertoimen arvo voi vaihdella välillä $-1 \dots 1$, jossa negatiiviset arvot vastaavat negatiivista korrelaatiota ja positiiviset arvot positiivista korrelaatiota. Kertoimen itseisarvo kuvaa korrelaation voimakkuutta [8].

4. TUTKIMUSAINEISTO

Tutkimus kohdistettiin terveistä sydämistä saatuun EKG-signaaliin. Raat signaalit ladattiin avoimesta tietokannasta, minkä jälkeen näistä muodostettiin tarvittavat RR- ja QT-aikasarjat. Näitä aikasarjoista suodatettiin virheet pois ennen tarkemman analyysin tekemistä.

4.1 Physionet

Tutkimuksessa käytetty data on ladattu yleiseen fysiologiseen tutkimukseen perustetulta internet-sivustolta *PhysioNet* [15]. Sivusto tarjoaa laajan kokoelman erilaisia fysiologisia signaaleja sekä niiden käsittelyyn tarkoitettuja ohjelmistoja. Tässä tutkimuksessa käytetty aineisto on otettu ”The MIT-BIH Normal Sinus Rhythm Database” -tietokannasta [16]. Tietokanta sisältää EKG-nauhoitukset 18 eri henkilöltä. Tietokantaan valitut ihmiset ovat perusterveitä eikä heillä ole sydämen toimintaan vaikuttavia sairauksia. Tietokannan data on tallennettu sydämen sähkökäyrästä digitalisoituna signaalina. Nauhoitusten pituus on kullekin potilaalle noin 24 tuntia ja näytteenottotaajuus 125 hertsiä.

4.2 RR- ja QT-aikavälien laskeminen

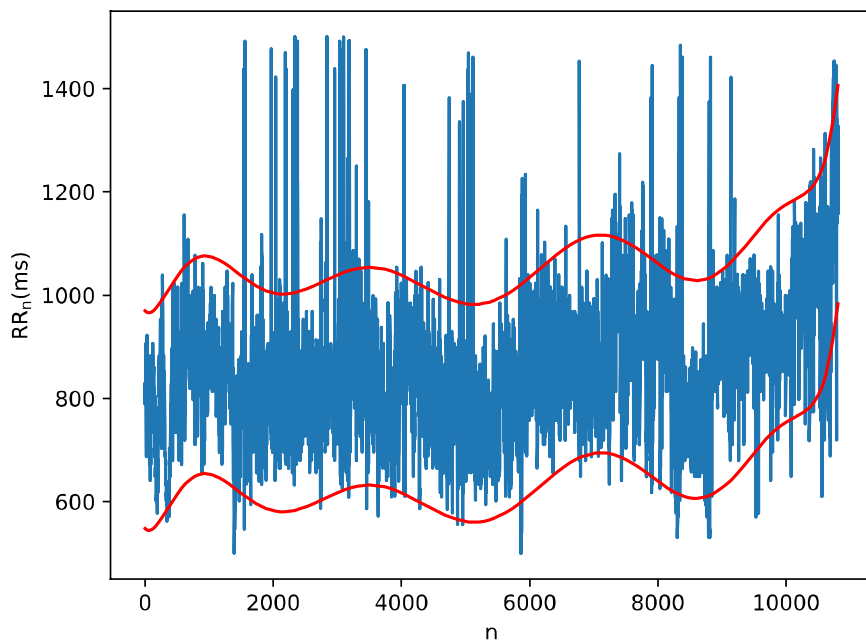
Ennen kuin dataa voidaan käsitellä luvussa 3 käsitellyillä menetelmillä, datasta on määritettävä sen sisältämät RR- ja QT-aikavälit. Tämä on toteutettu käyttäen erityistä PhysioNetin tietokantojen käsittelyyn kehitettyä, ”The WaveForm DataBase” -ohjelmistoa (WFDB) [17]. Ohjelmiston *ecgpuwave*-menetelmää käyttäen on mahdollista ajoittaa digitalisoidusta signaalista sydänsähkökäyrän eri osien alkamisajankohdat. Menetelmä kirjoittaa nämä ajat erilliseen *annotation*-tiedostoon. Tämän jälkeen voidaan käyttää toista WFDB:n sisältämää menetelmää, *rdann*-funktiota, joka lukee ajankohdat *annotation*-tiedostosta ja muodostaa niistä tietoja vastaavan tekstitiedoston.

Tässä työssä kehitettiin ohjelma, joka käy läpi *rdann*-funktion luomat tekstitiedostot. Ohjelma tunnistaa tiedostoista RR- ja QT-aikavälien laskemiseksi tarvittavat

ajankohdat sekä poistaa virheellisesti tunnistetuista aalloista syntyviä häiriöitä. Ohjelma on kirjoitettu Pythonilla ja liite A sisältää käytetyn lähdekoodin. RR-aikavälit lasketaan kahden perättäisen QRS-kompleksin huipun välillä olevasta ajasta ja QT-aikavälit taas saman sydämen lyönnin QRS-kompleksin alusta T-aallon loppuun kuluva ajasta. Muodostetut aikavälit tallennetaan uuteen tekstitiedostoon pysyriiveihin kahteen sarakkeeseen, jolloin ne ovat helposti ladattavissa eri ohjelmien käyttöön aikasarja-analyysin tekemistä varten.

