University of Ljubljana

Faculty of Electrical Engineering

Domen Novak

ADAPTIVE INTEGRATION OF PSYCHOPHYSIOLOGICAL VARIABLES FOR ROBOTIC TRAINING

Doctoral dissertation

Ljubljana, 2011

University of Ljubljana

Faculty of Electrical Engineering

Domen Novak

ADAPTIVE INTEGRATION OF PSYCHOPHYSIOLOGICAL VARIABLES FOR ROBOTIC TRAINING

Doctoral dissertation

Advisor: prof. dr. Matjaž Mihelj

Ljubljana, 2011

Acknowledgments

I would like to thank my advisor, Matjaž Mihelj, for his guidance during the course of my work. His excellent ideas, attention to detail, encyclopedic knowledge of robotics and great love and respect for psychology were an invaluable help.

Furthermore, this entire work was performed as part of the EU 7th Framework Programme project MIMICS and thus could not have been completed without the rest of my project colleagues. Jaka Ziherl and Andrej Olenšek prepared the ball-catching virtual scenario and the HapticMaster control, laying the technical foundations for psychophysiological research. Maja Milavec assisted with psychological methodology while Marko Munih watched over the entire project. Other members of the Laboratory of Robotics also helped, each in his own way. Nika Goljar, Metka Javh and the other staff at the University Rehabilitation Institute of the Republic of Slovenia made clinical studies possible. Abroad, other MIMICS partners enabled a useful exchange of ideas and objective evaluation of the work. Alexander Koenig and Robert Riener deserve special thanks in this respect.

A large number of volunteers, both healthy and post-stroke, participated as experiment subjects. While there were too many to list here, each one deserves my special thanks for participation in sometimes tedious and aggravating tasks. Anja Podlesek of the university's Department of Psychology provided valuable advice regarding the statistical analysis of the gathered data.

On a lighter note, this work also could not have been completed without a sufficient dose of coffee, which was luckily provided by a not insignificant number of people. The group includes Marko and Jernej (2008), Matic (2009-present), Jure (2010-present) and Dejan (2008-present) as its most proficient members here at the Faculty of Electrical Engineering. Of course, I also thank all my other friends who stood by me during the last few years.

Finally, I would like to thank my parents for everything. And I hope that my grandfather will now believe that I am really an engineer, even if I still can’t fix his radio.

Table of contents

Abstract

Robotic interfaces are becoming increasingly common in motor rehabilitation. In the long term, exercise with such devices yields results comparable to intensive exercise with a therapist. Additionally, they offer an objective estimation of the patient's motor performance and functional improvement. Frequently, they are combined with virtual environments in order to make rehabilitation more interesting and motivational. Patient-cooperative robots, which can recognize the patient's movement intentions and motor abilities, also adapt the robotic assistance to the activity (or passivity) of the patient. This dissertation extends the concept of patient-cooperative robotics to biocooperative robotics, where psychophysiological measurements (measurements of a person's physiological response to psychological stimuli) are used to estimate how difficult the rehabilitation task is for the patient and adapt the difficulty to make it moderately challenging without inducing boredom or stress.

The dissertation covers the use of four psychophysiological measurements in upper extremity rehabilitation: heart rate, skin conductance, respiration and skin temperature. The research can be divided into two parts. The first covers the analysis of rehabilitation-specific factors which could adversely affect psychophysiological responses: the presence of physical activity and pathological conditions.

The effects of physical activity on psychophysiological responses in haptic interaction were examined in a study where 30 healthy subjects performed an inverted pendulum balancing task with a haptic robot at two levels of physical workload and three levels of cognitive workload. Heart rate and skin conductance were primarily influenced by physical workload. Neither respiration nor peripheral skin temperature were significantly affected by physical workload. Respiratory rate variability decreased from baseline during the moderately cognitively challenging condition while skin temperature decreased during the cognitively overchallenging condition. This suggests that respiration and skin temperature are effective for the estimation of cognitive workload in haptic interaction.

The effects of pathological conditions on psychophysiological responses in rehabilitation were examined in a study where 23 stroke and 23 control subjects performed a virtual

rehabilitation task and a simple cognitive task (the Stroop word-colour interference task). Significant differences between stroke and control groups were found especially for heart rate and peripheral skin temperature, with the stroke group exhibiting weaker responses to both the rehabilitation task and the cognitive task. Skin conductance appears to be the most useful psychophysiological signal in stroke subjects, as there is a significant correlation with self-reported arousal as well as a significant difference between different difficulty levels of the virtual rehabilitation task. However, we must keep in mind that it is also affected by physical activity.

The second part of the dissertation deals with psychophysiological sensor fusion and biocooperative control in upper extremity rehabilitation. A review of the existing psychophysiological literature was performed, and a number of different dimension reduction and classification methods were selected for implementation. Furthermore, the method of adaptive discriminant analysis, which had previously only been used in electroencephalography, was transferred to the signals used in the dissertation and used to perform online adaptation of the sensor fusion rules.

Sensor fusion was first implemented with a physically undemanding cognitive task performed by 20 healthy subjects. This allowed different data fusion methods to be tested in a relatively controlled setting. The subjects performed the Corsi block-tapping task at different difficulty levels for six two-minute periods while task performance and psychophysiological signals were recorded. At the end of each period, they stated whether they would prefer difficulty to increase or decrease. Various features were extracted from the signals, and the aforementioned dimension reduction and classification methods were trained to classify these features into an estimate of whether difficulty should increase or decrease, and the accuracy rate of a method was defined as the percentage of times that it matched the subject's opinion. In cross-validation, the highest accuracy rate for psychophysiological features was 86.7% using adaptive discriminant analysis, which outperformed other classifiers. The highest accuracy rate using both psychophysiological features and task performance was approximately the same. This validates the sensor fusion approach in a non-rehabilitation setting.

Sensor fusion was then implemented with the upper extremity rehabilitation task previously used in the analysis of the effects of pathological conditions. 24 healthy subjects and 11 patients performed the task at different difficulty levels for six two-minute periods while task performance, biomechanical signals and psychophysiological signals were recorded. At the end of each period, they stated whether they would prefer difficulty to increase or decrease. The same dimension reduction and classification methods were used again on the same extracted features. In cross-validation, the highest accuracy rate was obtained for task performance (approximately 82% for both healthy subjects and patients), with psychophysiological measurements yielding relatively low accuracy (approximately 60%). The adaptive approach increased accuracy of psychophysiological measurements to approximately 77% for both healthy subjects and patients. Combining psychophysiology with task performance yielded an accuracy rate of 84.7% for healthy subjects and 89.4% for patients.

Finally, a biocooperative feedback loop was implemented in the upper extremity rehabilitation task. The measurements from the aforementioned 24 healthy subjects and 11 patients were used to train the data fusion rules, and the system then automatically adapted the task difficulty according to its own estimates. 10 healthy subjects and 6 patients tested the feedback loop and provided their own opinions. The decisions taken by the biocooperative controller matched the subjects' opinions in 88.3% of all cases for healthy subjects and 88.9% of all cases for patients.

Results of the dissertation suggest that psychophysiological measurements are not reliable as a primary data source in motor rehabilitation, especially since they are influenced by both physical activity and pathological conditions. They can provide supplementary information that complements task performance and biomechanical measurements. However, at this point it is questionable whether the amount of additional information justifies the increased complexity of the system, though several possibilities for improvement are presented in the final discussion.

Key words: psychophysiological measurements, rehabilitation robotics, biocooperative robotics, stroke, sensor fusion, machine learning

Razširjeni povzetek

Uvod

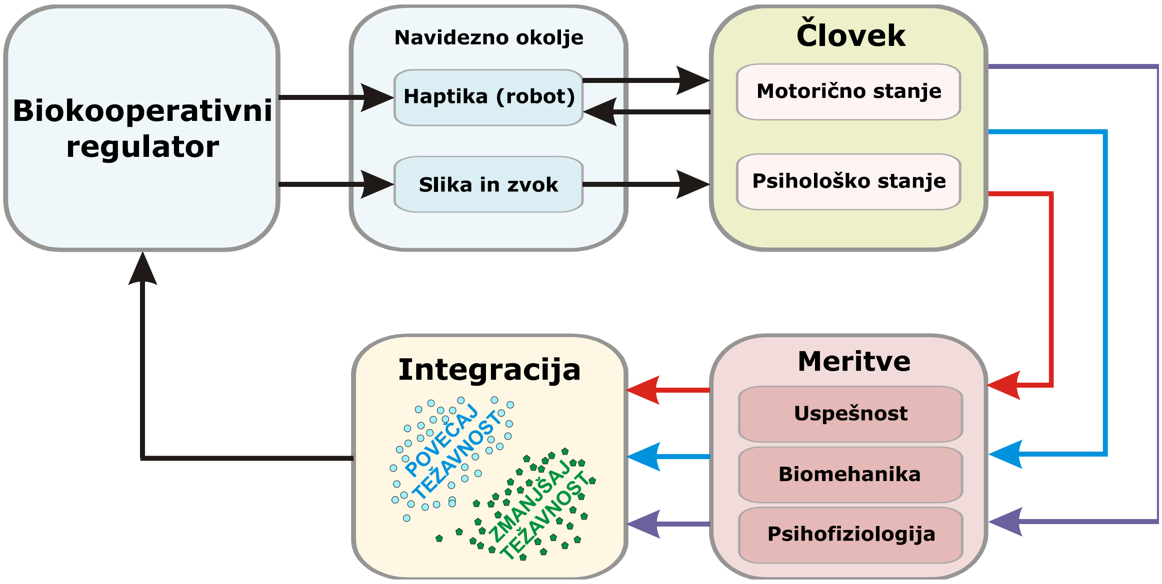
Rehabilitacijski roboti so naprave, ki pacientom z oslabljenimi zmožnostmi gibanja (kot posledica možganske kapi, poškodbe hrbtenjače ali drugih poškodb) pomagajo pri okrevanju in ponovnem učenju gibov. Ti roboti imajo več prednosti. Natančni senzorji sil in pozicij omogočajo objektivno ocenjevanje pacientovih gibalnih sposobnosti [1], aktivni motorji pa lahko pacientu pomagajo izvajati preproste ali tudi bolj zapletene gibe ter tako razbremenijo terapevta [2]. Rezultati dolgoročne vadbe s takimi napravami so primerljivi z rezultati vadbe s pomočjo terapevta [3]. Rehabilitacijski roboti se pogosto uporabljajo v kombinaciji z navideznimi okolji, ki nudijo bolj zanimivo in motivacijsko vadbo [4].

V rehabilitaciji so trenutno najbolj uveljavljene takoimenovane »pacientu prijazne« metode, ki zaznavajo pacientove gibalne sposobnosti in hotene aktivnosti ter robotsko pomoč prilagodijo gibom, ki jih izvaja pacient. Uspešno se uporabljajo pri rehabilitaciji tako spodnjih [5, 6] kot zgornjih okončin [7, 8]. Koncept robotske pomoči in prilagajanja je bil nedavno razširjen na biokooperativno robotiko, ki poleg izmenjave (mehanske) energije med pacientom in robotom upošteva tudi psihološke faktorje (npr. dolgčas, stres). V biokooperativni rehabilitacijski nalogi se parametri naloge samodejno spreminjajo tako, da naloga stalno predstavlja zmeren izziv, ki pacienta motivira in ne povzroča nepotrebnega stresa oziroma bolečine. Osnovna ideja biokooperativne robotike je sicer že dobro definirana [9, 10], vendar pa delujoča implementacija še ne obstaja, saj je merjenje psiholoških faktorjev veliko težje od merjenja sil ali hitrosti. V psiholoških raziskavah se za ta namen sicer najpogosteje uporabljajo vprašalniki, vendar le-ti niso dobra rešitev za rehabilitacijo, saj zahtevajo prekinitev vadbe in tako nudijo le informacije o že preteklih dogodkih. Obetavna možna rešitev bi bila uporaba takoimenovanih psihofizioloških meritev.

