

Tiivistelmä: Tfy-99.3275 Biosignal processing

Vastuu lukijalla. Mielipahaa ei korvata.

Syksy 2011

Kirjattuja muistiinpanoja luentojen ja luentokalvojen perusteella.

Mika Mäntykangas

LUENTO 1: Johdanto biosignaaleihin

-signaali on jokin informaatiota siirtävä aineen tai bittijonon ominaisuus, joka on kiinnitetty yleisimmin aikaan $f(t)$ tai toisinaan esimerkiksi paikkaan $f(\mathbf{x})$.

-biosignaalit ovat peräisin jostain biologisesta järjestelmästä, tässä kurssissa ihmisestä.

-biosignaalien käsittely (prosessointi): kaikki biosignaalien käsittely, joka tapahtuu niiden synnyinsijan (elimistö, mittauslaitteisto) ja saatujen mittaustulosten välillä (datankäsittely).

-biosignaalien käsittelyssä käytetään yleisiä signaalikäsittelyn menetelmiä, mutta niitä sovelletaan biologisiin signaaleihin.

-biosignaaleja voidaan käsitellä seuraavasti: joko tutkitaan ilmiötä joka on jo tiedossa ja saadaan karakterisoitua tätä TAI tutkitaan signaalia ja löydetään sieltä jotain mielenkiintoista (tutkimustyyppinen tai katsoa miten jokin ilmiö/tekeminen (esim. lihasvaikutus, stressi yms) vaikuttaa johonkin muuhun ominaisuuteen).

-datan kerääminen ei niinkään ole haasteellista, mutta sen muuttaminen ensin informaatioksi (merkityksellinen data) ja sitä tiedoksi (knowledge, asioiden ymmärryksen parantuminen) ja sovelluksiksi (utility, eli tiedon käyttökohteet) on.

-joitain kohteita (fundamentaaleja lajeja joissa biosignaalien prosessointia käytetään): TIEDONHAKU (biologisen systeemin ymmärtäminen), DIAGNOSOINTI (vikojen, sairauksien, epäjohdonmukaisuuksien tunnistaminen), MONITOROINTI (jatkuva-aikainen tiedonkeräys ja analysointi), KONTROLLOINTI (tiedon perusteella nykyisen tilan muuttaminen eli tietynlainen takaisinkytkentä biologiseen järjestelmään), DATANKÄSITTELY (datan objektiivinen vertailu muihin mahdollisuuksiin, laadunvalvonta ja tarkastelu, mahdollisen hoidon hyödyt ja haitat).

-biosignaaleja voidaan luokitella monilla eri tavoilla: signaalin OMINAISUUDET (etenkin raa-kadatassa, esimerkiksi sähköinen suure, mekaaninen suure, taajuuskäyttäytyminen, satunnaisuus), SIGNAALILÄHTEEN suhteen (eli mistä signaali saa alkunsa, esimerkiksi sydänperäinen, hengitysar artefaktit, erilaiset rytmiset käyttäytymiset, solujen vai organismin toiminta), SOVELLUSKOHTEET (jatkuva-aikainen monitorointi, rutiinitarkastukset, urheilusuorituksen seuranta, eli eri kohteilla on erilainen merkitys).

-biosignaaleja on seuraavia päätyyppejä: BIOELEKTRONISET signaalit (johtuvat neuronien ja lihassolujen sähköisestä käyttäytymisestä) ja BIOMAGNEETTISET signaalit (periaatteessa sama syntypaikka, mutta magneetikentän muodossa, voi antaa lisäinformaatiota sähköisille

ilmiöille esim. aivotutkimuksessa), BIOIMPEDANSSISET signaalit (sähkönjohtavuus antaa tietoa solujen ja kudoksen ominaisuuksista ja esimerkiksi veren määrästä ja jakautumisesta), BIO-AKUSTISET signaalit (esimerkiksi veren virtaaminen luo äänen ja turbulenssilla virtauksella on usein negatiivinen merkitys. tämä voidaan kuulla. myös keuhkojen toimintaa voidaan kuunnella), BIOMEKAANISET signaalit (liike ja siihen liittyvät jännitystilat elimistössä, verenkiertoelimistön paineenvaihtelut, keuhkopaine), BIOKEMIAALLISET signaalit (ionikonsentraatiot ja niiden vaihtelun merkitys biologisissa järjestelmissä, soluissa, veressä jne.), BIO-OPTISET signaalit (esimerkiksi veren hapettumisasteen mittausta voidaan nähdä tietyillä valon aallonpituuksilla), MUUT signaalit (esimerkiksi hoitohenkilökunnan silmämääräiset huomautukset silmämääräiset muutokset henkilöiden käyttäytymisessä, väsyneisyysaste jne.).

-signaalien luokittelu onnistuu myös seuraavasti: SATUNNAISUUS (käytetään myös nimitystä stokastinen signaali, termisestä kohinasta satunnaisiin ihmisen päätöksiin), HETKELLISET signaalit (kestävät vain tietyn ajan, lihaksen jäykistyminen), SÄÄNNÖLLISET signaalit (näillä on aaltomuoto, joka toistuu "loputtomasti", sydämen syke).

-yleisiä lainalaisuuksia, joita biosignaaleja mittaavan laitteen tulee ottaa huomioon: mittarin tulee olla kalibroitu halutulle mittausalueelle, sen tulee näyttää mitattavaa kohdetta tarpeeksi suurella nopeudella ja pystyä seulomaan mielenkiintoinen data kohinasta ja muusta sisällöstä. lisäksi siinä tulee olla filtereitä ja sen verran älyä, että biologisesti mahdottomia tapahtumia ei tulkita tapahtumiksi (kahta sydämenlyöntiä ei merkitä 1 ms välein eli tapahtumilla tulee olla tietyt karensialueet). nämä pohjautuvat statistiseen informaatioon. signaalinkäsittelyssä sen tulee erottaa halutut piirteet signaalista ja tunnistaa haluttuja lainalaisuuksia. lopuksi se mallintaa mahdollisimman yksinkertaisesti saadun "arvon" mallina jolla on tunnettu skaala (esimerkiksi luku välillä 0–100).

-EEG-signaali on hyvä esimerkki signaalista, joka voi näyttää kaoottiselta (random, satunnainen, vaikkakin kaikille signaalille lienee aivoissa perusta ja tarkoitus, se vain on niin monimutkainen), mutta josta löytyy paljon lainalaisuuksia: esimerkiksi erilaiset aivoaaltoaajuuksien alueet (alfasta gamma-aaltoihin) ja intensiivinen ajattelutyö. näiden aivoaaltojen taajuuksien amplitudisuhteet kertovat kulloisestakin tilasta ja esimerkiksi unen syvyyttä voidaan mitata. aivosähkökäyrä on erilainen riippuen missä unen tasossa ihminen on.

-toisinaan signaalien mittaamisessa on aukkoja, eli signaalia ei ole näytetty tasavälisesti. tämä on usein ongelma ja varsinkin silloin, kun kaikkia haluttuja signaaleja ei ole näytetty samanaikaisesti. ongelman ratkaisemiseksi tulee aukot täyttää esimerkiksi keskiarvoistamalla mittauspisteitä puuttuvien ympärillä olevien avulla.

-biosignaalien prosessointi tulee hiljalleen yhä lähemmäs ja lähemmäs ihmisten omaakin arkea, sillä mittalaitteet integroituvat esimerkiksi älypuhelimiin. sykemittarit ovat jo kauan olleet urheilijoiden käytössä. lisäksi biologisiin signaaleihin yhdistetään paikkatietoa, käyttäjien antamaa palautedatata, kellonaika, tottumuksia (esim. ruokailutottumukset) ja ne voidaan helposti jakaa eteenpäin sosiaalisissa verkoissa ja mahdollisesti lääkäreiden käyttöön.

-ongelmia biosignaalien datan analysoinnissa: KÄYTTÖOIKEUDET (käyttäjäturvallisuus ja yksityisyys), KÄYTTÖTURVALLISUUS (tutkittavan turvallisuus, esimerkiksi joitain mittauksia ei voida tehdä ilman radioaktiivisia varjoaineita), VARIAATIO (biologiset järjestelmät ovat mitä suurimmassa määrin yksilöllisiä, joten vaihtelut ovat suuria), RISTIKKÄISET KYTKÖKSET MUUHUN ELIMISTÖÖN (mitattu suure voikin olla peräisin jostain aivan muualta tai muusta syystä), MITTAUSJÄRJESTELY (mittausjärjestelyn tulisi olla mahdollisimman huomaamaton eikä saa vaikuttaa itse mitattavaan suureeseen), DATANKÄSITTELY (jokaisella tutkijalla voi olla myös mielipide siitä mitä datasta halutaan nähdä jolloin se ohjaa mittausjärjestelyä enemmän kuin itse löydökset, lisäksi datan tuotto on yhteistyötä, joka ei aina onnistu, kaikkia mittauksia ei myöskään tehdä standardisoidusti, jotta muut voivat sen samanaikaisesti toistaa).