4.3 Datan suodatus

Ennen kuin aikasarjoille voidaan tehdä tarkempaa analyysia, niistä on suodatettava pois mahdolliset virheelliset arvot. Dataan on saattanut vaikuttaa esimerkiksi sähköisiä häiriöitä mittaustilanteessa, jolloin signaaliin on saattanut muodostua ylimääräisiä piikkejä, tai sydämen lyöntejä on saattanut jäädä mittauksessa havaitsematta. Myös signaalin RR- ja QT-aikavälien tunnistamisessa on voinut tapahtua virheitä, jolloin lopullisessa datassa voi olla virheellisesti tunnistettuja sydämen lyöntejä. Nämä näkyvät dataa tutkiessa poikkeuksellisen pitkinä tai lyhyinä aikaväleinä.



Kuva 4.1 Esimerkki RR-sarjan suodatuksesta, jossa punaisten raja-arvojen ulkopuoliset mittauspisteet poistetaan.

Virheellisen datan poistamiseksi RR- ja QT-aikaväleille on asetettu aluksi ylä- ja alaraja, joiden ulkopuolelle jäävät arvot poistetaan datasta. Jäljelle jäävään aikasarjaan sovitetaan sen jälkeen n -astetta oleva polynomi, joka seuraa likimain aikasarjan

arvoja. Arvo n on tyypillisesti 5 tai 10 ja se valitaan sarjan heilahtelevuuden mukaisesti. Tämän jälkeen valitaan rajat, kuinka paljon aikasarjan arvot saavat poiketa sovitteesta. Kaikki ulkopuolelle jäävät arvot poistetaan, jolloin jäljelle jäävä data vastaa paremmin todellista mittausta. Kuvassa 4.1 on yksi esimerkki signaalista ja polynomisovitteesta. Jäljelle jäävään aikasarjaan voidaan sovittaa uusi sovite ja poistaa tämän perusteella lisää virheellisiä tuloksia datasta, mikäli tämä on tarpeen. Lopulta suodatuksen tuloksena saadaan aikasarja, jossa esiintyvät heilahtelut ovat pääosin sydämen ominaisuuksista johtuvia, ja aikasarja sopii käytettäväksi aikasarja-analyysin menetelmissä. Esimerkiksi DFA soveltuu yleensä analyysiin hyvin, vaikka datasta olisi poistettu merkittävä osa (jopa kymmeniä prosentteja) pisteitä.

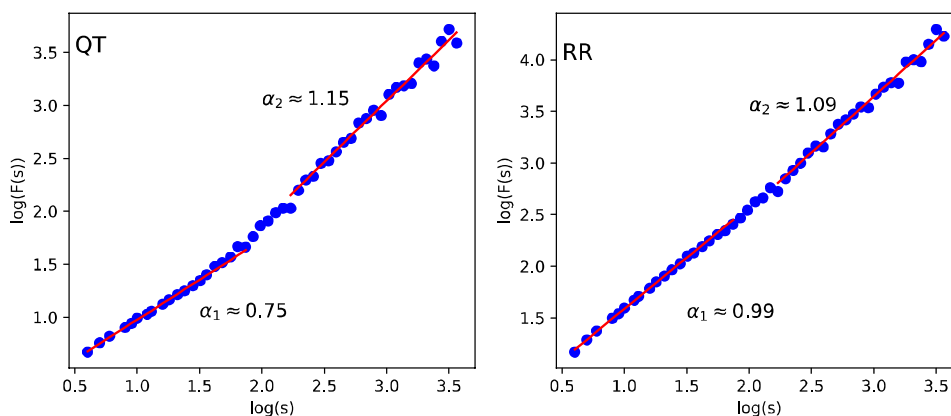
Liiallisella suodattamisella voi kuitenkin olla vaikutusta aikasarjan ominaisuuksiin ja menetelmistä saataviin tuloksiin. Aina kun aikasarjasta poistetaan jokin arvo, sitä ennen ja sen jälkeen olevat sarjat liitetään yhteen. Tästä syntyy aikasarjaan virhe, kun liitoskohdan arvot tulkitaan myös alkuperäisessä sarjassa viereisiksi arvoiksi, vaikka todellisuudessa näin ei ole. Näin ollen jos suodatuksessa sarjasta poistuu liikaa arvoja, ei jäljelle jäävää sarjaa kannata enää käyttää analyysissa. Suodatuksessa perusteella viisi eri signaalia sisälsi niin paljon virheitä, että nämä oli jätettävä tutkimuksen ulkopuolelle, jolloin jäljelle jäi 13 tarpeeksi hyvälaatuista näytettä käytettäväksi analyysissa.