Psihofiziologija je znanost, ki preučuje odzive telesa na psihološke dražljaje. Dobro znan primer psihofiziološkega odziva je naprimer povečano znojenje in spremenjen srčni utrip kot posledica stresa. Fiziološke odzive lahko merimo hitro, poceni in nemoteče za merjenca, zato predstavljajo priročen in objektiven način ocenjevanja človekovega psihološkega stanja brez njegovega aktivnega sodelovanja. Največjo oviro predstavlja interpretacija merjenih odzivov. Povezave med fiziološkimi odzivi in psihološkimi stanji namreč niso dobro raziskane, obstaja pa tudi več motečih faktorjev: vpliv okolja (npr. vročina), vpliv gibanja in drugih fizično zahtevnih aktivnosti, lezenje signalov, razlike med posamezniki itd.

Psihofiziološke meritve so bile najprej (in so še vedno) uporabljene v laboratorijskih eksperimentih, kot so naprimer študije fizioloških odzivov na slike in zvoke [11], vendar pa so se zaradi svojih prednosti kmalu pojavile tudi v aplikativnih raziskavah. Pogosto se naprimer uporabljajo za preučevanje miselne obremenitve v okoljih, kot je nadzor zračnega prometa [12], simulirano letenje [13] ali vožnja avtomobila [14]. Primerne so tudi za analizo stresa in anksioznosti v situacijah, kot je javno nastopanje [15]. Najdemo pa jih tudi v manj stresnih okoljih, naprimer v računalniških igrah [16] in robotiki [17, 18].

Psihofiziološke meritve bi bile lahko zelo uporabne v biokooperativni rehabilitacijski robotiki, saj bi z njimi lahko hitro in objektivno zaznali nezaželena stanja, kot so zdolgočasenost ali stres. Vendar pa se v rehabilitaciji pojavlja več zahtevnih izzivov za psihofiziologijo (npr. prisotnost fizične aktivnosti in patoloških stanj), zato je bilo na tem področju izvedenih le malo raziskav. Disertacija tako preučuje potencialno uporabnost psihofizioloških meritev v rehabilitaciji zgornjih okončin, še posebej v biokooperativni povratni zanki, ki bi samodejno spreminjala parametre naloge, tako da bi bila naloga čim bolj primerna za pacienta. Glavni cilj disertacije torej ni ugotavljanje psiholoških stanj iz psihofizioloških meritev, marveč le ugotavljanje primernosti naloge za pacienta. Princip biokooperativne povratne zanke je prikazan na sliki 1.



Slika 1: Biokooperativna povratna zanka.

Disertacija se osredotoča na štiri psihofiziološke odzive, ki jih oživčuje vegetativno živčevje: srčni utrip, prevodnost kože, dihanje in temperatura kože. Iz surovih izmerjenih signalov izluščimo več psihofizioloških značilk. Iz srčnega utripa izračunamo povprečno frekvenco srčnega utripa ter več ocen spremenljivosti srčnega utripa v časovnem in frekvenčem prostoru. Iz prevodnosti kože izračunamo nivo prevodnosti kože, frekvenco odzivov prevodnosti kože ter povprečno amplitudo prevodnosti kože. Iz dihanja izračunamo povprečno frekvenco dihanja ter spremenljivost frekvence dihanja. Iz temperature kože izračunamo temperaturo kože ob koncu vsakega časovnega obdobja. Vse to so standardne značilke, ki se pogosto uporabljajo v psihofizioloških raziskavah [19]. V vseh haptičnih in rehabilitacijskih nalogah je uporabljen haptični robot HapticMaster [20].

Raziskovalno delo lahko razdelimo na dva dela. Prvi predstavlja analizo vplivov dveh motilnih faktorjev, specifičnih za rehabilitacijo, drugi pa senzorno integracijo in biokooperativno vodenje.

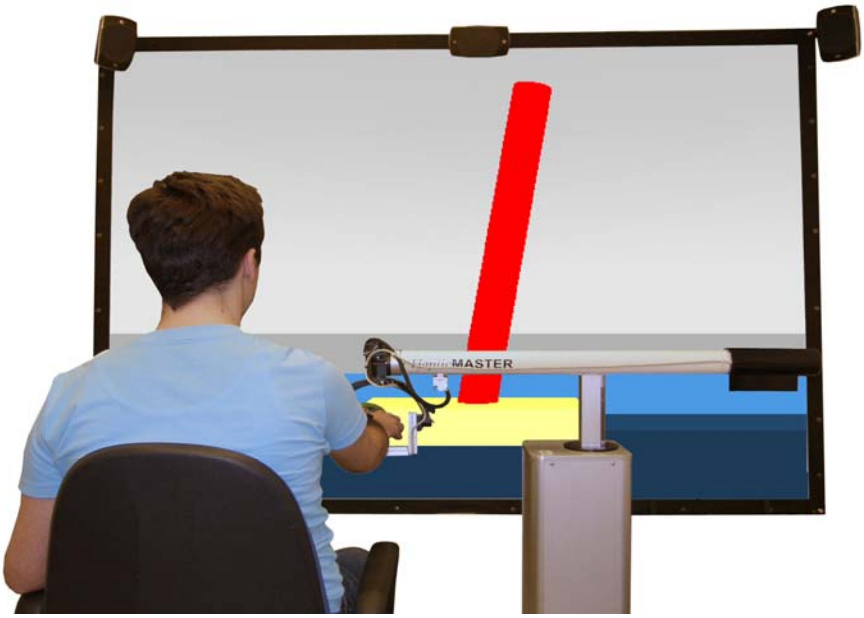
Vpliv motilnih faktorjev na psihofiziološka merjenja v rehabilitaciji

V motorični rehabilitaciji je bilo izvedeno le malo psihofizioloških raziskav, saj dva faktorja otežujeta analizo in interpretacijo meritev. Prvi faktor je prisotnost močnega fizičnega napora, ki vpliva na fiziološke odzive. V večino psihofizioloških študij je zato fizični napor nezaželen in zmanjšan na minimalni možni nivo. V rehabilitaciji to ni možno, saj je fizični napor bistven del vadbe. Več študij je že preučevalo vpliv kombinacije fizičnega in miselnega napora na psihofiziološke odzive, vendar pa so se osredotočale na situacije, kjer sta miselni in fizični napor popolnoma ločena (npr. računanje med vožnjo s kolesom) [21, 22]. Merjenci so tako obenem opravljali dve nepovezani nalogi. Pri delu s haptičnimi roboti pa ena sama naloga zahteva tako fizični kot miselni napor, kar bi lahko privedlo do drugačnih psihofizioloških odzivov. Tako smo želeli ugotoviti, ali lahko s psihofiziološkimi odzivi ločujemo med različnimi nivoji miselnega napora pri različnih nivojih fizičnega napora.

Drugi faktor, ki otežuje psihofiziološke raziskave v rehabilitaciji, je prisotnost poškodb vegetativnega živčnega sistema v merjencih. Posledice možganske kapi so naprimer dolgotrajne motnje znojenja in bitja srca [23]. Podobno tudi poškodbe možganov zaradi udarcev oslabijo psihofiziološke odzive [24]. Vendar pa psihofiziološki odzivi pacientov še nikoli niso bili preučevani med motorično rehabilitacijo samo. Tako smo želeli pred začetkom senzorne integracije z analizo ugotoviti, kateri psihofiziološki odzivi so v pacientih oslabljeni oziroma sploh odsotni. Zaradi omejenega števila razpoložljivih pacientov smo se osredotočili na paciente po možganski kapi.

Vpliv fizične aktivnosti na psihofiziološka merjenja

Analiza vpliva fizične aktivnosti je zajemala trideset merjencev, ki so opravljali fizično in miselno zahtevno haptično nalogo pri različnih stopnjah fizičnega in miselnega napora. Za nalogo smo izbrali navidezno obrnjeno nihalo, pritrjeno na voziček (slika 2). Obrnjeno nihalo je samo po sebi nestabilno in pade, če tega ne preprečimo s stalnim premikanjem vozička levo in desno. Voziček se premika v isto smer in z isto hitrostjo kot haptični robot HapticMaster. Če nihalo pade na tla, se samodejno vrne v skoraj navpičen položaj. Naloga ima dva možna nivoja fizične zahtevnosti. Višji nivo za enakovreden premik robota zahteva petkrat višjo silo kot nižji nivo. Naloga ima tudi tri možne nivoje miselne zahtevnosti. V nizkem nivoju nihalo nikoli ne pade, merjenec pa mora le premikati voziček levo in desno z zmerno hitrostjo. V srednjem nivoju so parametri fizičnega modela nastavljeni tako, da je uravnovešanje nihala zmerno zahtevno. V visokem nivoju miselne zahtevnosti na nihalo deluje bistveno močnejša težnost, nihalo pa se tudi slabše odziva na premike vozička. Nadalje je med premik vozička in vpliv premika na nihalo dodana polsekundna zakasnitev, s čimer uravnovešanje nihala postane veliko bolj zahtevno.

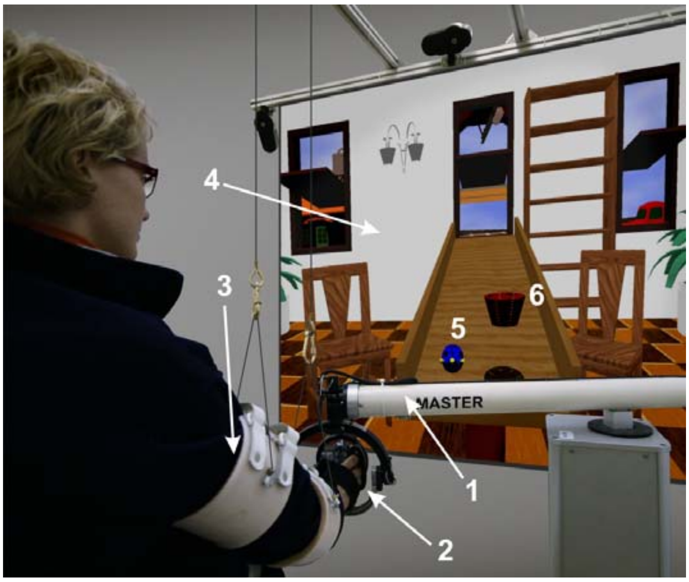


Slika 2: Merjenec opravlja nalogo z obrnjenim nihalom in robotom HapticMaster.

Vsak merjenec je opravljal nalogo v šestih različnih pogojih (2 nivoja fizične x 3 nivoji miselne zahtevnosti), vmes pa smo merili psihofiziološke odzive. Pri analizi rezultatov se je pokazalo, da fizična in miselna zahtevnost nista bili popolnoma ločeni (v nizkem nivoju miselne zahtevnosti so bili merjenci bistveno bolj fizično aktivni kot v ostalih dveh nivojih), vseeno pa smo pridobili relevantne informacije. Analiza je pokazala statistično značilen vpliv fizičnega napora predvsem na značilke, izluščene iz srčnega utripa in prevodnosti kože. Ta vpliv lahko popolnoma prekrije fiziološke odzive, povezane z miselnim naporom. Vpliv fizičnega napora na dihanje in temperaturo kože ni bil statistično značilen, čeprav smo pri dihanju opazili manjše neznačilne razlike med nivojema fizične zahtevnosti. Spremenljivost frekvence dihanja je ločevala med srednjim in visokim nivojem miselne zahtevnosti, saj se je le v srednjem nivoju statistično značilno zmanjšala z vrednosti v mirovanju. Temperatura kože je ločevala med visokim nivojem in drugima dvema nivojema miselne zahtevnosti, saj se je le v visokem nivoju statistično značilno zmanjšala z vrednosti v mirovanju. Tako bi lahko dihanje in temperatura kože bila učinkovita pokazatelja miselnega napora v fizično zahtevni haptični interakciji. Seveda pa so te ugotovitve veljavne le za zdrave osebe, ne za paciente po možganski kapi.