-biosignaalien prosessoinnissa on kyse laajan matemaattisen signaalianalyysityökalupakin käytöstä laajaan määrään käyttökohteita aina välillä tietämättä mitä halutaan itse asiassa nähdä. se tekee haastavaksi keksiä mitä tekniikkaa oikeastaan kannattaa kulloinkin käyttää. valitse siis oikea tutkimusmetodi ongelmaan (tämän valitsemiseen kannattaakin käyttää aikaa).

-tyypillinen signaalianalysointiprojekti koostuu seuraavista vaiheista: SUUNNITTELE JA TEE TUTKIMUS (tarkka kohdentaminen mitä ja millä järjestelyllä halutaan mitata, useita pilottimittauksia, informatiivisen datan kerääminen (eli merkityksellisen datan)), ANALYSOI SIGNAALIT (löydä mitatun signaalin ja halutun ilmiön väliset lainalaisuudet), SUUNNITTELE ALGORITMI (tee algoritmi, jolla voit nopeasti saada mitatusta datasta halutun suureen esiin ja mahdollisimman yksinkertaiseen muotoon myös muiden käyttöön), SOVELLA ALGORITMI (siirrä algoritmi johonkin oikeaan mittaussuureeseen, jotta siitä saadaan mahdollisesti reaaliaikainenkin vaste, käytä järjestelyä ja algoritmia oikein).

-ongelma on aina joko SUORAVIIVAINEN (forward problem, tiedetään ilmiö ja halutaan mitata se) tai EPÄSUORA (backward problem, käytössä on mittaussensori, mutta tutkitaan mitä sillä on saatu jotain ärsykettä vastaan mitattua).

-ennen signaalin syöttämistä algoritmiin tehdään usein seuraavia välivaiheita: mitataan haluttu ilmiö jollakin suureella (esitetty edellä), käytetään vahvistimia ja suodatusta (biologiset signaalit ovat usein melko heikkoja ja suodatus on pakollinen signaalikäsittelyssä korkeiden taajuuksien poistamiseksi), tehdään digitaalimuunnos (A/D-muuntimen käyttö, käytännössä on pakko hyödyntää tietokoneiden laskentatahoa) ja esiprosessoidaan data (pätkitään, hyväksytään käsiteltäväksi, on esimerkiksi turhaa ottaa käsittelyyn mittaustietoa missä tiedetään ettei ole mitään). vasta näiden vaiheiden jälkeen on mielekästä syöttää mitattu data käsiteltäväksi luotun algoritmiin.

-A/D-muunnin tekee signaalille hieman hallaa: se näytteistetään ja näytepisteistä tulee diskreettejä (ei-jatkuva pisteistö) ja kvantisoituja (ei voi saada mitä vain arvoa, esimerkiksi vain kokonaisluvut voivat olla hyväksytyjä).

-signaalinkäsittelyyn vaikuttaa valtavasti miten tiheään se on näytteistetty. mitä tiheämpi näytteistys, sitä korkeampia taajuuksia voidaan luotettavasti näyttää (puolet näytteistystaajuuksista). täytyy kuitenkin tehdä ero mielekkään taajuusalueen esityksen, datan määrän ja käytettävän laskentakapasiteetin suhteen.

-kvantisoinnissa syntyy virheitä, jolle voidaan esittää kvantisointivirhe. tämä johtuu siitä, että esimerkiksi 12-bittinen järjestelmä voi antaa ulos vain $2^{12} = 4096$ eri arvoa, mutta oikea arvo on jotain muuta (jonka suurempikaistanleveysjärjestelmä ehkä voisi antaa tarkemmin. virheen hyväksyminen on aina kompromissi ja välttämätön).

-datasta voidaan vähentää pisteitä (decimation) ja lisätä pisteistöä (interpolation), mutta etenkin jälkimmäinen harvoin on mielekästä tai hyväksyttävää, koska se ei voi tuoda lisää informaatiota systeemistä.

-Shannonin näytteenottoteoreeman mukaan jatkuva signaali voidaan täysin rekonstruoida näytteistetyistä signaaleista, mikäli näytteenottotaajuus on vähintään kaksi kertaa niin paljon kuin korkein signaalissa esiintyvä taajuus eli Nyquist-taajuus. mikäli ehto ei toteudu (signaali pitää alipäästö-suodattaa!), syntyy laskostumista (aliasing) eli muodostuu signaaliin kuulumattomia taajuuksia, sillä spektri kertautuu näytteenottotaajuuden välein (yleensä tuottaen korkeille taajuuksille ylimääräisiä piikkejä).

-vääränlainen näytteistäminen on aina mahdotonta korjata signaalista jälkikäteen. on vaikeaa myös olla ottamatta näitä signaaleja omassa analyysissä huomioon (etenkin jos ne ovat muodostaneet harmoneja muiden signaalien kanssa ja kertautuvat useaan paikkaan spektrissä).

-systeemi on lineaarinen silloin, kun ulostulo on sama riippumatta siitä syötetäänkö sisääntuloon yksi kahdesta eri vakiolla kerrotusta signaalista koostuva kokonaissignaali vai nämä kaksi signaalia summataan ja kerrotaan vakioillaan vasta ulostulon jälkeen.

- aikainvariantti systeemi voidaan muuttaa mihin tahansa mittauspisteeseen kertomalla se aika-riippuvasti z -muunnoksella z^{-m} , missä m on ajanhetki jolla mittaus on tehty. systeemi ei siis muuta toimintaansa ajassa.
- kausaalisissa systeemissä järjestelmän ulostulo riippuu ainoastaan siihen syötetystä datasta (joko vain senhetkisestä tai sitten myös aikaisemmista, molemmissa tapauksissa signaalin ulostulo on rajoitettu ajassa kunhan systeemiin syötetään äärellinen signaali).
- stabiililla järjestelmällä on aikarajoitettu ulostulo mikäli myös sisäänmeno on aikarajoitettu (esimerkiksi impulssivaste on rajoitettu stabiilissa järjestelmässä).

LUENTO 2: Perus signaalikäsittelymenetelmät, aika- ja taajuus-alue, suodattimet ja tehospektri

- signaalit esitetään yleensä aika-alueessa, jolloin se on näytteistykseen jälkeen heti käsiteltävissä ja antaa yleensä parhaiten havainnollistettavan kuvan signaalin merkityksestä.
- signaali voidaan muuntaa myös taajuusalueeseen, jonka perusteella voidaan tehdä päätelmiä. yleensä toistuvat ilmiöt saadaan paremmin esiin ja karakterisoitua näin, kohinan osuus pienenee visuaalisessa esityksessä ja monet laskennalliset tehtävät helpottuvat taajuusalueessa (esimerkiksi aika-alueen konvoluutio vastaa taajuusalueessa huomattavasti yksinkertaisempaa kertolaskua).
- näytteistyksestä johtuen tulee taajuusaluemuunnoksessa hyödyntää diskreettiä Fourier-muunnosta. Fourier-muunnos on tehokas työkalu erityisesti sinimuotoisille signaaleille taajuuksien läpikäymiseen (eksponenttitermi purkautuu trigonometrisiksi funktioiksi) ja siitä saadaan myös vaiheinformaatio.
- Fourier-muunnoksessa (tai taajuusalueen esitystä etsittäessä) on tärkeää huomioida Shannonin teoreema ennen muunnosta, koska muunnos ei tätä tietenkään voi huomioida erikseen.
- monet mielenkiintoiset ilmiöt liittyvät taajuusalueen esityksen muutoksiin. nämä tietenkin pohjautuvat aika-alueeseen, mutta vasta ikkunoimalla aika-alueella ja tekemällä useita taajuusanalyysyjä voidaan helposti esittää muutokset, joita on vaikea aika-alueesta nähdä.
- DFT:n (diskreetti Fourier-muunnos) vaikutuksia: JAKSOLLISUUS taajuusalueessa (spektri kertaantuu näytteenottotaajuuden välein), VUOTAMINEN (mikäli esimerkiksi kanttipulssi (rect-pulssi) muutetaan taajuustasoon, syntyy ääretön spektri (ei tietenkään mikäli näytteistetään, mutta laskostumista tapahtuu)), AIKA-ALUEEN JAKSOLLISUUS (käänteinen ominaisuus taajuusalueen jaksollisuudelle, mikäli esimerkiksi taajuusalueen esitys on tasajakautunut kaikkialle, saadaan aika-alueessa impulssijono).
- vuotaminen seuraavalle spektrin näytteenottomoninkerralle yritetään vähentää erilaisilla ikkunointifunktioilla. yleisimmät näistä ovat Bartlettin, Hanningin ja Hammingin ikkunat). yhteistä näille on näytteistysikkunan reunoilla matalat kertoimet sisääntulovalle aika-alueen signaalille, jolloin signaali voidaan ajatella äärettömän pitkäksi samoista ikkunoista koostuvaksi (toisin kuin pelkkä suorakaideikkuna, missä signaali tuskin koskaan on jatkuva tai ainakaan derivoituva ikkunan reunoilla).
- nopea Fourier-muunnos (FFT, fast Fourier transform) on tietokonealgoritmi DFT:n laskemiseksi, jolla säästetään valtavasti laskenta-aikaa, kun laskennan välivaiheita voidaan vähentää samojen välivaiheiden huomioimisen ansiosta. FFT käytännössä vaatii, että näytteiden määrä on jokin kakkosen potenssi.
- Fourier-muunnos tehdään kaikille ikkunoille aikatasossa, jolloin saadaan muutokset esiin. ikkunointia halutaan yleensä myös pistää osittain päällekkäin (overlap).