5. TULOKSET

Edellisen luvun mukaista aineistoa on tarkasteltu luvussa 3 selostetuilla menetelmillä. Menetelmiä käytettiin sekä RR- että QT-aikasarjoihin. Tulokset on laskettu käyttäen Pythonia sekä TampeRE Time Series Analysis (TRETSA) -työkalulaatikkoa [18].

5.1 DFA

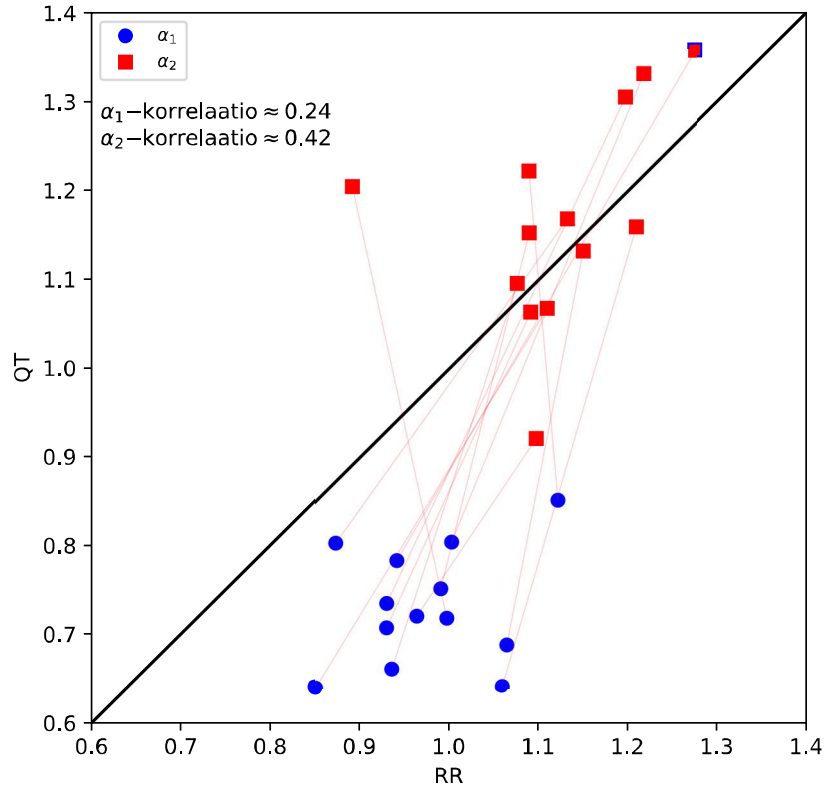
RR- ja QT-aikaväleistä on laskettu DFA:n mukaiset flukтуаatiot eri ikkunaleveyksillä. Kuvassa 5.1 on tyypillinen esimerkki DFA:n tuloksesta. Flukтуаatioiden arvot on sijoitettu kuvaajaan logaritmiselle asteikolle ja datajoukkoon on sijoitettu kaksi eri lineaarista sovitetta. Näistä ensimmäinen kuvaa lyhyen kantaman (<20 lyönnin) ja toinen pitkän kantaman (>50 lyönnin) korrelaatioita. Kuvasta voidaan tehdä monia mielenkiintoisia huomioita aikasarjojen ominaisuuksista. QT:n skaalauseksponentti on lyhyellä kantamalla huomattavan paljon pienempi kuin RR-aikasarjalla. Tämä kertoo vain vähäisistä lyhyen kantaman korrelaatioista QT-aikasarjassa. Pitkällä aikavälillä eksponentin arvo kuitenkin kasvaa, ja on molemmissa aikasarjoissa hieman yli yhden.



Kuva 5.1 Esimerkki DFA:lla lasketuista flukтуаatioista QT- ja RR-aikasarjoille aikaskaalan funktiona logaritmisella asteikolla.

Kuvasta on nähtävissä myös luvussa 3.1 esitellyt DFA:n heikkoudet. Kaikista pienimmillä ikkunaleveyksillä flukтуаatio on hieman lineaarisovitususta pienempää joh-

tuen trendinpoiston vaikutuksesta. Suurilla leveyksillä fluktuaation arvo vaihtelee voimakkaasti peräkkäisten mittausten kesken. Lineaarisovitukseen käytettävien pisteiden suuren lukumäärän vuoksi tämän ei kuitenkaan pitäisi merkittävästi vaikuttaa lasketun skaalauseksponentin arvoon.