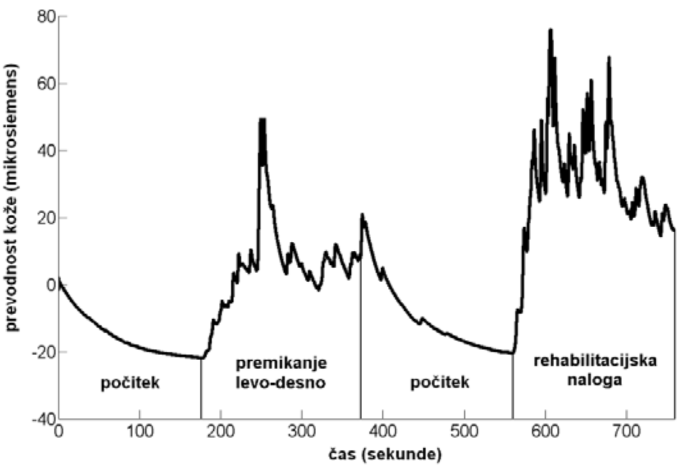
Primerjava odzivov zdravih oseb in oseb po možganski kapi

Analiza posledic možganske kapi je zajemala 23 pacientov v subakutnem obdobju po možganski kapi ter kontrolno skupino 23 zdravih oseb istega spola in starosti. Opravljali so tri naloge: haptično nalogo brez miselnega napora (premikanje haptičnega robota levo in desno), fizično in miselno zahtevno rehabilitacijsko nalogo ter miselno zahtevno nalogo brez fizičnega napora (Stroopov interferenčni test [25]). Rehabilitacijska naloga je prikazana na sliki 3. Na sredini ekrana je miza, nagnjena navzdol proti merjencu. Na vrhu mize se pojavi žoga, ki se začne kotaliti navzdol. Naloga merjenca je, da s pomočjo haptičnega vmesnika prime žogo, preden le-ta doseže spodnji rob mize. Ko je žoga prijeta, se nad mizo pojavi koš. Merjenec mora žogo držati in jo postaviti v koš. Ko je žoga v košu ali pa pade z mize, se na vrhu mize pojavi nova žoga in naloga se nadaljuje. Uporabili smo običajno in težjo različico naloge. Pri težji različici je premik robota v levo premaknil navidezno roko na ekranu v desno smer in obratno, zato je bil za uspešno opravljanje naloge potreben dodaten miselni napor. Pacientom, ki niso zmogli samostojno opraviti naloge, je bila na voljo tudi haptična pomoč.

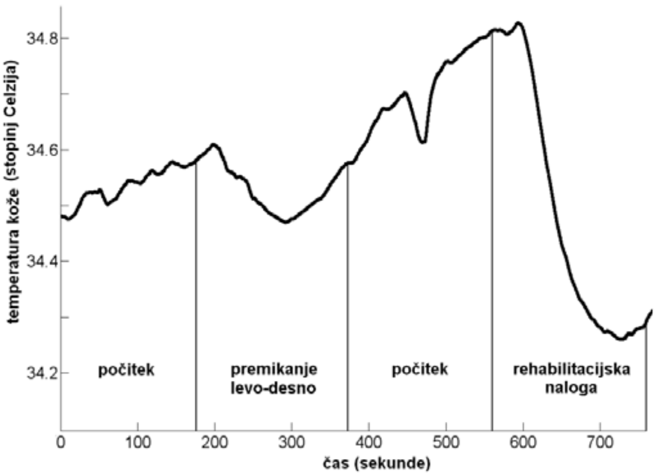


Slika 3: Merjenka opravlja rehabilitacijsko nalogo s haptičnim robotom (1) in prijemalom (2), medtem ko je njena roka podprta z manšetama. Na ekranu (4) so vidni nagnjena miza, žoga (5) in koš (6).

Postopek merjenja je bil sledeč: počitek, premikanje robota levo in desno, počitek, rehabilitacijska naloga, težja rehabilitacijska naloga, počitek, Stroopov test. Vsak interval je trajal tri minute. Po vsakem intervalu je merjenec tudi izpolnil kratek vprašalnik. Tipična poteka prevodnosti kože in temperature kože med merjenjem sta prikazana na slikah 4 in 5.



Slika 4: Primer poteka prevodnosti kože med počitkom, premikanjem robota levo in desno, počitkom in rehabilitacijsko nalogo. Začetna vrednost je bila definirana kot ničelna.



Slika 5: Primer poteka temperature kože med počitkom, premikanjem robota levo in desno, počitkom in rehabilitacijsko nalogo.

Za najbolj uporaben psihofiziološki signal v pacientih se je izkazala prevodnost kože. Nivo prevodnosti kože je statistično značilno ločeval med premikanjem robota levo in desno, rehabilitacijsko nalogo in težjo rehabilitacijsko nalogo. Korelacija med frekvenco odzivov prevodnosti kože in samooceno miselne budnosti (ang. *arousal*) je bila prav tako statistično značilna ($\rho = 0.60$). Temperatura kože je statistično značilno ločevala med premikanjem robota levo in desno, rehabilitacijsko nalogo in težjo rehabilitacijsko nalogo pri kontrolni skupini, ne pa pri pacientih. Tudi v Stroopovem testu se je temperatura kože statistično značilno spremenila od mirovanja do konca naloge samo pri kontrolni skupini. Povprečna frekvenca bitja srca je bila pri pacientih višja kot pri kontrolni skupini, spremenljivost bitja srca pa nižja. Tudi razlike v bitju srca med mirovanjem in opravljanjem nalog so bile pri pacientih manjše kot pri kontrolni skupini.

Iz analiz vpliva fizične aktivnosti in posledic možganske kapi smo tako ugotovili, da na vse merjene psihofiziološke odzive vpliva bodisi fizična aktivnost bodisi kap. Vseeno smo se odločili za nadaljnje raziskave senzorne integracije in biokooperativnega vodenja, vendar pa smo se pri tem zavedali, da bo senzorna integracija pri pacientih zaradi posledic kapi verjetno bolj zahtevna kot pri zdravih osebah. Nadalje smo pričakovali, da bo fizična aktivnost vplivala na psihofiziološke odzive, vendar pa še nismo vedeli, ali bo zaradi fizične aktivnosti težje ali lažje določiti primernost naloge za pacienta. Fizična aktivnost bi lahko namreč po eni strani zabrisala informacije o psiholoških stanjih, po drugi strani pa bi lahko nudila dodatne informacije o primernosti naloge.

Senzorna integracija in biokooperativno vodenje

Končni cilj senzorne integracije in biokooperativnega vodenja je povratna zanka, ki bi iz psihofizioloških in drugih meritev ocenila primernost naloge za pacienta ter spremenila parametre naloge, tako da bi bila le-ta bolj primerna za pacienta. Za to moramo najprej določiti pravila za interpretacijo psihofizioloških meritev, vendar pa kljub več desetletjem raziskav na tem področju še vedno ne obstaja splošno sprejet nabor pravil, po katerih bi lahko iz psihofizioloških meritev določili psihološko stanje posameznika. Prvi korak pri senzorni integraciji je bil torej pregled literature in identifikacija več možnih metod senzorne integracije za nadaljnjo implementacijo. Pregledali smo tudi že obstoječe primere psihofizioloških povratnih zank na drugih področjih (zunaj rehabilitacije).

Ugotovili smo, da večina znanstvenih študij s področja integracije psihofizioloških spremenljivk uporablja nadzorovano učenje: učenje na podlagi učne množice primerov s pripadajočimi izhodnimi vrednostmi (tj. psihološkimi stanji). Nadalje v psihofiziologiji uporabljajo predvsem klasifikacijo: razporeditev meritev v enega od možnih diskretnih razredov. Če imamo na voljo učno množico, lahko na njej preizkusimo več različnih klasifikatorjev in ugotovimo, kateri je v našem primeru najbolj učinkovit. Implementirali smo sledeče v psihofiziologiji že uveljavljene klasifikatorje:

- linearno in kvadratično diskriminantno analizo,
- diagonalno linearno in kvadratično diskriminantno analizo (tip naivnega Bayesovega klasifikatorja),
- metodo najbližjih sosedov,
- klasifikacijsko drevo,
- metodo podpornih vektorjev.

Ker imamo na voljo veliko psihofizioloških značilk, je pred klasifikacijo smiselno zmanjšati dimenzijo podatkov. To smo storili na dva načina: s preslikavo vhodnih značilk na manjše število značilk s pomočjo linearne transformacije (metoda analize glavnih komponent) oziroma z izločanjem manj pomembnih značilk (metoda sekvenčnega iskanja značilk – ang. *sequential feature selection*).

Poleg že uveljavljenih klasifikatorjev smo uporabili tudi adaptivno diskriminantno analizo [26], ki je bila predtem uporabljana le v elektroencefalografiji. Prednost te metode je, da se začetna klasifikacijska pravila nauči iz učne množice (torej že izmerjenih pacientov), nato pa jih postopoma prilagaja posameznemu pacientu s pomočjo Kalmanovega filtra. Ker pa je adaptivna diskriminantna analiza v osnovi različica nadzorovanega učenja, mora za prilagajanje klasifikacijskih pravil pridobiti mnenje pacienta oziroma terapevta. V praksi se želimo temu izogniti, zato smo predstavili tudi novo različico adaptivne diskriminantne analize, ki ne potrebuje nadzorovanega učenja.

Vse zgoraj naštete metode smo najprej uporabili v fizično nezahtevni nalogi z zdravimi merjenci, saj smo jih tako lahko preizkusili na psihofizioloških podatkih v nadzorovanem okolju brez vpliva fizične aktivnosti in kapi. Nato smo jih z enakim eksperimentalnim protokolom uporabili v rehabilitacijski nalogi s pacienti in tam implementirali tudi biokooperativno povratno zanko.

Senzorna integracija v nerehabilitacijski nalogi

Metode senzorne integracije smo najprej preizkusili z 20 zdravimi osebami, ki so opravljale računalniško verzijo Corsijeve naloge [27]. Na ekranu je razporejenih devet kvadratkov. Eden za drugim se obarvajo, merjenec pa mora nato z miško ponoviti prikazano zaporedje. Ko uspešno ali neuspešno ponovi zaporedje, se prične novo zaporedje iste dolžine. Prednost Corsijeve naloge je, da lahko njeno težavnost preprosto spreminjamo s spreminjanjem dolžine zaporedja, ki ga mora merjenec ponoviti.

Vsak merjenec je nalogo začel z zaporedji dolžine 5 kvadratkov. Naloga je trajala šestkrat po dve minuti (skupaj 12 minut). Znotraj vsakega dvominutnega intervala je bila dolžina zaporedja konstantna, ob koncu intervala pa smo merjenca vprašali, ali bi raje videl/a, da se težavnost poveča ali zmanjša. Dolžina zaporedij je bila nato v naslednjem intervalu večja ali manjša glede na željo merjenca. Za vsak interval smo ločeno izračunali psihofiziološke značilke ter uspešnost v nalogi. Nato smo zgradili učno množico, kjer so vhodi različne značilke, izhod pa merjenčeva želja (povečaj / zmanjšaj težavnost). S tem se senzorna integracija prevede na klasifikacijo v dva možna razreda. Za klasifikacijo smo uporabili vse zgoraj omenjene metode in princip navzkrižne validacije. Točnost klasifikacije je bila definirana kot odstotek primerov, kjer sta bila izhod klasifikatorja in merjenčeva želja enaka.

Točnost klasifikacije je bila primerljiva za vse v psihofiziologiji že uveljavljene metode. Točnost klasifikacije na podlagi psihofizioloških značilk z uveljavljenimi metodami je bila 75.0%, točnost klasifikacije na podlagi uspešnosti v nalogi pa 80.8%. Točnost klasifikacije na podlagi psihofizioloških značilk in uspešnosti skupaj je bila 85.0%. Adaptivna diskriminantna analiza, ki do zdaj v psihofiziologiji ni bila uporabljana, je izboljšala točnost klasifikacije na podlagi psihofizoloških značilk na 86.7%.

Iz rezultatov lahko sklepamo, da so izbrane metode primerne za ocenjevanje primernosti naloge iz psihofizioloških značilk in uspešnosti naloge, če jih uporabimo v nadzorovani laboratorijski nalogi z relativno homogeno skupino merjencev. Nadalje adaptivna diskriminantna analiza lahko opazno izboljša točnost klasifikacije na podlagi psihofizioloških značilk. Še vedno pa je bilo potrebno preučiti učinkovitost senzorne integracije v rehabilitaciji, kjer nastopajo bolj heterogene skupine merjencev, fizična aktivnost in patološka stanja.