-suodattimia on kahdelaisia: FIR-suodattimet (finite impulse response, impulssivaste on rajattu ja ulostulo filterissä riippuu ainoastaan mittauspisteistä itsestään eli vain nykyisistä tai mahdollisesti myös aikaisemmista mittauksista) ja IIR-suodattimet (infinite impulse response, impulssivaste on aikatasossa ääretön, systeemin siirtofunktio riippuu signaalin itsensä lisäksi myös aikaisemmista suodattimen antamista arvoista, joten vaikka sisääntulossa ei olisi mitään, filteri jatkaa arvojen luomista "itsenäisesti").

-suodattimesta voidaan numeerisessa signaalinkäsittelyssä luoda alipäästötyyppinen esimerkiksi keskiarvoistamalla aikaisempia mittauspisteitä. tällöin nopeat muutokset poistuvat. jos taas summataan vierekkäisten mittauspisteiden eroja keskenään, saadaan korkeat taaajuudet esille ja suodatin on ylipäästötyyppinen.

-lineaarisen aikainvariantin systeemin (LTI system) voi karakterisoida siirtofunktiolla $H(f)$, mistä voidaan johtaa amplitudivaste itseisarvona $|H(f)|$ ja vaihevasteena $\theta(f)$.

-LTI-systeemin siirtofunktioesitys perustuu z-muunnokseen, missä ulostulofunktio $Y(f)$ jaetaan sisääntulofunktiolla $X(f)$.

-yleensä on mielekästä muuttaa amplitudivasteet logaritmiselle asteikolle, joka määritellään siirtofunktiolle $20\log_{10}(H(f))$. mikäli kyse on tehosta, käytetään kertoimena kymmentä. logaritmi on aina kymmenkantainen.

-signaalinsuodatuksen perusmallit ovat ali-, yli-, kaistanesto- ja kaistanpäästösuodattimet. lisäksi määritellään päästökaista (maksimissaan -3 dB:n vaimennus reunalla), siirtokaista ja estokaista.

-liian kapea tai asteluvultaan korkea kaistanestosuodatin alkaa käyttäytyä hyvin omituisesti eikä se ole enää stabiili halutulla alueella. tämän vuoksi suodattimen ominaisuudet pitää aina mallintaa.

-mikäli taajuusvaste on lineaarinen, ei synny vaihemuutoksia ja kaikki signaalit viivästyvät saman verran. tämä on merkityksellistä mikäli vaiheella on systeemiä analysoitaessa merkitystä (usein halutaan laskea mikä on eritaajuisten aaltojen ajallinen ero tai muu aikataason analyysi, jota on mahdotonta (tai vaikeaa) tutkia epälineaarilla vaihevasteella). voidaan myös luoda suodatin, millä vain haluttu taajuusalue on vaihevasteeltaan (lähes) lineaarinen (parhaimmassa tapauksessa vaakasuora).

-erilaisilla suodatinperheillä on erilaisia ominaisuuksia. yleisimmin käytössä on Butterworth-, Hamming- tai Bessel-suodatin. näillä on eroa esimerkiksi esto- ja päästökaistan aaltoilussa ja siirtokaistan leveydessä. käyttö riippuu halutuista ominaisuuksista.

-FIR- ja IIR-filtterin vertailua: FIR-filttereillä on lineaarinen vaihevaste ja ne ovat aina stabiileja (tästähän FIR-sana tulee). lisäksi ne ovat usein päästökaistan reunoilla tarkempia ja laskenta on nopeampaa (samalla asteluvulla). kuitenkin FIR-filtterit tulee tehdä korkeammalle asteluvulle kuin vastaava IIR-filteri.

-kaksisuuntainen filteröinti voidaan tehdä vain offline-prosessoinnissa. siinä filteröinti tehdään ensin ajassa myötäsuuntaan, mutta lisäksi takaperin. tuloksena on täysin vaihevääristymätön signaali kaikilla lineaarisilla filtereillä (myös IIR), mutta amplitudispektri on neliöitynyt eli kertautunut itsellään (tehospektri).

-tehospektri määritellään autokorrelaatiofunktion Fourier-muunnokseksi. tämä on helppo ymmärtää seuraavasti: valkoisessa kohinassa energia on jakautunut tasaisesti kaikille taajuuksille. kun kohinalle tehdään autokorrelaatio, saadaan yksi ainut nollasta poikkeava arvo kohdassa $t=0$. impulssin Fourier-muunnos taas on vakio koko taajuustasossa - siis energia jakautuneena kaikille taajuuksille.

-tehospektriä (PSD, power spectrum density) voidaan arvioida ei-parametrisillä tai parametrisillä metodeilla: Blackman-Tukey (ei-parametrinen, laske PSD määritelmään perustuen mutta käytä äärellistä aikaa ja korrelaatiofunktion miniminormiestimaattia), Welchin metodi (ei-parametrinen, aikakeskiarvoistettu Fourier-muunnoksen keskiarvo antaa tuloksen ilman auto-

korrelaatiota), AR-mallit (autoregressive, parametrinen, luotettavampi metodi, missä annetut parametrit (kunhan luotettavia ja soveltuvia) mallintavat aika-alueen ikkunan mittauspisteiden lineaarikombinaatiota, jolloin ei tarvita tietoa systeemin stationaarisuudesta).

-aivan kuten lukujonolle yleisesti, taajuustasossa signaalin spektrille voidaan laskea mediaanitaajuus, keskimääräinen taajuus, ilmoittaa korkeimman amplitudin taajuus ja korkein mitattu (ei kohinasta johtuva) taajuus (SEF, spectral edge frequency, voidaan ilmaista myös esim. SEF90, eli taajuus jota pienempiä vastaa 90 % taajuusjakaumasta).

-aikatasosta ikkunoinnilla laskettujen taajuusesityksien tehospektien kuvaaminen kolmessa ulottuvuudessa (kolmas akseli aika) voi tuottaa mielenkiintoista lisäinformaatiota. tämän voi tehdä myös esimerkiksi spektrogrammilla, missä käytetään värin luminanssia eri taajuuksien voimakkuuden erottimena tai CDSA:lla tämän (colour-density spectral array) saman ilmoittamisessa eri väreillä.

-ei-parametristen PSD-estimointien ominaisuuksia: laskennallisesti helppo, mutta estimaatit eivät ole statistisesti luotettavia. datan ikkunoinnin vuoksi taajuusalue on rajatumpi ja ikkunointi rajaa monia muitakin ikkunan ulkopuolisia ilmiöitä pois laskuista. parempia malleja on kuitenkin olemassa (esimerkiksi parametriset AR-mallit).

LUENTO 3: Biosignaalien laadun parantaminen artefaktien ja kohinan eliminoimisella

-signaalien esiprosessoinnilla tarkoitetaan laaduntarkastusta, missä katsotaan että signaali on silmämääräisesti halutunlaista, siitä poistetaan liian korkeat taajuuskomponentit, vähennetään kohinaa sekä etsitään ja poistetaan selkeät artefaktit.