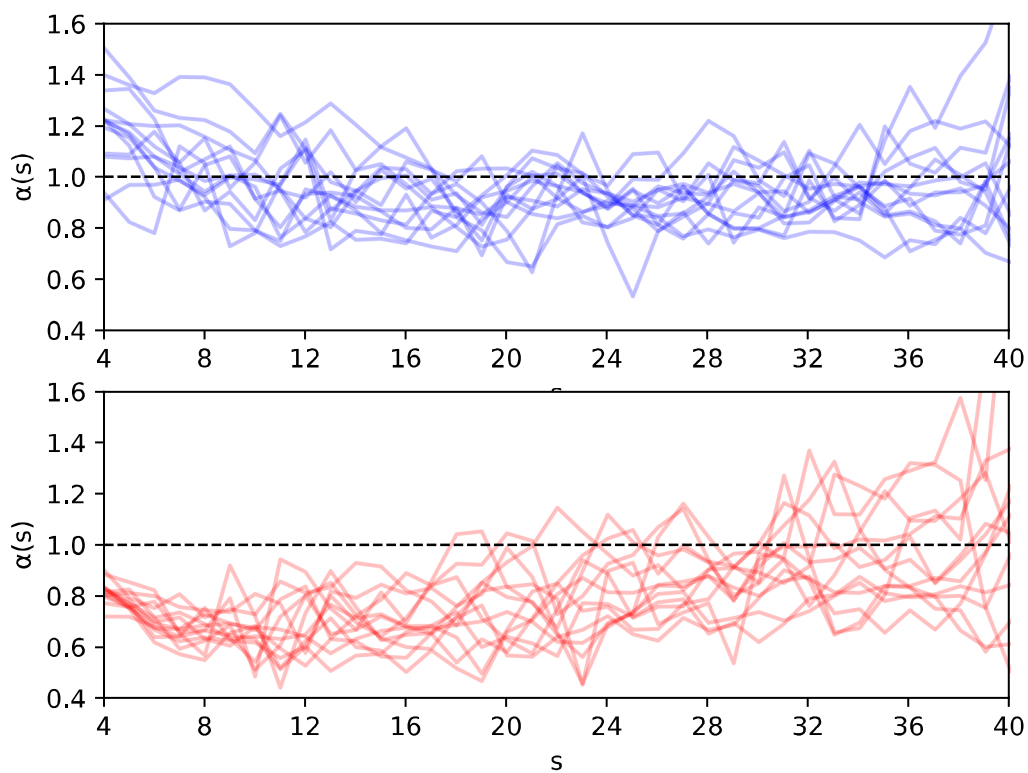


Kuva 5.2 Sydämen RR- ja QT-aikaväliden skaalauseksponentit koko tutkittavalle aineistolle (13 henkilöä).

Skaalauseksponentit on laskettu erikseen jokaiselle aineistossa olevalle aikasarjalle. Näistä tuloksista on muodostettu kuvaaja 5.2, joka havainnollistaa paremmin tyypillisiä arvoja eri skaalauseksponenteille. Lyhyen (ympyrät) ja pitkän kantaman skaalauseksponentit (neliöt) muodostavat kaksi ryhmää jotka ovat selvästi erossa toisistaan. RR-sarjalla on voimakkaammat korrelaatiot lyhyillä aikaväleillä kuin QT-sarjalla, mikä näkyy sinisen joukon sijainnista selvästi diagonaalien alapuolella. Pitkällä aikavälillä sekä QT:n että RR:n skaalauseksponenttien arvot kasvavat ja erot näiden välillä tasoittuvat. Kuvaan on laskettu lisäksi myös korrelaatio aikasarjojen skaalauseksponenttien välillä. Niin lyhyellä kuin pitkällä kantamalla RR:n ja QT:n skaalauseksponenttien korrelaatio on suhteellisen heikko, alle 0,5.