Senzorna integracija in biokooperativno vodenje v rehabilitaciji

Metode senzorne integracije smo preizkusili v nalogi, ki smo jo uporabili že za analizo posledic možganske kapi. Implementirali smo sedem težavnostnih nivojev, ki so se med seboj razlikovali po hitrosti in velikosti žoge, ki jo je bilo potrebno ujeti. Pri nizkih težavnostih je bila žoga velika in počasna, pri visokih težavnostih pa majhna in hitra.

Raziskavo smo razdelili na odprtozančni del in zaprtozančni del. V odprtozančnem delu je sistem le meril signale in beležil merjenčeve želje. Iz pridobljenih podatkov smo nato zgradili pravila za senzorno integracijo in jih v zaprtozančnem delu uporabili za samodejno prilagajanje težavnosti naloge. V odprtozančnem delu je sodelovalo 24 zdravih oseb in 11 pacientov, v zaprtozančnem delu pa 10 zdravih oseb in 6 pacientov.

Merilni postopek za oba dela je bil podoben. Vsak merjenec je nalogo začel pri zmerni težavnosti. Naloga je trajala šestkrat po dve minuti (skupaj 12 minut). Znotraj vsakega dvominutnega intervala je bila težavnost konstantna, ob koncu intervala pa smo merjenca vprašali, ali bi raje videl/a, da se težavnost poveča ali zmanjša. Težavnost naloge je bila nato v naslednjem intervalu večja ali manjša glede na željo merjenca (odprtozančni del) ali glede na oceno klasifikatorja (zaprtozančni del). Za vsak interval smo ločeno izračunali psihofiziološke značilke, biomehanske značilke ter uspešnost v nalogi. Nato smo zgradili učno množico, kjer so vhodi različne značilke, izhod pa merjenčeva želja (povečaj / zmanjšaj težavnost). S tem se senzorna integracija znova prevede na klasifikacijo v dva možna razreda. Za klasifikacijo smo uporabili vse zgoraj omenjene metode in princip navzkrižne validacije. Točnost klasifikacije je bila definirana kot odstotek primerov, kjer sta bila izhod klasifikatorja in merjenčeva želja enaka.

Za rezultate odprtozančnega dela smo opravili primerjavo različnih tipov podatkov. Pri klasifikaciji na podlagi uspešnosti v nalogi je bila točnost približno 80% tako za zdrave osebe kot za paciente. Pri klasifikaciji na podlagi biomehanskih značilk je bila točnost približno 75% tako za zdrave osebe kot paciente. Najmanj točna je bila klasifikacija na podlagi psihofizioloških značilk – približno 63% tako za zdrave osebe kot paciente pri uporabi v psihofiziologiji uveljavljenih klasifikatorjev. Z adaptivno diskriminantno analizo lahko izboljšamo točnost klasifikacije na podlagi psihofizioloških značilk na približno 77%, kar pa je še vedno nižje od točnosti klasifikacije na podlagi uspešnosti v nalogi. Tako psihofiziološke meritve niso primerne kot zanesljiv primarni vir informacij v biokooperativni rehabilitacijski robotiki. Z integracijo različnih tipov podatkov (uspešnost, biomehanika in psihofiziologija) in uporabo metod zmanjševanja dimenzije podatkov lahko točnost klasifikacije povečamo na 84.7% za zdrave osebe in 89.4% za paciente. Najpomembnejša značilka je bila uspešnost v nalogi, vendar pa lahko psihofiziološke značilke nudijo dodatne informacije, ki izboljšajo klasifikacijo.

V zaprtozančnem delu poskusa smo implementirali biokooperativno vodenje na podlagi vseh možnih značilk (uspešnost, biomehanika in psihofiziologija), zmanjševanja dimenzije podatkov in klasifikacije z diskriminantno analizo. Točnost klasifikacije v zaprtozančnem delu je bila 88.3% za zdrave osebe in 88.9% za paciente. S tem smo uspešno implementirali in preizkusili biokooperativno povratno zanko, ki prilagaja težavnost naloge na podlagi senzorne integracije psihofizioloških in drugih spremenljivk.

Zaključki

Disertacija se je osredotočala na dokaj neraziskano področje psihofiziologije v motorični rehabilitaciji, še posebej na uporabo psihofizioloških spremenljivk v biokooperativni povratni zanki: sistemu, ki na podlagi psihofizioloških in drugih spremenljivk prilagodi nalogo, tako da je le-ta čim bolj primerna za trenutnega pacienta. Merili smo štiri psihofiziološke odzive: srčni utrip, prevodnost kože, dihanje in temperaturo kože.

Z analizo vplivov fizične aktivnosti in posledic kapi smo ugotovili, da na vse štiri merjene psihofiziološke odzive vpliva bodisi fizična aktivnost bodisi možganska kap. Merjene odzive smo nato uporabili v senzorni integraciji in preizkusili več klasifikacijskih metod. Za najboljšo se je izkazala adaptivna diskriminantna analiza, ki lahko bistveno izboljša točnost klasifikacije na podlagi psihofizioloških značilk in ima primerljivo točnost kot druge metode v primeru nepsihofizioloških značilk. Bistvena je tudi primerna izbira vhodnih značilk. To izbiro lahko opravimo z metodami zmanjševanja dimenzije podatkov.

V rehabilitacijski nalogi senzorna integracija psihofizioloških značilk ni bila zelo točna. Bistveno boljšo točnost smo dosegli z upoštevanjem uspešnosti v nalogi ter biomehanskih značilk. Psihofiziološke meritve tako niso primerne kot zanesljiv primarni vir informacij v biokooperativni rehabilitacijski robotiki, lahko pa služijo kot sekundarni vir informacij, ki poveča točnost. Vseeno pa je vprašljivo, ali dodatna točnost, ki jo nudijo psihofiziološke meritve, odtehta povečano ceno in kompleksnost sistema. Psihofiziološke meritve bi bile morda najbolj uporabne v nalogah in okoljih, kjer uspešnosti in biomehanskih značilk bodisi ni mogoče meriti bodisi niso povezane z merjenčevim počutjem.

Kljub ne najbolj spodbudnim rezultatom pa je možnih še veliko izboljšav, ki smo jih predlagali v poglavju Razprava in so lahko zelo preproste ali pa tudi zelo kompleksne. V disertaciji smo opravili prve korake k uvedbi biokooperativnega vodenja v rehabilitaciji, vključno z razvojem biokooperativne povratne zanke, ki vključuje psihofiziološke meritve. Biokooperativno vodenje predstavlja nadgradnjo “pacientu prijaznih” metod in poskuša robota približati vlogi fizioterapevta oziroma delovnega terapevta. Terapevt ima namreč celosten vpogled v pacientovo biomehansko, psihološko in fiziološko stanje, s “pacientu prijaznimi” metodami pa ima robot vpogled le v biomehansko stanje. Naša biokooperativna povratna zanka poskuša pridobiti tudi vpogled v psihološko in fiziološko stanje. Vprašanje, ali se bo ideja biokooperativnega vodenja uveljavila v klinični praksi, tako ostaja odprto, kljub deloma nespodbudnim rezultatom pa verjamemo, da bi bilo samodejno prilagajanje naloge pomembna izboljšava rehabilitacijskih robotov, ki bi lahko morda vodila v boljši izid rehabilitacije.

Izvirni znanstveni prispevki doktorske disertacije

Izvirni znanstveni prispevki disertacije so:

- Analiza psihofizioloških odzivov zdravih oseb na kombinacijo psihološke in fizične aktivnosti v haptični interakciji človeka in robota.
- Analiza psihofizioloških razlik med zdravimi osebami in hemiparetičnimi pacienti v kliničnih rehabilitacijskih okoljih.
- Integracija psihofizioloških senzorjev za ocenjevanje primernosti nalog v rehabilitacijski robotiki s pomočjo različnih metod.
- Adaptivna metoda, ki se lahko prilagodi na psihofiziološke razlike med posamezniki.
- Biokooperativni regulator, ki lahko prilagodi parametre rehabilitacijske naloge na podlagi adaptivne integracije psihofizioloških, biomehanskih in drugih senzorjev.

Ključne besede: psihofiziološke meritve, rehabilitacijska robotika, biokooperativna robotika, senzorna integracija, strojno učenje

List of figures

List of tables

List of abbreviations

ADIM – adaptive information matrix

ANN – artificial neural network

ANOVA – analysis of variance

BAS – behavioral activation system

BIS – behavioral inhibition system

ECG – electrocardiogram

FIM – Functional Independence Measure

HF – high-frequency (in this dissertation, specifically referring to heart rate variability)

HRV – heart rate variability

KALDA – Kalman adaptive linear discriminant analysis

kNN – *k*-nearest neighbors

LDA – linear discriminant analysis

LF – low-frequency (in this dissertation, specifically referring to heart rate variability)

MMSE – Mini-Mental State Examination

PCA – principal component analysis

pNN50 - percentage of interval differences of successive NN-intervals greater than 50 ms

QDA – quadratic discriminant analysis

RMSSD – square root of the mean squared differences of successive NN-intervals

SAM – Self-Assessment Manikin

SCL – skin conductance level

SCR – skin conductance response

SDNN – standard deviation of NN-intervals

SFFS – sequential floating forward selection

SVM – support vector machine

VR – virtual rehabilitation

1 Introduction

1.1 Rehabilitation robotics

Rehabilitation robots are devices that assist the recovery of patients whose motor functions are impaired as a result of stroke, spinal cord injury or other condition. Their benefit is twofold. First, they offer accurate sensors for measurement of forces and positions, thus providing a method of objectively evaluating the patient's motor performance [1]. Second, robots with active motors can help the patient train simple or complex movements, taking some of the strain off therapists [2]. Training with such robots yields long-term results comparable to exercise with a therapist [3]. Frequently, they are combined with virtual environments in order to make rehabilitation more interesting and motivational [4].

The first device that provided robotic training in rehabilitation was the MIT-Manus, a 2-degree-of-freedom system that supports planar movements using an impedance controller [28]. After its introduction, several rehabilitation systems based on different robots were developed, such as the GENTLE/s [29] or the ARMin [30]. However, while early rehabilitation robots were able to provide active assistance to the patient, they did not adapt their movement to the activity (or passivity) of the patient. Rather, the affected limb was moved along a predefined, fixed trajectory. The patient was also not informed about his or her activity and contribution to the movement. This problem was addressed by patient-cooperative or "assist as needed" control techniques.

By recognizing the patient's movement intentions and motor abilities, patient-cooperative techniques adapt the robotic assistance to the activity (or passivity) of the patient. They can thus control the amount of physical workload patients need to perform: essentially, the amount of energy they need to actively exert with their arms and legs. Patient-cooperative techniques have been successfully used for rehabilitation of both the lower extremities [5, 6] and upper extremities [7, 8]. However, physical workload is not the only important factor in rehabilitation. Cognitive workload, the amount of working memory and problem-solving skills that needs to be applied to a task in order to successfully complete it, also needs to be considered. If cognitive workload is too low, the patient will become bored. On the other hand, if cognitive workload is too high, the patient will become stressed and frustrated.

The concept of patient-cooperative robotics has thus recently been extended to biocooperative robotics, which take into account not only the bidirectional flow of energy between the patient and the robot, but also psychological factors. In a biocooperative rehabilitation task, the parameters of the task are automatically adjusted so that the patient is challenged in a moderate but engaging and motivating way without causing undue stress or harm. However, while this basic idea of biocooperative robotics has been defined [9, 10], no working implementation exists yet since measuring psychological factors is more difficult than measuring forces or velocities. While commonly used in psychological research, questionnaires are not a good solution for rehabilitation since they require therapy to be interrupted and only provide information ‘after the fact’. A promising potential solution would be the use of psychophysiological measurements.

1.2 Psychophysiological measurements

Physiology is the study of bodily function and how the parts of the body work. Psychophysiology is intimately related to it, but is also concerned with psychological phenomena – the experience and behavior of organisms in the physical and social environment [31]. It has been defined as “any research in which the dependent variable (the subject’s response) is a physiological measure and the independent variable (the factor manipulated by the experimenter) a behavioral one” [32]. Put more simply, psychophysiology is the study of the body’s responses to psychological stimuli. For instance, one classic psychophysiological response is the increased sweating and changes in heart rate as a response to stressful situations.