-suodattaminen voidaan tehdä ennaltamäärätysti optimaalisen tuloksen saamiseksi (sopii stationaarisille signaaleille, joiden ominaisuudet eivät muutu ajassa) tai aikamuuttuvasti (adaptoivasti, mukautuvasti, tarvitaan ei-stationaarisille eli muuttuville signaaleille) optimaalisen suodatuksen saamiseksi kullekin signaalille, jonka ominaisuuksia ei ennalta tunneta.

-kohinanpoisto on helpointa mikäli suurin osa siitä on mielenkiintoisten taajuuksien ulkopuolella. muuten tullaan tilanteeseen, missä kohinanpoisto hävittää myös osan signaalista.

-erilaisia signaalinsuodatuksen optimointikriteerejä voidaan käyttää ja tällöin päädytään johonkin optimointiongelmaan, tarkemmin sanottuna minimointiin tai maksimointiin. seuraavia ainakin on käytössä: neliöllisen keskivirheen minimointi (kun halutaan signaali mahdollisimman puhtaana, johtaa Wienerin suodattimeen), signaali-kohinasuhteen maksimointi (signaalin muodot saadaan näin esiin, johtaa matched filter -suodattimeen, jota käsitellään seuraavassa luennossa).

-Wienerin suodattimessa tarvitaan ennalta tietoa signaalin tehospektristä niin saadussa signaalissa x kuin fysiologisessa signaalissa s . ideana on, että optimaalinen suodattimen siirtofunktio tehdään PSD:n taajuusalueen jakolasku $s:n$ ja $x:n$ välillä, mikä kertoo korrelaatiosta kullakin taajuudella. jos korrelaatio on vahva, signaali pääsee läpi kertoimella yksi, mutta huonosti kolloivat signaalit suodattuvat pois. toisin sanoen, mikäli saadussa signaalissa x on kohinaa vaikka näin ei tulisi olla fysiologisessa mielessä funktiossa s , tulee taajuudella f huono korrelaatio ja sen taajuuden kerroin on lähellä nollaa.

-artefaktit ovat mitattuja signaaleja, jotka eivät ole peräisin mitattavasta fysiologisista prosesseista. ne syntyvät siis muissa laitteiston osissa tai ihmisen virheiden seurauksena.

-artefaktit voidaan jakaa kahteen luokkaan: fysikaaliset artefaktit (syntyvät esimerkiksi käytetystä laitteistoista, huonoista mittauslaitteiden kontakteista, muuttuneesta mittausasetelmasta)

tai biologiset artefaktit (lihaskiväily, hampaiden purenta, liikkeet, silmänliikkeet).

-yleinen sääntö mittausvirheiden minimoimiselle pätee aina: on helpompaa estää artefaktien muodostuminen kokeen aikana harkitulla koeasetelmalla ja hyvällä laitteistolla kuin poistaa ne signaalinkäsittelyvaiheessa.

-toisinaan biologisista artefakteista voi olla myös hyötyä, fyysikaalisista sen sijaan harvemmin. esimerkiksi jatkuva lihastonius voi kertoa stressistä ja yleisestä huonovointisuudesta ja aivosähkökäyrän artefaktit hermostosairauksista. siksi toisen mittauksen artefaktit voivat olla toisen mittauksen haluttu signaali. silmänliike-elektrodeja käytetään merkitsemään data, joka ei luultavasti ole käyttökelpoista signaalia käsitellessä (toisaalta silmäliikedatalla voi tehdä kaikkea muuta, sillä silmä on kuin dipoli, mistä on helppo karakterisoida mihin suuntaan dipoli osoittaa).

-artefaktien seulonta (detection) perustuu ainakin seuraaviin: HUIPPUARVOJEN seuraaminen (usein poistetaan datasta kohdat jotka ovat amplitudiltaan liian suuria oletettavissa olevaan mittausdataan nähden, ongelma on rajojen asettamisessa eivätkä kaikki artefaktit luopu piikkejä mittausdataan, esimerkiksi pitkäperiodiset signaalit), KULMAKERTOIMEN mittaaminen (liian nopeat ajalliset muutokset ovat usein virheellistä dataa), ALHAISET AMPLITU-DIT (signaalin lasku alhaisille amplitudeille ei liene normaalia monissa sovellutuksissa. myös keskimääräisen amplitudin selvä nousu voi olla artefakti).

-artefaktien poistamisrajojen määrittäminen on haastavaa. fysiologisissa järjestelmissä voidaan antaa mahdottomat rajat (esim. sykke ei voi olla 300 bpm), mutta tämä on usein liiankin korkea raja signaalinkäsittelyn ja todellisten rajatapausartefaktien poiston kannalta. jakauma signaalin normaaleista arvoista tulee tietää.

-artefaktien poistamisessa tulee myös huomioida, että on (etenkin lääketieteellisissä sovellutuksissa) vaarallisempaa jättää jokin todellinen ilmiö huomioimatta kuin poistaa se.

-Gaussisesti jakautunut prosessi on helppo, sillä tällöin voidaan käyttää esimerkiksi six-sigma-periaatetta, missä 99,7 % mittauspisteistä osuu keskiarvostaan kolmen hajonnan päähän. artefakti voi tuki olla hajonnan sisällä, jolloin se huomioidaan mittauspisteinä, mutta huomattavan paljon jää myös ulkopuolelle. jakauman normaalius on aina tarkistettava tilannekohtaisesti.

-koejärjestelyjen muuttuessa muuttuvat usein myös artefaktien hylkäysrajat. tällöin ne tulee mitata ja suunnitella uudestaan. myös artefaktin paikka taajuustasossa tai esimerkiksi mittausdataan eksyneen virheellisen jännitepiikin muoto voi helposti muuttua toisenlaiseksi pienellä mittausjärjestelyn muutoksella.

-mittausdatan yksittäiset piikit voidaan poistaa lineaarisella interpoloinnilla off-line-analyysissä esimerkiksi mediaanisuodattimella. näin saadaan pois etenkin yksittäiset mittausvirhepisteet, mutta ei voida poistaa askelmaista muutosta signaalin trendissä.

LUENTO 4: Biosignaalien analysoinnista signaalin merkitykseen: ominaisuuksien etsiminen, luokittelu ja päätöksenteko

-usein mitataan useita eri muuttujia, joista halutaan rakentaa soveltuva järjestelmä halutun mittausarvon ulossaamiseksi. mitattavien parametrien valinta ei välttämättä ole selkeä, ja osasta voidaan saada vain sen verran vähän tietoa, että niitä ei kannata monimutkaistumisen vuoksi lopullisessa mallissa käyttääkään. tämän vuoksi tarvitaan ominaisuuksien valintaperusteita ja karsimista.

-ominaisuuksilla tarkoitetaan eri signaalinkäsittelyvaiheissa eri asioita. alkuvaiheessa nämä ovat esimerkiksi amplitudi ja vaihe, josta johdetaan taajuustason esitys ja regressorit. näillä taas on

ominaisuuksinaan esimerkiksi kulmakerroin, “piikikkyys” tai jokin matemaattinen malli.

- ominaisuuksien karsiminen (feature extraction) vähentää syntyvän datan määrää, kun kaikkea dataa ei halutakaan tallentaa tai ottaa analyysiprosessiin mukaan.
- ominaisuuksien valitsemisella (feature selection) tarkoitetaan, että vain kaikista tärkeimmät ominaisuudet valitaan mitattavaksi/laskettavaksi vaikka niistä voisi laskea paljon enemmänkin tunnuslukuja (on turha laskea vaihevastetta, jos vain amplitudivaste tarvitaan).
- ominaisuusvektorit kuvaavat kunkin ominaisuuden merkityksellisyyttä loppupäätelmään nähden. nämä voidaan johtaa eri mittausmekanismille tai siitä laskettaville parametreille. vektorit voidaan sijoittaa omin luokkiinsa, jotka kuvaavat samankaltaisia ominaisuuksia. saman luokan vektorit korreloivat usein voimakkaasti.

-eri ominaisuuksien havainnointiin on erilaisia tekniikoita (pattern recognition techniques), jotka voidaan jakaa kolmeen pääläjiin: STATISTINEN (luokittelee dataa statistisin keinoin), SYNTAKTINEN (karakterisoi signaaleita niiden muodon ja ominaisuuksien perusteella eikä niinkään numeeristen arvojen) ja KEINOÄLYLLINEN tekniikka (AI, artificial intelligence, pyrkii luokittelemaan dataa loogisten päättelyiden kautta (=symbolic AI, symbolinen AI) tai karakterisoi dataa opittujen esimerkkien kautta (=connectionist AI, yhdistelmäkeinoäly)).