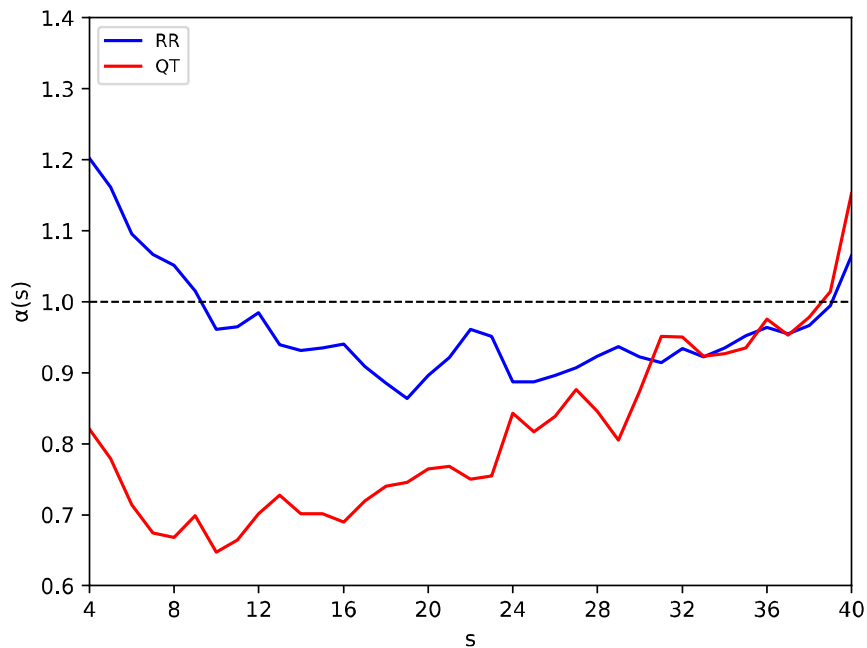
5.2 Alfa-beeta-suodatin

Edellä lasketuista flukтуаatioiden arvoista on muodostettu myös alfa-beeta-suodatuksen mukainen jatkuva kuvaaja skaalauseksponentin arvoista eri ikkunaleveyksillä. Laskennan tulos on esitetty kuvassa 5.3, jossa on nähtävissä skaalauseksponentin muutos sekä RR- että QT-aikasarjoille. Lisäksi kaikkia mittauksia käyttämällä on laskettu keskiarvoistettu skaalauseksponentin muutos, joka kuvaa selvemmin RR ja QT-aikasarjojen välillä olevia eroja. Keskiarvoistettu käyrä on esitetty kuvaajassa 5.4.



Kuva 5.3 Alfa-beeta-suodattimen skaalauseksponentti yksittäisille mittauksille.

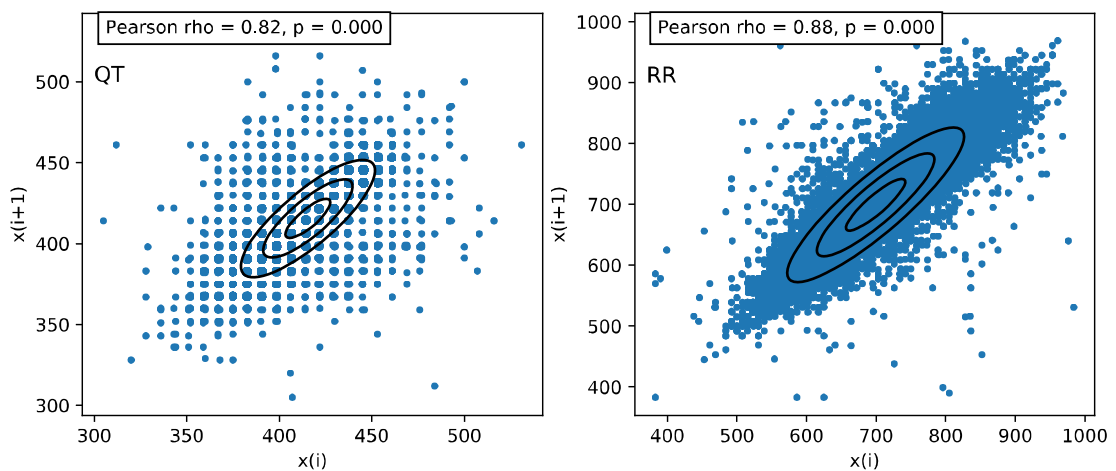
Kuvaajat ovat linjassa luvussa 5.1 tehtyjen havaintojen kanssa. RR-aikasarjan eksponentti on aluksi selvästi QT:n eksponenttia suurempi, mutta ikkunanleveyden kasvaessa molemmat kasvavat lähes samansuuruisiksi. Molemmilla aikasarjoilla tapahtuu merkittävä eksponentin arvon lasku pienimmillä ikkunanleveyksillä. Tämä on todennäköisesti seurausta DFA:n trendien poistamisesta aiheutuvasta virheestä.



Kuva 5.4 Alfa-beeta-suodattimen tuottama jatkuva skaalausekspONENTIN arvo keskiarvoistettuna 13 otoksesta.

5.3 Poincarén kartta

Aikasarjojen välittömiä sisäisiä korrelaatioita on tutkittu muodostamalla niistä Poincarén kartta. Kartat on esitetty kuvassa 5.5, johon on laskettu myös Pearsonin korrelaatiokerroin. Lisäksi karttoihin on lisätty kolme ellipsiä, jotka esittävät mitä arvoja jakauma saa kaikista todennäköisimmin. QT:n kartassa näkyvä ruudukkomaisuus on seurausta käytettävissä olevan aineiston matalasta resoluutiosta.



Kuva 5.5 Esimerkki RR- ja QT-aikaväleistä muodostetuista Poincarén kartoista.

Poincarén kartat tukevat aiempaa tietoa aikasarjojen korrelaatioista. QT-aikasarjan arvot muodostavat karttaan leveähkön ylöspäin nousevan jakauman, mikä kertoo positiivisten korrelaatioiden olemassaolosta. RR:lle jakauma on taas huomattavasti kapeampi, jolloin perättäiset aikasarjan arvot ovat korreloituneet paljon voimakkaammin. Tämä on myös nähtävissä Pearsonin korrelaatiokertoimien arvoista: RR:n kertoimen arvo on selvästi QT:n vastaavaa suurempi.