In its early years, psychophysiology mainly focused on the physiological responses and organ systems innervated by the autonomic nervous system (e.g. Ekman et al. [33]). More recently, psychophysiologists have been equally interested in the central nervous system, exploring techniques such as electroencephalography, magnetoencephalography, functional magnetic resonance imaging and other neuroimaging methods. Nonetheless, responses of the autonomic nervous system such as cardiorespiratory and electrodermal responses are still more popular in applied studies since they can be measured cheaply, quickly and unobtrusively, providing a convenient and objective method of estimating a person's psychological state without his or her active cooperation. Of course, though psychophysiological measurements are easy to measure, they are difficult to interpret for a number of reasons: uncertain connections between physiological responses and psychological states, intersubject variability, effects of the environment (e.g. heat), effects of physical activity, signal drift etc.

Though first (and still) used in laboratory experiments such as studies of physiological responses to isolated images and sounds [11], the advantages of psychophysiological measurements have allowed them to appear in many applied fields. For instance, they are frequently used to study mental workload in situations such as air traffic control [12], simulated flight [13] or driving [14]. They are also popular for analysis of stress and anxiety in, for example, public speaking [15]. On a lighter note, they are used to evaluate users' experiences with computer games [16] and robots [17, 18].

Psychophysiological measurements represent a promising potential addition to biocooperative rehabilitation robotics, as they would allow undesirable states such as stress or boredom to be objectively detected in real time. The basic idea of using psychophysiological measurements in motor rehabilitation has already been raised [34], but little concrete research has been carried out in this direction due to a variety of rehabilitation-specific challenges such as the presence of physical activity and pathological conditions. Furthermore, despite extensive research, there is still no standard method of interpreting psychophysiological measurements in general psychophysiology. Aside from theoretical limitations to inferring significance from psychophysiological data [35], there are entirely practical disagreements among psychophysiologists. One disagreement is how a subject's psychological state can be described. Here, there are two main approaches. The first tries to classify psychophysiological measurements into one of several basic emotions (anger, sadness, fear, surprise, happiness…) [36]. The second posits that a person's psychological state is multidimensional and thus described with multiple variables. The most popular multidimensional model in psychophysiology is the arousal-valence model [37]. Valence (sometimes also called pleasure) is defined as positive versus negative affective states (e.g., humiliation, disinterest, and anger at one end versus excitement, relaxation, and tranquility at the other) while *arousal* is defined in terms of mental alertness and physical activity (e.g., sleep, inactivity, boredom, and relaxation at the lower end versus wakefulness, tension, exercise, and concentration at the higher end) [38]. In applied psychophysiological studies, however, it is also common to find researchers using neither basic emotions nor arousal and valence, but other psychological quantities such as stress, frustration, mental workload etc.

The dissertation will thus be concerned with the exploration and interpretation of psychophysiological measurements in a motor rehabilitation setting, particularly in a biocooperative closed loop that could adjust the parameters of a rehabilitation task to make it optimally challenging for the patient.

1.3 Dissertation structure

The ultimate goal of the dissertation is to combine psychophysiological measurements with biomechanical and other measurements in a biocooperative feedback loop for rehabilitation robotics. This does not require precise identification of psychological states (e.g. angry, surprised, sad), but requires us to determine how suitable the task is for the patient and how the task parameters should be adjusted so that the patient is challenged in a moderate but engaging and motivating way without causing undue stress or harm. The basic blocks of such a biocooperative feedback loop are shown in Figure 1.1.

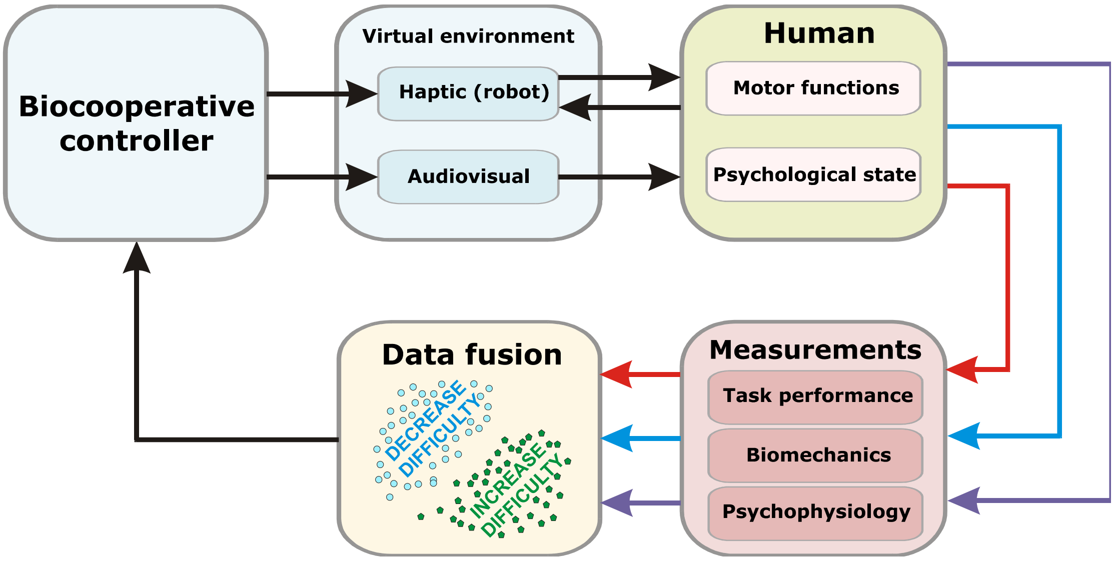


Figure 1.1: The principle of a biocooperative feedback loop. The human interacts with a haptic robot and virtual environment while different measurements are taken. These measurements are fused into an estimate of how difficulty should be changed, and the biocooperative controller changes the parameters of the virtual environment accordingly.

The dissertation focuses primarily on psychophysiological, data fusion and biocooperative challenges. The design and implementation of robots and virtual environments for rehabilitation, while a significant challenge, is not a part of this dissertation; rather, the dissertation makes use of already existing robots and environments (with some modifications necessary for implementation of biocooperative control). As previous work at the University of Ljubljana was mainly done on upper extremity rehabilitation, the dissertation also focuses on the upper extremities.

2 Hardware, psychophysiological measurements and questionnaires

2.1 HapticMaster and audiovisual display

The HapticMaster robot [20], developed by Moog FCS, was used as the haptic interface. Shown in Figure 2.1, this robot offers movement with three degrees of freedom. The first joint allows vertical translation, the second allows rotation around a vertical axis, and the third allows horizontal translation. The robot's end-effector also contains a three-axis force sensor. A two-axis gimbal with a wrist support mechanism and a two-degree-of-freedom passive grasping module instrumented with force cells (Figures 2.2 and 2.3) is attached to the end-point of the robot [39]. The subject places his/her hand on the wrist support mechanism, and the arm is then fixed to the grasping device at the wrist. While the wrist is thus fixed during movement, the subject can freely move his/her fingers, elbow and shoulder joints. The arm is additionally supported using two cuffs fastened above and below the elbow. These cuffs were connected to electric motor pulleys using Kevlar cables. The pulleys applied a constant torque in order to compensate for the gravity acting on the subject's arm.

A 1.4x1.4-meter screen was suspended from the ceiling, and a projector behind the screen displayed the image onto it. The bottom edge of the screen was approximately 50 centimeters from the ground. Subjects sat approximately 1.25 meters in front of the screen, with the robot situated between the seat and the screen. Additionally, five speakers and a subwoofer were arranged around the screen and subject in the Dolby 5.1 layout: two speakers to the left and right of the screen, one speaker atop the screen, two speakers to the left and right behind the subject and the subwoofer below the screen.

The HapticMaster has previously served as the basis of the GENTLE/s rehabilitation platform [29] and can thus be considered a suitable choice for robot-aided upper extremity rehabilitation. Though not yet extensively used in rehabilitation, the grasping module has proven promising in preliminary trials with chronic stroke patients [39] and should thus also be suitable for this research.

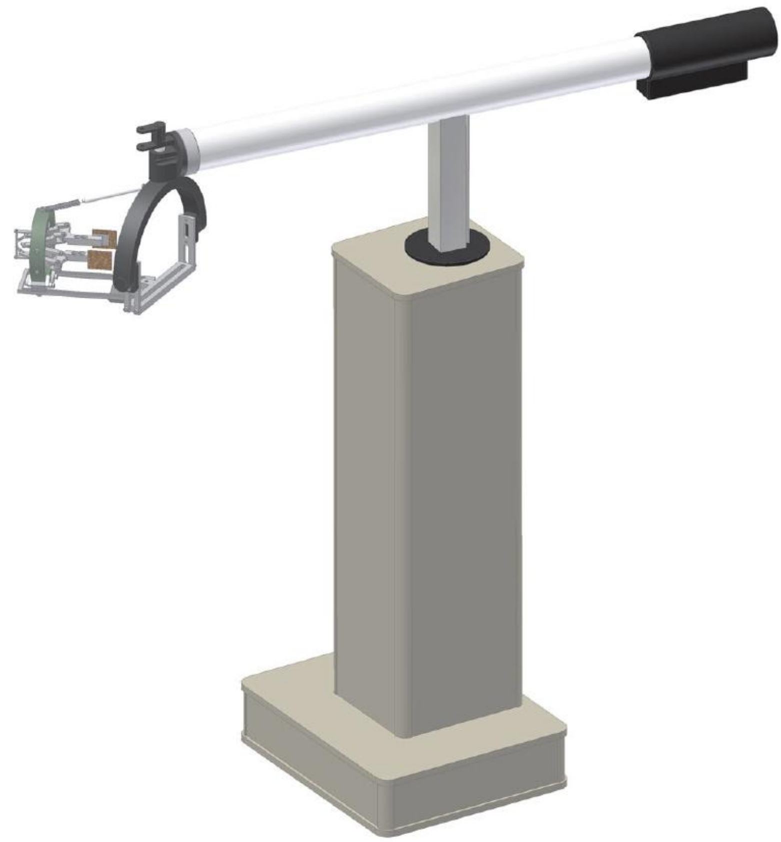


Figure 2.1: The HapticMaster.

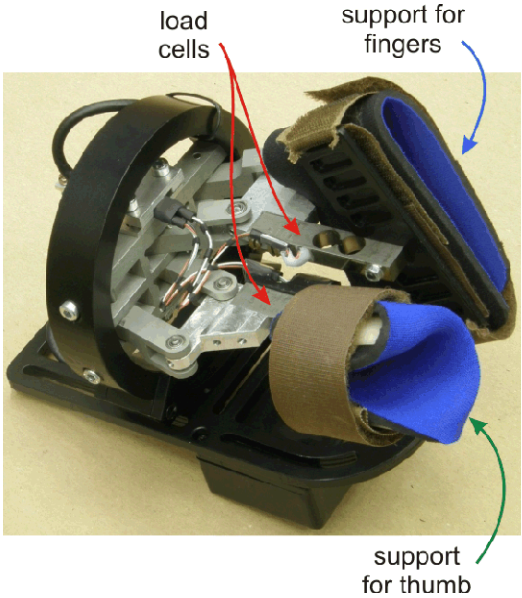


Figure 2.2: A photograph of the HapticMaster's grasping device [39].