-statistisessa luokittelussa ominaisuusvektorit sijaitsevat N-dimensioisessa omassa avaruudessaan, missä N on vektorien lukumäärä, jolloin ne voidaan ilmaista ortogonaalina systeeminä. ne määrittävät päätösalueita (decision region), joilla mallinnetaan kutakin luokkaa. ongelma on saada luotua rajat näille alueille päätösfunktion (decision function) avulla. alueita luodessa voidaan käyttää esimerkiksi miniminormiestimaatteja, toisinaan taas datan klusteroitumismalleja.

-syntaktisessa ominaisuuksien havainnollistamisessa samantyyppisiä signaalimuotoja (ei niinkään niiden laskennallisia arvoja) laitetaan samoihin ryhmiin. ryhmiinjakaminen tapahtuu ominaisuusvertailuilla, esimerkiksi vertaamalla kulmien ja käyrien samankaltaisuutta tunnettuihin lajittelumuotoihin. näin voidaan esimerkiksi karakterisoida ja tunnistaa sydänsähkökäyrästä P-aalto ja QRS-kompleksi. tämä tekniikka on herkkä kohinalle.

-keinoälytekniikka keskittyy kahteen pääkohteeseen: tiedon esittämiseen järkevästi sekä uuden oppimiseen ja varmuuden parantumiseen toistojen myötä.

-symbolinen AI (myös expert system) on osa keinoälytekniikkaa ja pyrkii esittämään oikean datan käyttäjäkomentojen avulla loogisten päättelyketjujen kautta (hieman kuten Wolfram Alpha tietää mitä sinne syötetään vaikkei kyseessä olisikaan Mathematica-syntaksi). systeemi rakentuu tietopankista (lainalaisuudet, mallit), datapankista (mitattu data), heuristiikkakoodista (tulkitsee mitä käyttäjä haluaa) ja käyttäjärajapinnasta (ulkoasu, ohjeet). toiminta ongelmanratkaisussa voi olla suoraviivainen (ongelman suoriin lainalaisuuksiin haetaan ratkaisut, kunnes päädytään lopputulokseen) tai epäsuora (aloitetaan ongelman tutkiminen seurauslähtöisesti ja palataan rekursiivisesti takaisinpäin kohti syitä). loogiset päättelyketjut on aina mahdollista tarkastaa käsin (ohjelman tulisi antaa tästä palaute). tiedon varastointi, muuttaminen ja päättelyketjun muutokset ovat haastavia toteuttaa. myös niiden formuloiminen tietokoneiden ymmärtämään muotoon on haastavaa.

-yhdistelmäkeinoäly (connectionist AI) on yleinen ongelmanratkaisumalli neuroverkoille (artificial neuronal networks, ANN). niiden avulla voidaan etsiä ratkaisuja ongelmiin, missä vain vähän esitietoa ongelmasta on saatavilla, mutta mistä on olemassa samantyyppisiä esimerkitapauksia (esimerkitapauksista siis luodaan ongelmanratkaisua vaikka ne eivät olisikaan juuri käsitellyn ongelman osana).

-ANN tarkoittaa suurta määrää yksinkertaisia epälineaarisia järjestelmiä, jotka on yhdistetty toisiinsa (kuten ihmisen neuroniverkko). näiden siirtämässä informaatioissa tietoa sisällytetään kytkösten välisiin painokertoimiin. äly neuroverkoissa liittyy näiden painokertoimien muuttamiseen kulloisenkin ongelman mukaan (esimerkiksi henkilön ikä voi vaikuttaa tiettyyn ongelman

ratkaisun suhteen vahvasti korreloivasti, jolloin se saa suuren painoarvon). tarkoitus on luoda ulostulossa järjestetty lista eri ratkaisuista ongelmaan todennäköisyyksien mukaan.

-ANN-tyyppisiä neuroverkkoja tulee opettaa esimerkkidatalla, kun tiedetään millainen ulostulo niistä tulisi kullakin sisäänmenolla saada ulos. näin ollen painokertoimet määritellään halutuiksi useilla iteraatioilla. uuden oppiminen myöhemmin voi tapahtua automaattisesti.

-toisenlaisiakin ANN-verkkoja on käytössä, joiden äly perustuu esimerkiksi stokastisiin prosesseihin tai sumeaan logiikkaan.

-ANN-verkkojen kouluttamisessa datan puuttuminen on usein ongelma. esimerkiksi terveitä referenssitietoja voi olla saatavilla paljonkin, mutta kouluttamisessa on hyötyä jos olisi myös sairastapauksia opettamassa verkkoa.

-expert systemin ja ANN-tyyppisten neuroverkkojen pääero on, että expert system pyrkii pääsemään mahdollisimman vähällä laskentatyöllä loogisten päättelyketjujen perusteella (mikäli potilaalla on tämä tauti ennestään, sillä varmaankin on toinen tähän liittyvä sairaus mikäli tuntomerkit täsmäävät) ja siten rajata mahdollisuuksia. ANN-verkkojen tehtävänä taas on laskea kaikki kytkökset ja käyttää esitietojen (opetusdata perusteella laskettuja) painokertoimia todistamaan eri mahdollisuuksien todennäköisyydet hyvin pieniksi, mutta ei karsimaan näitä kuitenkaan kokonaan pois laskennasta.

LUENTO 5: Terveysteknologisia sovelluksia

-terveyden ylläpitämisessä on monia tasoja, jotka vaihtelevat liikkumisen hyvinvoinnista, kotihoitosta, sairaushoidosta aina akuuttiin sairaalahoitoon.

-akuutissa sairaalahoidossa (tai sairaanhoidossa) on tärkeää saada ihmiselimestöstä jatkuvaa tietoa sen tilasta. näin esimerkiksi elimistöä tulee monitoroida ja vaarallisten muutosten tulee antaa hälytys. saatavilla tulee olla oikea henkilökunta (tai laskentateho) informaation päättämiseen ja tiedonluontiin.

-terveysteknologiassa ollaan tekemisissä äärimmäisen lopullisten asioiden kanssa, joten niiden tulee olla erittäin vikavarmoja, helppokäyttöisiä kaikelle henkilökunnalle, nopeita (miehellään lähes reaaliaikaisia) ja turvallisia potilaalle.

-terveysteknologisten laitteiden haasteita ovat esimerkiksi henkilöiden yksilölliset erot mittauksissa, ei-normaalien mittausympäristön haasteet (potilas mitatetaan sairaalan tiloissa, joka varmasti vaikuttaa mittauksiin kotimittauksiin nähden), tiedonpuute, subjektiiviset (tai yrityksen sisäiset) näkemykset miten laitteisto pitää rakentaa ja pitkät validointiajat uuden laitteen käyttöönotossa.

-mikään järjestelmä ei korvaa hoitajien ja lääkäreiden työtä (ja läsnäoloa), mutta voi helpottaa niitä tuottamalla nopeasti informaatiota heidän käyttöönsä.

-potilasmonitoroinnilla on isot markkinat. laitteistojen tulee näyttää informaatio, tuottaa siitä tunnuslukuja, etsiä myös pitkäaikaisia trendejä potilaan tilan muutoksessa, tallentaa data, olla käytettävissä etänä, raportoida tilanne myös tekstimuodossa, verrata dataa statistisesti ja tuottaa uutta informaatiota tietokantoihin.

-yleensä biosignaalien prosessoinnissa halutaan tehdä jotain (tai kaikkea) seuraavista: PALOITTELU (segmentointi, tiedon jakaminen osaväleihin), TAPAHTUMIEN HAVAINNOINTI (event detection, esim. sydämenlyönti), SYNKRONOITU KESKIARVOISTAMINEN ja MUUTOSTEN HAVAITSEMINEN (pitkälläkin aikavälillä).

-segmentoinnissa eli paloittelussa tai ikkunoinnissa ikkunanpituus pitää valita niin, että mahdollisimman paljon informaatiota välittyy (liian lyhyt ikkuna ei anna juuri tietoa), mutta ikkuna ei ole niin pitkä ettei muutoksia huomata. ikkunointi voidaan tehdä samanpituuisina pätkinä tai

tilanteeseen mukautuvina. myös vertailuikkuna (reference window) voi olla mittauksen edetessä yhtä pitkä tai kasvaa aina mittauksen alusta asti (laskennallisesti kasvava rasiutus).