6. YHTEENVETO

Tässä työssä on tutkittu korreloituneita aikasarjoja ja niiden käyttäytymistä. Tutkimuskohteena on ollut sydämen sähkökäyrästä määritettävät RR- ja QT-aikasarjat. Korrelaatioiden karakterisoimiseksi on käytetty trendit poistavaa flukтуаatioanalyysia (DFA) eri variaatioilla pitkän kantaman korrelaatioihin ja Poincarén karttoja välittömiin korrelaatioihin.

Käytetyt menetelmät soveltuivat hyvin työn tutkimukseen. Menetelmiä on käytetty laskennallisessa fysiikassa aikasarja-analyysiin perustuvissa tutkimuksissa onnistuneesti aiemmin ja niiden ominaisuudet tunnetaan hyvin. Tässä työssä sekä DFA että Poincarén kartat tuovat esille aikasarjojen korrelaatioiden olemassaolon ja niiden voimakkuudet. Menetelmät ovat myös ajallisesti tehokkaita, joten laskenta oli mahdollista toistaa useita kertoja erilaisille aikasarjoille.

Käytetyn aineiston laadussa on joitakin puutteita, kuten matala resoluutio ja pieni otanta, mutta aineisto sopii kuitenkin käytettäväksi työn analyysissa. Aineiston 125 hertsin näytteenottotaajuudella voidaan havaita pienimmillään 8 ms vaihtelu eri aikasarjojen arvoissa. Kaikista pienimmät muutokset aikasarjassa voivat olla tätäkin pienempiä, jolloin osa aikasarjan sisältämästä informaatiosta jää matalan resoluution takia käyttämättä. Lisäksi aineiston rajoittuminen 13 henkilön sydän-dataan voi pienentää tulosten tilastollista merkittävyyttä. Suuremmalla otoskoolla olisi mahdollista tutkia tarkemmin tulosten jakautumista eri yksilöiden välillä. Aineiston hyviä puolia on EKG-nauhoitusten pitkä kesto, jolloin sähkökäyrät sisältävät tietoa sydämen toiminnasta vuorokauden ympäri. Tällöin esimerkiksi kellonajan tai koehenkilöiden erilaisten vuorokausirytmien vaikutus tuloksiin minimoituu.

Työn tuloksena on todistettu terveen sydämen RR- ja QT-aikavälien sisältämien korrelaatioiden olemassaolo sekä laskettu niiden voimakkuus. DFA:ta käyttämällä saatu RR-aikasarjan skaalauseksponentti on likimain yksi niin lyhyillä kuin pitemmilläkin aikaväleillä, mikä osoittaa aikasarjan fraktaalisuuden. QT-aikasarjalle taas lyhyen aikavälin skaalauseksponentti on huomattavasti matalampi, noin $0,7 \dots 0,8$, mikä kertoo selvästi suuremmasta kohinasta kuin RR-sarjalla. Suuremmalla aikavälillä skaalauseksponentti kasvaa ja saa samansuuruisia arvoja kuin RR-aikasarja.

Poincarén karttojen osoittama välitön korrelaatio on kummallekin aikasarjalle positiivinen ja melko voimakas. Kartat muodostavat tiheän ja ylöspäin nousevan pistejoukon erityisesti RR-aikasarjan tapauksessa, mikä viittaa välittömän korrelaation olemassaoloon. Lisäksi kartoista laskettava sisäisen korrelaation voimakkuutta kuvaava Pearsonin korrelaatiokerroin on tyypillisesti korkea, jopa yli 0,8 niin RR- kuin QT-aikasarjoillakin.

Tämän työn tuloksia voi käyttää monien jatkotutkimusten perustana. Eriteltyjä menetelmiä voidaan jatkossa käyttää esimerkiksi erilaisten sydänsairauksien tai lääkkeiden vaikutuksen alaisena olevien sydänten sähkökäyrään ja selvittää, mikäli niistä saadut tulokset poikkeavat jollain tavalla terveen sydämen toiminnasta. Sairauksista kerätystä tiedosta syntyy erityisesti terveysteknologialle potentiaalinen tilaisuus hyödyntää tutkimuksen tuloksia kaupallisesti. Automatisoimalla tutkimuksen prosessit voidaan valmistaa kuluttajille suunnattuja sykemittareita, jotka sykkeen mittaamisen lisäksi tunnistaisivat erilaiset häiriöt sydämessä ja voisivat ohjata käyttäjän jatkotutkimuksiin. Biohakkeroinnin aikakaudella tämänkaltaiset lisäominaisuudet ovat varma kilpailuvaltti yritykselle.