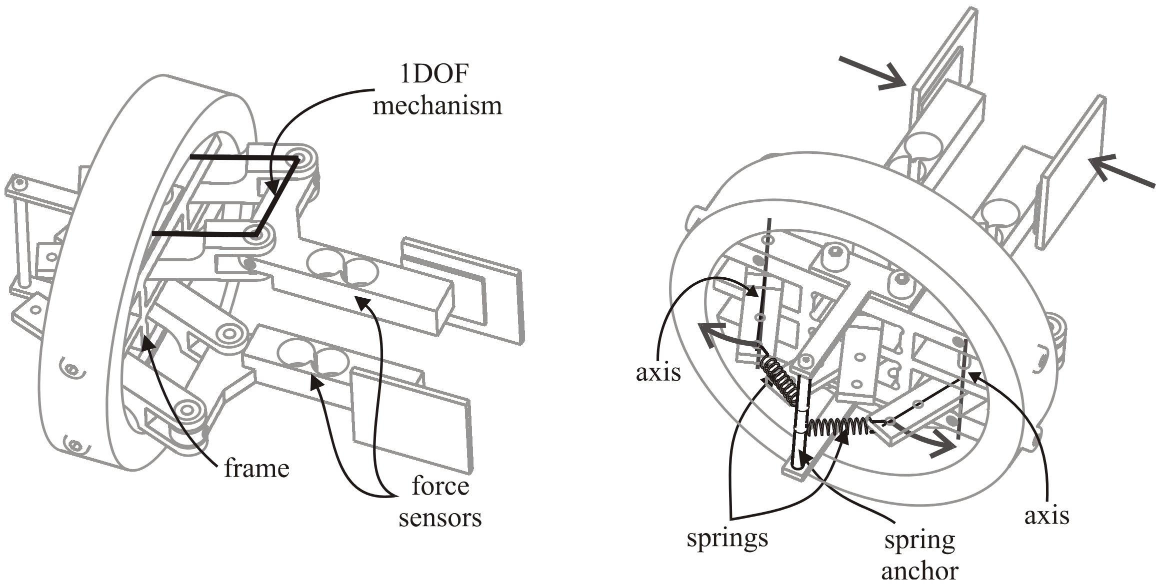


Figure 2.3: The schematics of the HapticMaster’s grasping device without the thumb and finger supports [39]. The user squeezes the pads of the device to grasp a virtual object. The left subfigure shows the basic mechanism while the right subfigure shows the back of the device where springs are attached to the mechanism and used for passive haptic rendering.

2.2 Psychophysiological sensors

Four measurements were chosen for use in the dissertation following a preliminary evaluation: electrocardiography, skin conductance, respiration and peripheral skin temperature. These four have seen extensive use in psychophysiology, as evidenced in numerous review papers [19, 40, 41].

All psychophysiological sensors were manufactured by g.tec (Graz, Austria). Skin temperature, respiration and skin conductance sensors are shown in Figure 2.4. The electrocardiogram (ECG) was recorded using four disposable surface electrodes placed in a configuration suggested by g.tec (one electrode on the left part of the chest, one on the right part of the chest, one on the left part of the abdomen, and a ground electrode on the upper left part of the back). Skin conductance was measured using a g.GSR sensor (g.tec). The electrodes were placed on the medial phalanxes of the second and third fingers of either the idle hand (for healthy subjects) or the nonparetic hand (for hemiparetic patients). The sensor generated a constant voltage between the two electrodes and measured the current between the electrodes in order to estimate skin conductance according to an established procedure [42]. Respiration was measured using a thermistor-based SleepSense Flow sensor placed beneath the nose. Peripheral skin temperature was measured using a g.TEMP sensor (g.tec) attached to the distal phalanx of the fifth finger of either the idle hand (for healthy subjects) or the nonparetic hand (for hemiparetic patients). All of these signals were connected to a g.USBamp signal amplifier (Figure 2.5). The sampling frequency was 2.4 kHz in the first part of the dissertation (Section 3, Analysis of rehabilitation-specific factors). It was, however, decreased to 1.2 kHz in the second part of the dissertation (Section 4, Data fusion and biocooperative control). This was done to save hard drive space and processing power since 2.4 kHz is unnecessarily high. Skin conductance, respiration and skin temperature can all be accurately measured with a sampling frequency below 20 Hz while a recommended ECG sampling frequency that allows accurate analysis of heart rate variability (HRV) is approximately 500 Hz for healthy adults [43] and 500-1000 Hz for subjects with pathological conditions [44]. These recommendations are based on a number of earlier studies, which found that the majority of the ECG's energy is located between 100 Hz and that there is no statistically significant difference between the power spectral density calculated at the recommended frequency and the power spectral density calculated at higher frequencies. Since the signal amplifier requires all signals to have the same sampling frequency and has a limited number of available sampling frequencies, the first available sampling frequency above 1000 Hz was selected (1200 Hz). Though stored at 1200 Hz, the skin conductance, respiration and skin temperature signals were downsampled to 24 Hz before processing.

Figure 2.4: Physiological sensors: temperature (top left), respiration (top right) and skin conductance (bottom). All manufactured by g.tec.

Figure 2.5: The g.USBamp signal amplifier (manufactured by g.tec).

Three other measurements were excluded after early trials: facial electromyography, blood oxygen saturation and finger photoplethysmography. Blood oxygen saturation is not a commonly used measurement in psychophysiology, and a recent review lists no examples of its use [19]. After no changes in blood oxygen saturation were noted during early trials, it was excluded from further use. Finger photoplethysmography is frequently used as a measure of heart rate, but this can be more accurately calculated from the ECG. It is also used together with the ECG to calculate pulse transit time (the time needed for the pulse pressure waveform to propagate through a length of the arterial tree), which has been previously used in psychophysiology [45]. However, early trials found the sensor to be very sensitive to movement, and it was decided to abandon its use, especially since pulse transit time is not nearly as commonly used in psychophysiology as other heart rate measures [19]. Facial electromyography, on the other hand, has been demonstrated to provide accurate information regarding emotional valence in laboratory studies [16, 17, 46]. It was eventually discarded because the electrodes needed for electromyography require precise positioning, are time-consuming to apply, and are considered fairly obtrusive by the subject (since they are placed around the eyes and along the jaw). Thus, facial electromyography was felt to be inappropriate for a clinical rehabilitation setting, though its usefulness in other settings is well-established.

2.3 Signal preprocessing and feature extraction

Feature extraction refers to extracting a number of relevant features from raw physiological, biomechanical or other signals. The ECG, for example, is a raw physiological signal from which a number of features such as mean heart rate or various measures of HRV can be extracted. The process is described for each signal separately in the following subsections.

Psychophysiological features are generally calculated over an interval of fixed length, with lengths from a few seconds to a few minutes being common in the literature (see Kreibig [19] for a review and a table of interval lengths). In this dissertation, several studies were performed with different interval lengths. The shortest intervals (used in section 4) were two minutes long while the longest (used in section 3.2) were five minutes long. Shorter intervals were not used since some features require an interval of at least two minutes to be calculated (e.g. some measures of HRV [43]). Additionally, some signals such as peripheral skin temperature respond fairly slowly to stimuli and cannot be properly evaluated over a short interval.

After feature extraction, a number of different features are obtained for each interval. Throughout this dissertation, a vector of different features from a single interval will be referred to simply as a 'data point'.

2.3.1 Electrocardiogram

Of the four psychophysiological signals included in the dissertation, the ECG is the most complex to process. Fortunately, many processing methods have been developed for it since it is commonly used in nonpsychophysiological applications. The first step is to filter the raw ECG to remove noise. Two filters were applied. The first was the optional 50 Hz notch filter available in the g.USBamp signal amplifier for reduction of power line interference. The second was a fourth-order Butterworth high-pass filter whose cutoff frequency was set at 0.5 Hz. Such filters are frequently used to reduce baseline drift and noise caused by mechanical movement [47].

After filtering, R-peaks need to be detected in the ECG. R-peaks correspond to ventricular depolarization and are the most prominent peak in the ECG. Because of this, they are usually used as the basis for heart rate calculation. Numerous algorithms, mostly based on the amplitude or derivative of the filtered ECG, are available for R-peak detection (e.g. Friesen et al. [48]), but a simple amplitude threshold followed by detection of signal peaks using the first and second derivative proved sufficient as long as the electrodes were properly positioned on the subject's body. In offline analysis, the ECG was additionally checked manually and any incorrectly detected peaks were corrected.

The length of time that passes between two normal (non-pathological) R-peaks is called a NN-interval (Figure 2.6). Heart rate is defined as the reciprocal value of the NN-interval. *Mean heart rate* served as the primary ECG-derived feature. However, several standardized time- and frequency-domain measures of HRV were also calculated [43]. These are not calculated directly from the ECG itself, but from the NN-intervals. The three time-domain features were the *standard deviation of NN-intervals* (*SDNN*), the *square root of the mean squared differences of successive NN-intervals* (*RMSSD*) and the *percentage of interval differences of successive NN-intervals greater than 50 ms* (*pNN50*). They are calculated as follows:

$$SDNN = \text{stdev}(NN) \tag{2.1}$$

$$RMSSD = \sqrt{\frac{\sum_{i=1}^{L_{NN}-1}(NN_{i+1}-NN_i)^2}{L_{NN}}} \tag{2.2}$$

$$pNN50 = \frac{\sum_{i=1}^{L_{NN}-1}((NN_{i+1}-NN_i)>0.05\ s)}{L_{NN}} \tag{2.3}$$

where NN is the time series of NN-intervals, L_{NN} is the number of elements of NN and stdev is the standard deviation function.

For frequency-domain measures, NN-intervals were converted into an instantaneous time series using cubic spline interpolation and the power spectral density of this time series was calculated using Welch's method of modified periodograms [49]. The power spectral density has two frequency bands of interest to us: the low-frequency band (LF) between 0.04 Hz and 0.15 Hz and the high-frequency band (HF) between 0.15 Hz and 0.4 Hz. Three frequency-domain HRV features were calculated: *total power in the LF band*, *total power in the HF band* (commonly referred to as respiratory sinus arrhythmia) and the ratio of the two (commonly referred to as the *LF/HF ratio*). These frequency-domain measures should be calculated over a time period of two to five minutes.

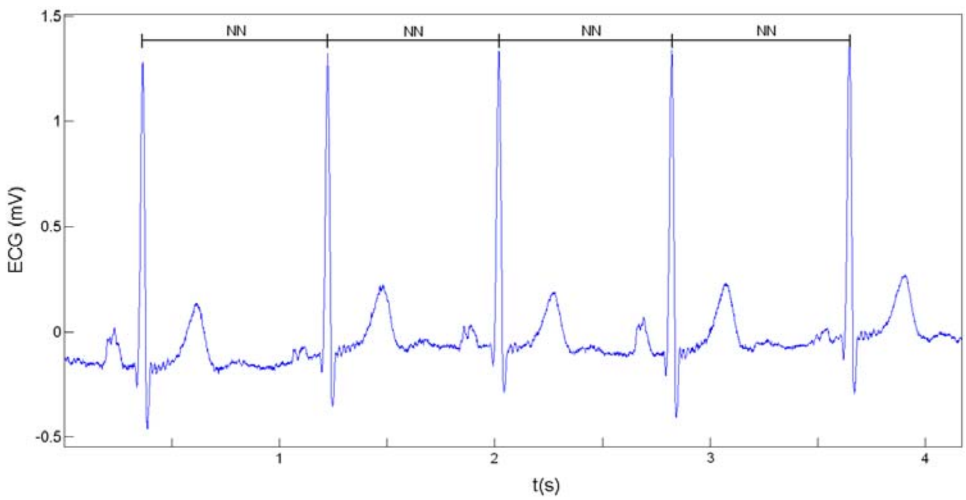


Figure 2.6: An example ECG signal with NN-intervals marked.

In total, seven features were derived from the ECG. Both heart rate and HRV have been previously connected to psychological changes. Heart rate increases and HRV decreases as a result of cognitive workload [12, 13, 14, 50]. Changes in both heart rate and HRV have also been linked to different emotions such as anger [51, 52] fear and sadness (a review for the last two is available in Kreibig et al. [53]), but results are still controversial, with different studies reporting different results. These controversies may be due to different methods of eliciting emotions or due to different types of emotions (e.g. hot or cold anger).

2.3.2 Skin conductance

An example of a skin conductance signal is shown in Figure 2.7. Two components of skin conductance are characterized, tonic and phasic [42, 54]. Tonic skin conductance is the slowly-changing baseline level of skin conductance, in the absence of any particular discrete environmental event, and is generally referred to as skin conductance level (SCL). Each person has a different SCL, which varies over time depending on psychological state and autonomic regulation. Phasic skin conductance consists of rapid skin conductance increases followed by a return to the tonic level. These changes occur in response to discrete environmental stimuli, but can also occur spontaneously in the absence of any specific stimuli. These rapid increases are generally referred to as skin conductance responses (SCRs). They are often also called galvanic skin responses, but this is considered to be an outdated term.