-myös potilaan tai koehenkilön tila voidaan segmentoida eri luokkiin. tähän voidaan näyttää apuna epälineaarista energiaoperaattoria (NLEO, non-linear energy operator), joka huomioi niin amplitudi- kuin taajuusmuutoksetkin ja näin saadaan kuva siitä milloin ihmisen tila on muuttunut (NLEO-kuvajassa on piikkejä eli signaalin energiajakauma on muuttunut). segmentointia voidaan havainnollistaa esimerkiksi aikajanalla, missä on väritetty eri alueet kertoen eri tiloista tai antaa tila uutena suureena toisella akselilla (esimerkiksi unen syvyys).

-tapahtumien havaitsemisessa halutaan nähdä/havaita/visualisoida missä tapahtuu jokin fysiologinen tapahtuma. nämä voivat olla hankalastikin kohinaiseen dataan piiloutuneita (aivosähköiset ilmiöt), helposti havaittavia (QRS) tai vasta pitkän ajan trendeissä nähtävissä olevia. usein tapahtumilla (eventeillä) on tietty tunnettu muoto, jota voidaan käyttää vertailussa ja siten havaitsemisessa.

-Pan-Tompkins-algoritmeilla voidaan havaita sydämensykkeestä QRS-kompleksi, sillä sen ominaisuudet muuhun signaaliin nähden tunnetaan ja laskennallisina välivaiheina se voidaan tuoda esiin. algoritmissa raaka EKG-data kaistanpäästetään QRS-taajuusalueella, tämä derivoidaan ja neliöidään ja lopuksi luetaan integraattori-ikkunalla. lopputuloksesta näkee helposti sydämen QRS-kompleksien paikat. lopputulos tosin on ikkunaintegroinnin myötä hieman viivästynyt.

-matched-filteeröinnissä pyritään maksimoimaan signaali-kohinasuhde (SNR, signal to noise ratio) ja siten nähdään kohinan seasta onko signaalissa haluttuja tapahtumia. ei ole niinkään oleellista minkälainen signaalin muoto on käsittelyn jälkeen, kunhan sen olemassaolo voidaan kohinan seasta todentaa. matched-filter perustuu konvoluutioon ja filteerin impulssivasteen muoto on peilikuva halutulle signaalille, jolloin konvoluutiotuloksen tulisi korreloida mahdollisimman hyvin todellista eventtiä.

-synkronoidussa keskiarvoistamisessa oletetaan, että kohinaa ei voida poistaa signaalista, koska se on samalla taajuusalueella itse vasteenkin kanssa. tämän vuoksi tulee tehdä keskiarvoistaminen, missä oletetaan vasteen toistuvan samanlaisena joka kerta. kun keskiarvoistaminen tehdään synkronoidusti aina samasta kohdasta stimuluksen alkaessa. valkoisen kohinan ominaisuuksien mukaisesti sen painoarvo pienenee keskiarvoissa ja itse vaste saadaan näkyviin kunhan signaali ja kohina eivät korreloi. yleinen aivosähköisten ilmiöiden mittaamisessa (heikkoja ja kohinaisia, signaalin osuus voi olla vain noin 1-5 μV , mutta kohinan jopa 10-100 μV). signaali-kohinasuhde paranee neliöjuurellisesti toistettujen kokeiden määrään (yhdeksänkertainen stimulusten määrä kasvattaa SNR:n kolminkertaiseksi).

-muutosten havaitsemisessa käytetään kahdenlaista havaitsemistapaa, jotta nähdään milloin systeemi ajautuu pois normaalitilasta (OOC, out of control -situation). merkitykselliset nopeat muutokset ja pitkän aikavälin trendit ovat molemmat merkkejä epänormaalista elimistön käyttäytymisestä. yleensä normaalialueella signaalit ovat keskihajonnan (tai six-sigman mukaisesti kuu-den hajonnan) sisällä ja muutoksia tutkitaan useimmiten statistisin keinoin.

-useimmiten OOC-tila nähdään muuttuneesta keskiarvosta, yksittäisestä mittauspisteestä, hitaasta muutoksesta tai signaalin huomattavasta varianssin muutoksesta. tilan seuraamista tulee tietenkin dataa ikkunoida ja valita keskimääräinen ikkunapituus kunkin tilan mukaan (ARL, average run length). jos potilaan tila on vakaa, voi ARL olla pitkäkin ja näin välttyään vääriä hälytyksiltä. jos taas tila ei ole vakaa, pitäisi ikkunan olla lyhyt nopeaa muutosten havaitsemista varten.

-muutoksia voidaan visualisoida Shewhartin kartoilla, missä on merkitty mittauksen oletettava keskiarvo ja kolme hajontaa kumpaankin suuntaan pystyakselilla. vaaka-akselille merkitään havaintojen keskiarvotulos, jolloin voidaan nähdä em. muutoksia silmämääräisesti.

-keskiarvon hitaita muutoksia taas voidaan havainnollistaa Cusumin kartoilla. siinä mitataan ensin tietty määrä mittauspisteitä referenssiksi, jonka jälkeen näistä laskettuja keskiarvoja ja

hajontoja aletaan käyttää seuraavien arvojen virtailukohtana. näin ollen ensimmäisen vaiheen kerätty data auttaa näkemään hitaat muutokset myöhemmässä vaiheessa. se ei kuitenkaan sovellu esimerkiksi yksittäisten mittausarvojen vaihteluiden seuraamiseen hyvin (siihen sopii Shewhart paremmin).

LUENTO 6: Hyvinvointiteknologian sovelluksia

-kehittyneissä maissa väestö ikääntyy nopeasti. tämä ja elämän teknillistymisen vuoksi erilaisille hyvinvointia mittaaville laitteille on lisäkysyntää.

-hyvinvointia mittaavien laitteistojen merkitys voi olla valtava kustannussäästö esimerkiksi vanhustenhoidossa, missä pyritään että mahdollisimman moni vanhus pystyisi toimimaan itsenäisesti. laitteilla on myös tarkoitus ennaltaehkäistä tulevia kustannuksia.

-laitteistoja jotka monitoroivat ihmisten tilaa ja antavat tietoa terveydentilasta ovat esimerkiksi verensokerimittarit, verenpainemittarit, sykemittarit ja lattialikeanturit (vanhustenseuranta).

-laitteistot voidaan jakaa pääasiassa kolmeen luokkaan. ubiikit anturit, joita ei tarvitse itse erikseen käyttää vaan ne mittaavat ihmisen läsnäollessa jatkuvatoimisesti (ambient sensors) esimerkiksi onko ihminen huoneessa tai nukkumassa pedissään. puettavat sensorit (wearable sensors) mittaavat esimerkiksi rannekellomaisesti ja laitetta tulee pitää mukana, ja varsinaisia mittalaitteita (measurement devices) käytetään vain mittaustapahtumassa.

-laitteistojen suunnittelussa tulee ottaa huomioon seuraavia seikkoja: käytettävyys, käyttäjystävällisyys, hinta, luotettavuus, kestävyys, toimintavalmiuden nopeus, huoltokustannukset, määräaikaishuollot, muotoilu, käyttäjätunnistus, tiedon tallennus, yhteydet mahdolliseen hoitohenkilökuntaan ja tietoliikenneyhteydet.

-laitteistojen tulee omalla algoritmillaan ja laskentatehollaan pystyä tekemään edellisessä luennoissa esitettyjä signaalikäsittelyn tehtäviä loppupäätelmien esittämiseksi.

-ongelmia datankäsittelyssä ovat esimerkiksi puuttuva data (käyttäjä ei ole käyttänyt laitetta sovitusti), alinäytteistäminen (liian alhainen näytteistystaajuus), muuttuva mittausympäristö ja datan laatu sekä kohinaisuus.

-yllättävänkin kohinaiselta näyttävät signaalit voivat olla oikeasti deterministisiä. esimerkiksi peräkkäisten arvojen suhteellinen erotus voi olla hyvin samankaltainen. tätä mitataan ns. Poincare-kuvaajilla, missä signaalia itseään viivästetään (eri viivästyksistä saadaan lisätietoa) ja nämä piirretään kahteen ulottuvuuteen. satunnainen data hajoaa, mutta deterministinen muodostaa suoralle $y = x$ tiheän pisteistön. tälle voidaan tunnuslukuna laskea keskihajonnat eri suuntiin (kertoen yleisestä mittausarvojen hajonnasta sekä aikaviivästettyjen arvojen hajonnasta).

-kun dataa ei ole saatavilla tasavälisellä näytteistyksellä (käyttäjä ei ole esimerkiksi mitannut painoaan joka päivä, vaikka pitäisi), tulee dataa manipuloida tasaväliseksi näytteistykseksi (luo, eli interpoloi ja keskiarvoistaa puuttuvat mittauspisteet. voidaan myös luoda muiden vastaavien ajanjaksojen perusteella eikä pelkästään kahden lähimmän mittauksen mukaan puuttuvan ympärillä) tai käyttää erikoisalgoritmeja, jotka huomioivat epäsäännöllisen näytteistuksen. dataa voidaan myös suoranaisesti hylätä, jolloin tältä ajalta ei saada myöskään tuloksia.