Toinen sydäntutkimuksen kannalta mielenkiintoinen jatkotutkimuskohde on menetelmien käyttäminen yksittäisistä sydänsoluista mittattuihin sähkökäyriin. Tällöin voidaan selvittää mitkä sydämen ominaisuudet ovat olemassa jo soluissa itsessään ja mihin tarvitaan taas solujen välisiä vuorovaikutuksia ja kytkentämekanismia. Tarkempi tieto sydämen toimintaan vaikuttavista tekijöistä auttaa ymmärtämään kuinka erilaiset sydänsairaukset syntyvät ja miten niitä kuuluisi hoitaa.

Kokonaisuutena työtä voi pitää hyvinkin onnistuneena. Laskennan tulokset ovat yhdenmukaisia muiden saman aihealueen tutkimusten kanssa ja QT- ja RR-sarjojen väliset yhteneväisyydet ja erot tulivat selkeästi esille. Menetelmien tehokkuus ja käytökelpoisuus tuli hyvin ilmi ja ne ovat helposti hyödynnettävissä myös tulevilla tutkimuksilla. Lisäksi työssä kehitetty menetelmä ladata ja käyttää PhysioNet-sivuston tietokantoja on osoittanut hyödylliseksi avuksi jatkotutkimuksia ajatellen.

LÄHTEET

- [1] S. H. Liu and L. Pietronero. *Fractals*. *AccessScience*, 2014. <https://doi.org/10.1036/1097-8542.270750>.
- [2] S. Schaefer, D. Levin, and R. Goldman. Subdivision schemes and attractors. In *Proceedings of the Third Eurographics Symposium on Geometry Processing*. Eurographics Association, 2005.
- [3] M. Li. *Fractal Time Series—A Tutorial Review*. *Mathematical Problems in Engineering*, 2010:6, (2010). <https://doi.org/10.1155/2010/157264>.
- [4] R. Perini and A. Veicsteinas. *Heart rate variability and autonomic activity at rest and during exercise in various physiological conditions*. *European Journal of Applied Physiology*, 90(3):317–325, (2003). <https://doi.org/10.1007/s00421-003-0953-9>.
- [5] J. J. Goy, J. C. Stauffer, J. Schlaepfer, and P. Christeler. *Electrocardiography (ECG)*. Bentham Science Publishers, Sharjah, 2013. <http://dx.doi.org/10.2174/97816080547941130101>.
- [6] I. Potapov, J. Latukka, J. Kim, P. Luukko, K. Aalto-Setälä, and E. Räsänen. *Information transfer in QT-RR dynamics: application to QT-correction*. to be published.
- [7] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley. *Fractal dynamics in physiology: Alterations with disease and aging*. *Proceedings of the National Academy of Sciences*, 99(1):2466–2472, (2002). <https://doi.org/10.1073/pnas.012579499>.
- [8] J. Kim, D. Shah, I. Potapov, J. Latukka, K. Aalto-Setälä, and E. Räsänen. *Scaling and correlation properties of RR and QT intervals at the cellular level*. page 7. to be published.
- [9] Y. Mandel, A. Weissman, R. Schick, L. Barad, A. Novak, G. Meiry, S. Goldberg, A. Lorber, M. R. Rosen, J. Itskovitz-Eldor, and O. Binah. *Human Embryonic and Induced Pluripotent Stem Cell-Derived Cardiomyocytes Exhibit Beat Rate Variability and Power-Law Behavior**Clinical Perspective*. *Circulation*, 125(7):883–893, (2012). <http://circ.ahajournals.org/content/125/7/883>.
- [10] J. Kuusela, V. J. Kujala, A. Kiviahho, M. Ojala, H. Swan, K. Kontula, and K. Aalto-Setälä. *Effects of cardioactive drugs on human induced pluripotent*

- stem cell derived long QT syndrome cardiomyocytes. SpringerPlus*, 5(1):234, (2016). <https://doi.org/10.1186/s40064-016-1889-y>.
- [11] S. Haverinen. *Epälineaarinen aikasarja-analyysi ja sykevälivaihtelu*. Master's thesis, Tampere University of Technology, 2017.
- [12] M. Molkkari et al. *Kalman Filter Approaches for Estimating the Scaling Exponent in Detrended Fluctuation Analysis*. to be published.
- [13] J. C. Echeverria, M. S. Woolfson, J. A. Crowe, B. R. Hayes-Gill, G. D. H. Croaker, and H. Vyas. *Interpretation of heart rate variability via detrended fluctuation analysis and filter. Chaos: An Interdisciplinary Journal of Nonlinear Science*, 13(2):467–475, (2003). <https://doi.org/10.1063/1.1562051>.
- [14] A. K. Golińska. *Poincaré Plots in Analysis of Selected Biomedical Signals. Studies in Logic, Grammar and Rhetoric*, 35(1):117 – 127, (2013). <https://content.sciendo.com/view/journals/slgr/35/1/article-p117.xml>.
- [15] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation*, 101(23):e215–e220, (2000). Circulation Electronic Pages: <http://circ.ahajournals.org/content/101/23/e215.full>.
- [16] The Arrhythmia Laboratory At Boston's Beth Israel Hospital (Now The Beth Israel Deaconess Medical Center). *The MIT-BIH Normal Sinus Rhythm Database [Data set]*. (1990). physionet.org. <https://doi.org/10.13026/c2nk5r>.
- [17] I. Silva and G. B. Moody. *An Open-source Toolbox for Analysing and Processing PhysioNet Databases in MATLAB and Octave. Journal of Open Research Software*, 2(1):e27, (2014).
- [18] TampeRE Time Series Analysis toolbox. <https://gitlab.com/qcad.fi/tretsa>. Viitattu 2.10.2018.