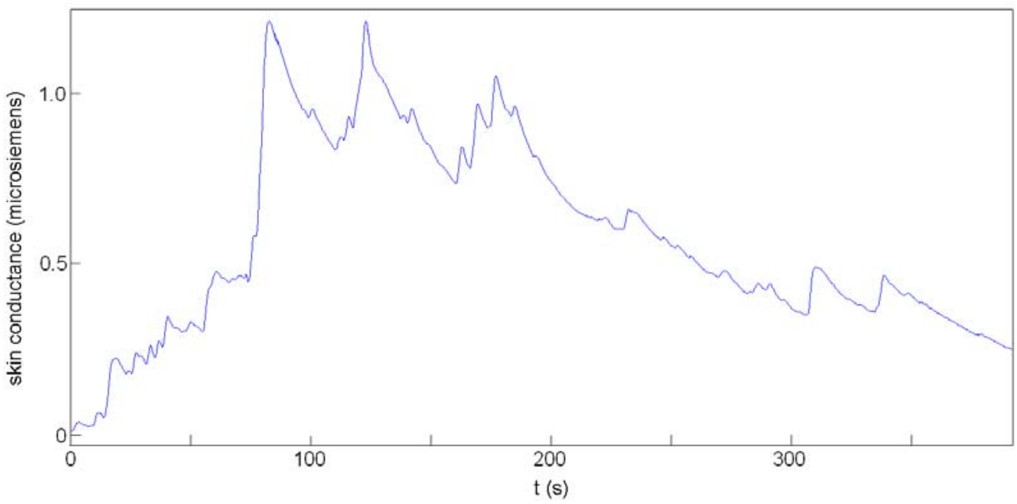


Figure 2.7: An example skin conductance signal with several skin conductance responses.

The skin conductance signal was first filtered with a fourth-order Butterworth low-pass filter whose cutoff frequency was set at 5 Hz. This is a commonly used cut-off frequency for removal of high-frequency noise [55]. Afterwards, two filters were separately applied to this signal to obtain tonic and phasic skin conductance. A fourth-order Butterworth low-pass filter with a cutoff frequency of 0.1 Hz was applied to the signal to obtain tonic skin conductance while a fourth-order Butterworth high-pass filter with a cutoff frequency of 0.1 Hz was applied to obtain phasic skin conductance. *Mean SCL* was calculated from tonic skin conductance. A transient increase in phasic skin conductance was detected as a SCR if its amplitude (from beginning of the increase to the peak) exceeded 0.05 microsiemens and its peak occurred less than 5 seconds after the beginning of the increase. These are commonly used criteria for SCR detection [56], if perhaps somewhat arbitrary. *SCR frequency* and *mean SCR amplitude* were calculated.

In total, three features were derived from the skin conductance signal. It is generally agreed that skin conductance is predominantly innervated by the sympathetic nervous system [31]. It thus increases with general psychological arousal and cognitive workload [12, 41, 57, 58]. Though skin conductance also increases as a result of emotions such as fear [53], it is poor at differentiating between positive or negative emotions.

2.3.3 Respiration

The signal obtained from the SleepSense flow sensor is a sine-like signal in which the troughs represent the beginning of inspiration and the peaks represented the beginning of expiration (Figure 2.8). As there is no standardized procedure for respiratory rate calculation from flow sensors, a procedure was defined following preliminary measurements. The signal was first filtered with a fourth-order Butterworth low-pass filter whose cut-off frequency was set at 5 Hz. Since the adult human respiratory rate is generally between 10-20 breaths per minute (approximately 0.16-0.33 Hz), it was felt that such a cut-off frequency would reduce high-frequency noise without removing useful information. Peaks in the signal were then detected with a simple algorithm based on the signal's first and second derivatives. Respiratory periods were calculated as the times between two peaks in the signal, and *mean respiratory rate* was calculated as the mean reciprocal value of the respiratory period. Furthermore, *respiratory rate variability* was calculated as the standard deviation of the reciprocal value of the respiratory period. Respiratory volume was not calculated since the amplitude of the signal from the sensor does not reflect respiratory volume; only the locations of the peaks and troughs are useful.

During measurements, subjects were generally encouraged to remain silent in order to avoid artefacts in the respiration signal that would be caused by speaking. If the subject spoke, this was noted down by the experimenter and any artefacts around the time of speaking were manually removed from the recorded signal prior to feature extraction.

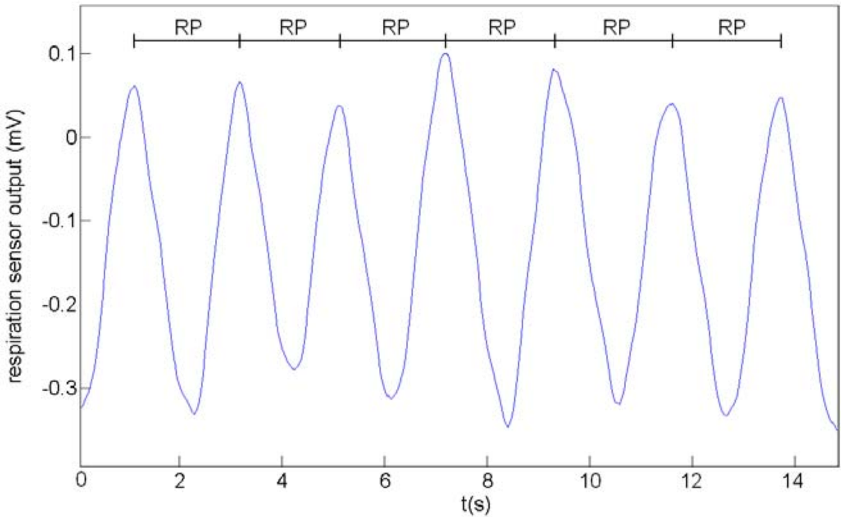


Figure 2.8: An example respiration signal with respiratory periods (RP) marked.

Compared to heart rate and skin conductance measurements, respiration is often overlooked in psychophysiology. Often, it simply serves as a secondary measure used to identify respiration-related artifacts [31]. Nonetheless, it has also been connected to psychological states. In a thorough review, Boiten et al. [59] concluded that respiratory activity is (in their own words) mainly affected by the continua of calm-excited and active versus passive coping. In other words, there is a primary influence of psychological arousal and activation. Boiten [60] also found that respiratory variability in general decreases with increased cognitive workload. Veltman and Gaillard [13] found that respiratory rate increases with cognitive workload and arousal. In a later study, Gomez et al. [61] confirmed a strong influence of arousal, but also noted an influence of emotional valence (positive vs. negative emotions). Furthermore, respiratory variability has been connected to anxiety [62].

2.3.4 Peripheral skin temperature

Peripheral skin temperature changes very slowly compared to the other three physiological signals. Responses to stimuli begin to occur at more than 15 seconds after the stimulus [63]. An observation of the signals recorded during the coures of the dissertation found that changes in skin temperature could begin to occur up to a minute after constant exposure to a stimulus and need an additional minute or two to reach the maximum deviation from the initial value. Because of this slowness, high-frequency noise was removed with a Butterworth fourth-order low-pass filter whose cut-off frequency was set at 1 Hz. The *final skin temperature* at the end of each period was calculated as the mean value over the last 5 seconds of the period. The final value rather than the mean value was chosen because of the slow, delayed response.

Changes in peripheral skin temperature are caused by changes in the microcirculation induced by stressors or other stimuli acting on the sympathetic nervous system [63]. Thus, like skin conductance (which is also innervated by the sympathetic nervous system), skin temperature changes are likely to primarily reflect general psychological arousal and cognitive workload [64], though skin temperature changes have also been noted in response to emotions such as anxiety [65], fear and sadness [53].

2.3.5 Biomechanics

Biomechanical features describe the movements and forces applied by the human onto the HapticMaster's end-effector. They were derived from the position of the robot's end-effector (which was calculated from the HapticMaster's internal joint position sensors through direct kinematics) and from the force signal measured by the 3-axis force sensor in the end-effector. The extracted features were defined so that they would both account for mean physical activity (e.g. mean absolute force, total work) as well as detect an increased amount of sudden jerky movements. Preliminary experiments with haptic tasks showed that increased task difficulty was characterized by jerkier movements. A total of eight features were extracted: *mean absolute force*, *mean absolute velocity*, *mean absolute acceleration*, *total work*, *mean frequency of the position signal*, *mean frequency of the velocity signal*, *mean frequency of the acceleration signal*, *and mean frequency of the force signal*. The mean frequencies were calculated through Welch's method of modified periodograms [49]. Total work was calculated as:

$$W_{total} = \int_C F_{int}\, dx \tag{2.4}$$

where F_{int} is the interaction force between the human and the robot's end-effector and C is the path along which the robot was moved. It should be acknowledged that this total work includes both the active physical effort exerted by the subject and the passive energy of the robot moving the subject's arm.

All biomechanical features were calculated only for movement in the horizontal plane. This was done due to the nature of the two tasks used in the dissertation: the inverted pendulum task and the ball-catching task. The inverted pendulum task (section 3.2) only includes horizontal movements. The ball-catching task (sections 3.3 and 4.3) does include vertical movement, but task difficulty is modulated by only adjusting the horizontal component of the task, so vertical movements were considered irrelevant.

Though the primary focus of the dissertation is on psychophysiological measurements, biomechanical measurements also have several uses. First, they can be used to evaluate the level of physical activity during the task. Second, since they are already available in the HapticMaster, it would be interesting to see if they could, by themselves, provide enough information about how suitable the task is for the patient. In such a case, psychophysiological measurements would be unnecessary.

2.3.6 Task performance

Task performance features describe how well a subject did at a particular task. As such, they are necessarily task-specific. Since several different tasks were used in the dissertation, task performance features for each task will be introduced together with that particular task in later sections.

Like biomechanical measurements, task performance measurements are already available in motor rehabilitation and might by themselves provide enough information about how suitable the task is for the patient. In data fusion (section 4), they were thus used as an alternative or complementary data source to psychophysiological measurements. In statistical analysis (section 3), they were used as validation: to determine whether differences actually exist between different conditions and to calculate correlations between psychophysiology and task performance.

2.4 Feature normalization

Psychophysiological features exhibit high intra- and intersubject variability as a result of age, gender, time of day and other factors. Normalization is primarily an attempt to reduce the effect of this variability prior to data fusion. For instance, if a training data set contains measurements from several subjects, some subjects may exhibit much larger responses than others or have different resting values for psychophysiological features (resting heart rate, for instance, can easily be anywhere between 60 and 80 beats per minute). This needs to be taken into account prior to data fusion. Furthermore, since different features are measured in different units, some features have much larger numerical ranges than others, which can be problematic for some data fusion methods (such as nearest-neighbor classification, described in section 4.1.2.1.1). Normalization also attempts to reduce this effect. Three normalization approaches are commonly used.

The first approach is to record psychophysiological responses in a neutral or 'baseline' conditions where the subject is not exposed to stimuli or is only exposed to basic, relaxing stimuli. Psychophysiological features from other conditions (where the subject is performing a task or exposed to affective stimuli) are then normalized by either subtracting the baseline value [66-68], dividing by the baseline value [69, 70], subtracting the baseline value and dividing the result by the baseline value [71-73], or a combination of these, with different options used for different features [74]. Subtraction of the baseline value is obviously aimed at reducing intersubject variability due to different baseline values while division is also partially aimed at reducing variability due to different response sizes. This approach can easily be used online, though it does require a baseline condition to be recorded first.

The second approach also begins by recording psychophysiological responses in a baseline condition. However, instead of subtracting or dividing the data from the 'task' or 'affective' conditions with the baseline data, the baseline features are added to the feature space as independent features - thus doubling the dimension of the feature space. This approach is called the 'baseline matrix' and has been previously used in several studies [46, 75]. Like the previous approach, this can be easily done online, though it requires a baseline condition to be recorded first.