-mahdollisimman huomaamattomasti dataa voi luoda Lomb-Scargle-metodilla, missä signaaliin sovitetaan mahdollisimman pienellä keskineliövirheellä sinikäyriä. tällöin vältetään tehospektarin muuttuminen uusien mittauspisteiden luomisen yhteydessä.

LUENTO 7: Tiedonkeruukanavien yhdisteleminen

-tämänhetkinen terveysteknologia suuntautuu sairauzehoidosta kohti ennaltaehkäisevää tekniikkaa. sen vuoksi useita tietoja pitää voida yhdistää ja tehdä näistä oikeat johtopäätökset.

-datankeruu ei taaskaan ole ongelma, mutta sen muuttaminen hyödylliseksi infomaatioksi on. miten esimerkiksi yhdistetään EEG- ja verenpaine-tieto (korkea verenpaine lisää kohinaisuutta ja tätä tietoa voidaan hyödyntää)?

-yleensä mittausrvojen yhdistäminen kannattaa tehdä tarkastelemalla (ensin visuaalisesti, esimerkiksi hajontakuviomatriiseilla (scatter plot)) eri mittauskomponenteista saatua dataa ja tarkastella mitkä voisivat olla sopivia yhdistettäviä parametreit näistä. kvantitatiivinen tarkastelu on nopeaa ja mahdollisesti hyvinkin palkitsevaa. tämän jälkeen kannattaa siirtyä laskennalliseen analyysiin.

-yhdistämisessä voi käyttää hyväksi korrelaatioita. tässä vaaditaan samanaikainen näytteistäminen. tarkastellaan miten paljon kukin ominaisuuspari/ryhmä korreloi keskenään tietyn ongelman ympärillä. hyvät korrelaatiotuloksetkin kannattaa kuitenkin silmämääräisesti pohtia tarkoin mielekkyyden kannalta (hukkumiset ja kuolemat korreloivat vahvasti, mutta tämä tuskin tuo lisäinformaatiota). kohinaisen datan joukosta korrelaatiot eivät välttämättä ole lukumääräisesti suuria, mutta ovat kuitenkin statistisessa mielessä merkittäviä.

-datat tulee myös normalisoida. normalisoinnit voi tehdä jokaisen koehenkilön omaa dataa vasten tai koko mittausryhmään nähden. valinta riippuu halutuista suureista ja algoritmit tulee kehittää niiden mukaan.

-merkittävien komponenttien analyysissa (PCA, principal component analysis) tarkoituksen on luoda lineaarikombinaatio ja tätä kautta ortogonaali kanta kaikista mittausarvoista. näiden vektoreiden pituudella on tietty merkitys mittauksen kokonaisvarianssiin. edelleen voidaan laskea miten paljon kukin komponentti korreloi kantamatriisissa muiden kanssa ja näistä voidaan luokitella ryhmiä (esimerkiksi pituus ja paino korreloivat luultavasti vahvasti, yhteinen tekijä on koko) sekä nähdä merkityksellisimmät komponentit.

-PCA-analyysissä Scree-visualisoinnissa kantamatriisin ominaisarvot sijoitetaan merkitsevyyssjärjestyksessä kuvaajaan, jolloin voidaan helposti nähdä kaikista merkityksellisimmät mittauksessa tarvittavat komponentit. on turhaa mitata 20 eri parametria, jos jo kolmen mittaamisella päästään 95 % siitä tulosvarmuudesta, johon kaikkien mittaamisella päästään.

-PCA-analyysi ei huomioi mahdollisuutta, että vain osa mittaustuloksista on käytössä. jos valitaan kolme kahdestakymmenestä, voi mittaustulos olla uudella järjestelyllä kuitenkin erilainen. siksi merkityksellisyyttä pitäisi mitata myös komponenttivalinnoilla, eli pyritään valitsemaan paras mahdollinen alijoukko mitattavia komponentteja kun muita komponentteja ei huomioida lainkaan. vie kuitenkin paljon laskentatehoa tehdä PCA-analyysi kaikille osajoukoille.

-osajoukkojen haravoimiseen on useita eri mahdollisuuksia. osassa esimerkiksi käydään läpi kaikki mahdollisuudet, osassa vain merkittävimmät. toisissa taas lisätään ja/tai poistetaan komponentteja yksi kerrallaan.

-tiedonkeruun esimerkkitapauksena on anestesia, jolla on kolme päätavoitetta: potilas ei saa olla tietoinen leikkaustapahtumasta, hän ei saa tuntea kipua ja lihasten tulee olla rentoja leikkauksen mahdollistamiseksi. liian heikko anestesia aiheuttaa potilaan valvotuneisuutta ja liian vahvana johtaa esimerkiksi verenkierron ja hengityksen ongelmiin.

-anestesia-aineet sisältävät monia aineita, joissa kunkin tehtävä on tehdä jotain yllämainituista. ne kuitenkin vaikuttavat suoraan tai elimistön oman toiminnan kautta lähes kaikkiin muihinkin. ongelmia tulee, mikäli esimerkiksi potilaan kivunsietoa pitää lisätä, mutta tämän mahdollistava aine rentouttaa sivuvaikutuksen muita lihaksia liikaa.

-tietoisuudelle on määritelty kuusiportainen ns. OAA/S-skaala (observer's assessment of alert-

ness/sedation, ulkoisten merkkien, esimerkiksi nimeen vastaamiskyvyn suhteen). tämän lisäksi EEG:llä on mahdollista mitata anestesian tasoa (koska aivosähkökäyrästä näkee melko helposti ihmisen tietoisuuden tilan) sataportaisella BIS-indeksillä (kuvaa unisuutta/nukutuksen syvyyttä).

-BIS-indeksi on epälineaarinen yhdistelmä kolmesta komponentista, joita painotetaan eri kertoimilla (EEG:n aikakäyttäytyminen, aivosähkökäytän taajuusalueessa logaritminen tehospektirisuhde tietyillä beta-aivoaaltoaajuuksilla sekä toinen EEG:n taajuusaluepiikkien suhdeluku). -toinen lähestymistapa voisi olla EEG:n epäjärjestyksen mittaaminen. mitä epäjärjestyneempi signaali on, sitä enemmän on aivotoimintaa ja sitä valvettuneempi ihminen on. spektraalientropia voidaan ilmoittaa yhdellä luvulla nollan ja yhden väliltä.

-kivun mittaaminen (analgesia) on haastava kvantisoida, koska se on subjektiivinen näkemys. lisäksi erilaiset lääkeaineet kehossa muuttavat kiputuntemusta. autonomisen hermoston kautta suureisiin päästään kuitenkin käsiksi, sillä autonomisen hermoston sympaattinen puoli aktivoituu kivussa. tällöin esimerkiksi verenpaine nousee, sydämensyke kiihtyy ja kasvojen lihakset supistuvat. näitä voidaan monitoroida, joten kiputilasta saadaan epäsuorasti tietoa.

LUENTO 8: Ideasta lopputulokseen

-tyypillisellä biosignaalien analysointiprojektilla on seuraavat vaiheet: ALKUMÄÄRITTELY (määritellään mitä halutaan tutkia ja luodaan hypoteesi sen todentamista varten), SUUNNITTELU (tehdään koeasetelma mahdollisimman tarkasti), KOKEIDEN TEKEMINEN (tehdään tarvittavat mittaukset ja kerätään data), PROSESSOINTI (luodaan mitattua dataa varten oikeat algoritmit ja analysoidaan tulos) sekä JULKISTAMINEN (tehdään loppupäätelmät ja julkaistaan tulos, mahdollisesti lehtiartikkelina).