LIITE A. RR- JA QT-ARVOJEN LASKEMISEEN KÄYTETTY PYTHON-KOODI

```
RAW_DATA_DIRECTORY = "raw_data/"
EXTRACT_DATA_DIRECTORY = "extraced_series/"

def extract_qt_rr(filename, minlength):
    data = [] # data consists of qt1, rr, qt2 and if skipped,
               ↪ skip_interval

    tempqts1 = []
    tempqts2 = []
    temprrs = []

    skip_start = None
    skip_end = None
    skip_interval = None
    max_skip = 0
    max_rr = 0

    total_waves = 0
    total_Ns = 0

    qt_start = None
    rr_start = None

    p_found = False
    N_found = False
    t_found = False
    error_happened = False

    file = open(RAW_DATA_DIRECTORY + filename + "qt.txt", 'r')
    for line in file:
        line = line.strip()
        mark = read_wave_mark(line)

        if mark == "p":
            if not (N_found or t_found):
```

```
        p_found = True
    else:
        error_happened = True

elif mark == ")":
    if p_found and not (N_found or t_found):
        qt_start = read_time(line)
    elif p_found and N_found and t_found:
        time = read_time(line)
        qt_int = calc_diff(qt_start, time)
        tempqts1.append(qt_int)
        if len(tempqts1) > 1:
            tempqts2.append(qt_int)
        p_found = False
        N_found = False
        t_found = False
        qt_start = None
        if len(tempqts2) == 0:
            skip_end = time
        if len(tempqts2) == minlength and skip_start is not
            ↪ None:
            skip_interval = calc_diff(skip_start, skip_end)
            if skip_interval > max_skip:
                max_skip = skip_interval
        if len(tempqts2) >= minlength:
            skip_start = time

    elif not p_found or (t_found and not N_found):
        error_happened = True

elif mark == "N":
    total_Ns += 1
    if p_found and not(N_found or t_found) and qt_start is not
        ↪ None:
        N_found = True
        time = read_time(line)
        if rr_start is not None:
            temprrs.append(calc_diff(rr_start, time))
        rr_start = time
```

```
        else:
            error_happened = True

    elif mark == "t":
        if N_found and p_found:
            t_found = True
        else:
            error_happened = True

    if error_happened:
        total_waves += len(tempqts2)
        if len(tempqts2) >= minlength:
            max_temp_rr = max(temprrs)
            if max_temp_rr > max_rr:
                max_rr = max_temp_rr
            if skip_interval is not None:
                data[-1] = data[-1] + " " + str(skip_interval)
            for qt1, qt2, rr in zip(tempqts1, tempqts2, temprrs):
                row = str(qt1) + " " + str(rr) + " " + str(qt2)
                data.append(row)

        tempqts1 = []
        tempqts2 = []
        temprrs = []

        qt_start = None
        rr_start = None

        p_found = False
        N_found = False
        t_found = False
        error_happened = False
        max_temp_rr = 0

    if len(tempqts2) >= minlength:
        for qt1, qt2, rr in zip(tempqts1, tempqts2, temprrs):
            row = str(qt1) + " " + str(rr) + " " + str(qt2)
            data.append(row)

    total_waves += len(tempqts2)
```

```
file.close()
print("Patient:", filename, "length: %5d Waves total: %5d N total
      ↪ : %5d, Largest skip: %8d Max rr: %8d" %\
      (len(data), total_waves, total_Ns, max_skip, max_rr))
return data

def read_wave_mark(line):
    if line[15] == "1":
        return line[30]
    else:
        return line[28]

def read_time(line):
    return int(line[10:13]) + 1000 * int(line[7:9]) + 60000 * int(
        ↪ line[4:6])

def calc_diff(time1, time2):
    res = time2 - time1
    if res < 0:
        res += 3600000
    return res

def write_qt_rr(filename, data):
    file = open(EXTRACT_DATA_DIRECTORY+ filename +"qtrr.txt", "w")
    for d in data:
        file.write(d + "\n")
    file.close()

def main():
    names = open("names.txt", "r")
    for name in names:
        name = name.strip()
        data = extract_qt_rr(name, 10)
        write_qt_rr(name, data)

main()
```