-alkumäärityksessä voidaan ottaa tutkimukselle jokin seuraavista suunnista: uuden datan LUOMINEN (muodostetaan dataa, jota ei ole ennen ollut ja tutkitaan jotain ilmiötä ilman, että varsinaisesti tiedetään mitä etsitään. voidaan hakea esimerkiksi dataa, jota tarvitaan jossain myöhemmässä tutkimuksessa. haluttu data kannattaa säilöä hyvin dokumentoituna sekä myös raakadatamuodossa.), PROOF OF CONCEPT (todistetaan, että jokin menetelmä voidaan toteuttaa uudella tavalla ja että tämä on jatkojalostettavissa), VERTAILU VANHAAN (kehitetään uusi mittausjärjestely ja verrataan sitä nykyiseen. pyritään tekemään helpompi, nopeampi tai parempi. kokonaisvaltainen aiempaan teknologiaan vertaaminen voi olla haastavaakin, mikäli esimerkiksi verrataan käyttökokemusta), uuden METODIN KÄYTTÖÖNOTTO (luodaan täysin uusi tapa ilmaista jokin ilmiö esimerkiksi uuden suureen tai indeksin avulla. haasteena on saada asiakkaat tietoisiksi siitä, että he tarvitsevat uutta tekniikkaa) tai AUTOMATISOINTI (luodaan tietokoneavusteinen metodi tehdä asia, joka aikaisemmin on tehty käsityönä. voiko algoritmi kuitenkin lopulta korvata ihmistä ja milloin järjestelmä on tarpeeksi hyvä ja valmis? voi kuitenkin onnistuessaan säästää paljon kustannuksia ja työvoimaa).

-suunnittelussa mietitään mikä on se kysymys, johon tutkimuksessa halutaan vastaus. lisäksi päätetään millaisia menetelmiä käytetään, paljonko koehenkilöitä tarvitaan ja pohditaan esimerkiksi kokeen eettisyyttä.

-kokeiden tekemisen vaiheessa suoritetaan aluksi pilottikokeita, joilla testataan kokeen toimivuutta. tässä vaiheessa myös hiotaan datankeruutavat, pyritään varautumaan mahdollisimman moneen erilaiseen ongelmaan ja lopulta suoritetaan mittaukset halutulle määrälle koehenkilöitä. mitä enemmän dataa saadaan kerättyä, sen varmempia tulokset tulevat olemaan. lisäksi paljon dataa joudutaan hylkäämään eri syistä (koehenkilö ei noudattanut ohjeita, laitteisto ei mitannut kaikkea, mittauksen tekijöiden virheet, käyttäjän virheet).

-prosessointivaihe on usein pisin osa prosessiketjua. siihen kuuluvat esimerkiksi datan hyväksyminen käsittelyyn, silmämääräinen tarkastelu, statistinen analyysi, sopivan datan valitseminen kaikista tarjolla olevasta, prosessointi kehitetyillä algoritmeilla ja alkuperäisen hypoteesin testaaminen. prosessin loppuvaiheessa on yleensä tarkoitus saada yksi ulostulo, vaikka prosessi saa tietoa kymmenistäkin lähteistä. tämän kurssin muut luennot ovat pääasiassa painottuneet tähän vaiheeseen.

-prosessointivaiheessa tulee ottaa kantaa virhetodennäköisyyteen. virheitä on paljon erilaisia, joita voidaan tutkia ns. sekoitematriisilla (confusion matrix). se ottaa huomioon virheiden määrän eri parametreilla ja/tai koehenkilöiden ominaisuuksilla. helppo ja vaarallinen esimerkki tästä on ihmisen luokittelu sairaaksi: järjestelmä voi luokitella sairaan sairaaksi ja terveen terveeksi, mutta myös tervees sairaaksi (uudella mittauksella varmistutaan onko tämä totta) tai sairaan terveeksi (vältettävä kaikin keinoin!).

-sekoitematriisissa on kaksi todennäköisyyttä, jotka kuvaavat herkkyyttä (sensitivity, kyky luokitella eri luokkiin) ja luokittelun tarkkuutta (specificity, kyky luokitella oikein). virheitä voi myös painottaa (esimerkiksi vähentääkseen terveiksi luokiteltujen sairaiden osuutta) sen mukaan paljonko kustannuksia kukin virhe tuottaa (yleisin tapa on mitata rahallisia kustannuksia, mutta muitakin kriteerejä voi käyttää). kustannukset halutaan tietenkin minimoida, joten näin saadaan optimointiongelma, mistä voidaan laskea painokertoimet eri virheille. kustannusten euromääräinen arvioiminen eri virhelähteistä on kuitenkin haastavaa.

-myös herkkyyden ja tarkkuusongelma on optimointia. kumpaakin ei voi maksimoida samanaikaisesti, vaan tulee valita kompromissi. tähän auttaa ROC-kuvaaja (receiver operating characteristic -plot), vaaka-akselilla on 1-tarkkuuden (1-specificity, huomaa vähennyslasku, eli tarkkuus on korkeimmillaan kun tällä akselilla ollaan nollassa) todennäköisyys ja pystyakselilla herkkyyden todennäköisyys. suora $y = x$ on satunnaisen valinnan todennäköisyys 0,5 ja kuvaajan vasen ylänurkka kuvaa täydelliseen luokitteluun kykenevän systeemin pistettä (herkkyyden (ei vääriä positiivisia) ja tarkkuuden (ei vääriä negatiivisia) todennäköisyydet molemmat yksi). todellinen kuvaaja kulkee useimmiten jossain satunnaisen valinnan lävistäjän vasemmalla puolella.

-kahden komponentin systeemiä voi ROC-kuvaajan lisäksi visualisoida myös todennäköisyysmassan jakautumisella osittain päällekkäin olevilla normaalijakaumilla (jakaumat, jotka eivät tosin välttämättä ole normaalijakaumia, edustavat todellisuudessa terveitä tai sairaita ihmisiä). siirtämällä tässä kuvaajassa toimintapistettä (vaaka-akselilla on kullekin koehenkilölle tutkitavan asian tai algoritmin antama numeerinen arvo, jonka perusteella jaottelu tehdään. jaottelun raja-arvoja voidaan siis siirtää), saadaan pinta-alojen suhteella tutkittua luokittelun todennäköisyyksiä (sairas, ei sairaas, epäsairas, epäterve).

-tarkoituksena on prosessoida algoritmi/koe myös mahdollisimman suorituskykyiseksi (käyttökokemus, nopeus jne.). koska on kallista tehdä aina muutos ja testata järjestelmää, tyydytään usein käyttämään datankäsittelykeinoja. mitatulla datalla on siten kaksi käyttötarkoitusta: kehittää järjestelmää "kouluttamalla" sitä tai saada tuloksia testaamalla sitä. usein on ongelmana mikä osuus uhrataan kumpaankin. kouluttaminen lisää lopputuloksen varmuutta, mutta tällöin tuloksien testaamiseen on vähemmän dataa käytettävissä. ristivalidoitimenetelmässä testidata erotetaan ristivalidoitavasta datasta, jota käytetään kiertävästi niin testaamiseen kuin järjestelmän kouluttamiseenkin (esimerkiksi 10 mittauksesta voidaan kulloinkin käyttää 90 kouluttamiseen ja 10 testaamiseen, jolloin tämä voidaan tehdä 10 kertaa ns. 10-fold cross-validationissa).

-lopullisen suorituskyvyn näkee aina vasta viimeisimmällä, reaalisella (tai ristivalidoinnissa koskematta jääneellä) datalla. suorituskykyä parantaakseen ei myöskään tule oikoa koejärjestelyissä tai tuoda omia mielipiteitä ja kulttuuriarvoja mukaan, sillä tämä vähentää kokeen uskottavuutta ja tieteellisen tiedon luomisen peruseriaatteita (toistettavuus, objektiivisuus).

-lopuksi vielä jotain ohjeita tutkimuksen tekemiselle: huolellinen suunnittelu (kohteen valinta ja koko tutkimusprosessin mahdollisimman selkeä toteuttaminen, myöhemmässä vaiheessa on vaikeaa lisätä koehenkilöille jokin tarvittava tieto, mitä ei ole kokeessa mitattu), kommunikointi (ei ole tyhmää kysyä mikäli ei ymmärrä. biologisia järjestelmiä käsiteltäessä useat eri alojen ammattilaiset eivät välttämättä tulkitse asioita heti samalla tavalla tai puhu samaa työkieltä), datan tarkastaminen ja dokumentointi (kun asiat ovat tuoreena mielessä, on helppo kirjoittaa ylös mitä tapahtui ja nopea silmäys dataan voi paljastaa jotain, mikä liittyy koeasetelmaan. esimerkiksi arvailuja EEG-käyrän suttuisuudesta on paljon vähemmän, mikäli muistiinpanoissa lukee koehenkilön yskineen tietyllä aikavälillä), avoin mieli (yritys ja erehdys kuuluvat elämään ja datasta voi löytyä jotain aivan muutakin kuin haluttiin nähdä) ja suorituskykyarvio (hienokkaan koejärjestely ei ole hyödyllinen, mikäli se antaa puolesta tapauksista väärän tuloksen).