The third approach includes no baseline recordings, but simply involves normalizing the data from each subject separately or across all subjects to a certain range (e.g. from 0 to 1 or from -1 to 1) [16, 76-78]. This is done for each feature separately by, for instance, subtracting the mean value of all data points and dividing the result by the standard deviation of all data points. If done for each subject separately, the goal is generally to reduce intersubject variability by scaling each person's features to a difference between their maximum and minimum value. If done across all subjects, the goal is simply to ensure that each psychophysiological feature has the same numerical range. In online data fusion, normalizing psychophysiological features without a baseline recording can be done by calculating the maximum and minimum value of each feature across the entire training data set, then scaling features online between that maximum and minimum value.

As there is no clear 'best' normalization approach in the literature, the first approach was chosen since it is the most common and since recording a baseline condition is not problematic. However, it was uncertain whether to simply subtract the baseline value or to also divide by the baseline value. Thus, for the first study in this dissertation (analysis of the effects of physical activity, section 3.2), most features were normalized by first subtracting the baseline value and then dividing the result by the baseline value. There was one exception: *mean SCL*, which is already measured as the difference from an initial value and was thus normalized by simply subtracting the baseline value. For later studies, however, it was felt that a mixed approach would be best. Thus, in all other studies, the "subtract and divide" method was used for *SDNN*, *RMSSD*, *LF/HF index*, *total LF power*, *total HF power*, *SCR frequency* and *respiratory rate variability*. *Mean heart rate*, *pNN50*, *mean SCL*, *mean SCR amplitude*, *mean respiratory rate* and *final skin temperature* were normalized by simply subtracting the baseline value. The choice to change normalization methods was based on a purely subjective opinion that such an approach would be more effective.

2.5 Questionnaires

A number of different questionnaires have been used together with psychophysiological measurements, among them the NASA-TLX [79], the Behavior Activation System / Behavior Inhibition System (BAS/BIS) scales [80], the Self-Assessment Manikin (SAM) [81] and a variety of study-specific questionnaires mainly consisting of multiple-choice questions [17, 53, 66, 72]. Though multiple studies have failed to find strong correlations between self-report questionnaires and psychophysiological features [82, 83], questionnaires remain the most popular and convenient method of validating psychophysiological measurements.

Two questionnaires were used in the dissertation: the SAM, which measures the current valence and arousal of the subject, and the BAS/BIS scales, which measure properties of the individual's innate motivational systems. The SAM was used for both studies described in section 3 as well as data fusion in a non-rehabilitation setting (section 4.2). The BAS/BIS scales were used for analysis of the effects of stroke (section 3.3) and data fusion in a non-rehabilitation setting (section 4.2). Although the BAS/BIS scales were also intended for use with data fusion in rehabilitation (section 4.3), they had to be omitted due to lack of time.

2.5.1 The Self-Assessment Manikin

The nine-point *arousal* and *valence* scales from the SAM [81] were chosen as the primary questionnaire. Shown in Figure 2.9, these scales allow subjects to rate their level of emotional valence and arousal graphically by choosing the picture that best represents their current mood. *Valence* (sometimes also called pleasure) is defined as positive versus negative affective states (e.g., humiliation, disinterest, and anger at one end versus excitement, relaxation, and tranquility at the other end) while *arousal* is defined in terms of mental alertness and physical activity (e.g., sleep, inactivity, boredom, and relaxation at the lower end versus wakefulness, tension, exercise, and concentration at the higher end) [38].

Valence and *arousal* were converted to numerical values for purposes of analysis. For *valence*, 1 represented extremely negative *valence* while 9 represented extremely positive *valence*. For *arousal*, 1 represented extremely low *arousal* while 9 represented extremely high *arousal*. The SAM was chosen over other questionnaires for two reasons. First, the physiological effects of arousal and valence are well-documented [41]. Second, the SAM is graphical in nature and thus very simple to use; in pretesting, some stroke patients had difficulty comprehending more complex self-report questionnaires. Despite this simplicity, the SAM has been shown to yield results similar to those of more complex self-report scales such as the semantic differential [81]. The SAM also contains a third subscale, dominance, but most psychophysiological studies omit it since it has never been reliably connected with physiological responses.

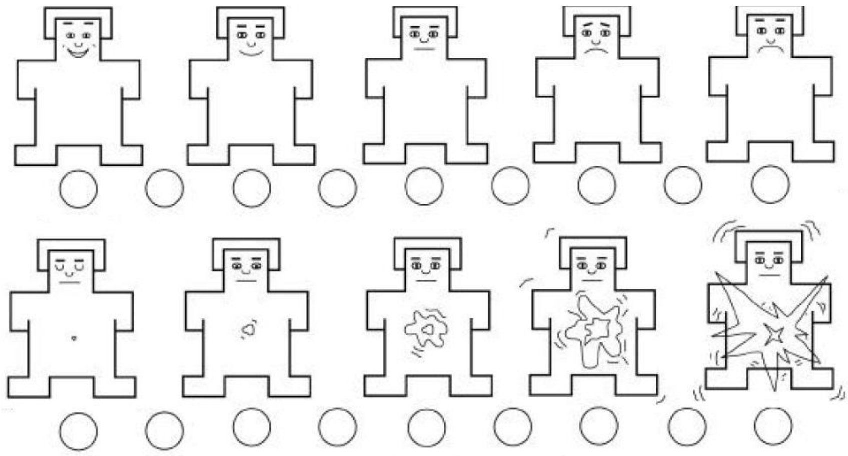


Figure 2.9: The valence (top) and arousal (bottom) scales of the Self-Assessment Manikin.

2.5.2 Behavioral Activation System / Behavioral Inhibition System Scales

The Behavioral Activation System / Behavioral Inhibition System (BAS/BIS) scales [80] describe two innate motivational systems governing appetitive and aversive behaviors. The behavioral activation system (BAS) is involved in simple reward-approach situations as well as in the initiation of behavior in active avoidance situations where the subject must respond to avoid punishment. The behavioral inhibition system (BIS) is viewed as an anxiety system and inhibits behavior in the presence of cues signaling that frustrative or anxiety-evoking aversive consequences would occur as a result of that behavior.

Passive avoidance is one situation that activates the BIS. It has been demonstrated that the BAS primarily influences heart rate while the BIS primarily influences skin conductance [84]. The BAS/BIS scales have already been used in virtual reality [85], making them a potentially useful addition to the dissertation.

The BAS/BIS scales themselves consist of 24 statements (such as “I often act on the spur of the moment” or “I feel worried when I think I have done poorly at something important”) with four possible choices for each (very true for me / somewhat true for me / somewhat false for me / very false for me). Four statements are filler and do not contribute to the result while the others contribute to four subscales: one *BIS* scale and three BAS scales (*BAS Drive*, *BAS Fun Seeking*, *BAS Reward Responsiveness*). According to Carver and White [80], the fact that there are three BAS-related scales and only one BIS-related scale was not planned or theoretically motivated. The factors emerged empirically, from an item set that was intended to capture diverse manifestations of the BAS, according to various theoretical statements.

Since the BAS/BIS scales have been linked to motivation (a very important factor in biocooperative rehabilitation) as well as to psychophysiological responses, they were included as a way of evaluating the effect that a person's innate psychological properties (as opposed to his/her current mood, evaluated by the SAM) affect his/her experience and psychophysiological responses.

3 Analysis of rehabilitation-specific factors

3.1 Introduction

As previously mentioned, little concrete psychophysiological research has been carried out in motor rehabilitation due to two rehabilitation-specific factors. This section is concerned with a statistical analysis of these two factors so that their influence can be better-understood in data fusion.

The **first** factor, studied in Section 3.2, is the presence of strenuous physical activity. Psychophysiological responses are affected not only by psychological stimuli, but also physical activity. Most studies consider this physical activity to be an undesired factor and attempt to limit it to a minimum, but this cannot be done in motor rehabilitation where physical activity is the integral component of the process.

A number of studies have examined psychophysiological responses to a combination of physical and cognitive workload, but have mainly focused on the effects of a mentally demanding task superimposed onto a physically demanding task (e. g. performing mental arithmetic while riding a bicycle) [21, 22]. Subjects in these studies were thus performing several unrelated tasks at once. During interaction with haptic robots, however, a single task frequently contains elements of both physical and cognitive workload. The interplay between cognitive and physical workload found in haptic human-robot interaction may result in different psychophysiological responses. While psychophysiological measurements have been applied to human-robot interaction [18, 86], they have never been studied specifically in the context of haptic interaction.

The question to be answered was thus simple: in haptic human-robot interaction, is it possible to use psychophysiological responses to differentiate between different levels of cognitive workload at different levels of physical workload? Our main hypothesis and its sub-hypotheses were:

H1: In in haptic human-robot interaction, physiological responses are affected by both cognitive and physical workload.

- H1.1: Cognitive workload causes heart rate, skin conductance and respiratory rate to increase. Furthermore, it causes skin temperature to decrease.
- H1.2: Physical workload also causes heart rate, skin conductance and respiratory rate to increase. Effects on skin temperature are uncertain.
- H1.3: The physiological effects of physical and cognitive workload are additive; the presence of both workload types causes a larger physiological response than the presence of a single type.
- H1.4: Both types contribute significantly to physiological responses; at the same level of cognitive workload, changing physical workload should significantly change physiological responses and vice-versa.

The **second** factor, studied in section 3.3, is the damage to the autonomic nervous system that is present in most patients undergoing motor rehabilitation. Stroke patients, for instance, are known to show long-lasting abnormalities in sweating and HRV [23], though some recovery occurs with time [87]. Similarly, traumatic brain injury also results in weakened psychophysiological responses [24]. Electrical nerve stimulation evokes significantly smaller SCRs in stroke patients than controls [88-89] and fails to evoke any SCRs at all on the limbs of some patients with spinal cord lesions [90].

Psychophysiological responses have not, however, yet been studied during motor rehabilitation itself. Prior to data fusion, an analysis should be performed with both patients and healthy controls to identify all weakened or absent psychophysiological responses during rehabilitation tasks. This should ideally be done with both an actual rehabilitation task as well as additional psychological tasks where the effects of physical activity are not present. Due to a limited availability of patients at the University Rehabilitation Institute of the Republic of Slovenia, it was decided to focus primarily on stroke patients. While there are different types of stroke, a group of only stroke patients should nonetheless exhibit less intersubject variability than, for instance, a group of both stroke and spinal cord injury patients. Our main hypothesis and its sub-hypotheses were:

H2: Stroke patients have weakened or even absent psychophysiological responses compared to control subjects.

- H2.1: Changes in skin conductance and heart rate variability in response to stimuli should be smaller in stroke patients than controls based on previous research. Effects of stroke on respiration and skin temperature are uncertain, but it is expected that these responses are also weakened.
- H2.2: When performing a purely cognitive task without any physical workload, stroke patients nonetheless exhibit one or more of the following: significant increases in heart rate, skin conductance or respiratory rate or a significant decrease in skin temperature.
- H2.3: When performing a motor rehabilitation task, stroke patients nonetheless exhibit one or more of the following: significant increases in heart rate, skin conductance or respiratory rate.

3.2 The effects of physical activity

3.2.1 Task

Subjects were presented with a virtual version of the classic inverted pendulum problem (visible on the screen in Figure 3.1). A thin pole with a weight at its top end is attached at its bottom to a moving cart. This vertical pendulum is inherently unstable; left alone, the pole will fall to the ground. However, if the cart is moved left or right, it will act upon the pole and either accelerate its fall or balance it. This system is referred to as the inverted pendulum and is a classic problem in control theory. Subjects were presented with a simulated cart and pole on a screen. They moved the cart left and right using the HapticMaster, with the goal of keeping the pole from falling. The cart moved in the same direction and with the same velocity as the end-effector of the HapticMaster. If the subjects failed to balance the pole and it fell to a horizontal position, it was immediately reset to a nearly vertical position. Force feedback was also implemented with the

HapticMaster, allowing the subjects to feel the reaction forces resulting from the movement of the cart.

The nonlinear differential equations describing the inverted pendulum system can be derived using standard laws of motion and are:

$$(M + m)\ddot{x} + b\dot{x} + ml\ddot{\theta}\cos(\theta) - ml\dot{\theta}^2\sin(\theta) = kF \tag{3.1}$$

$$(l + ml^2)\ddot{\theta} + mgl\sin(\theta) = -ml\ddot{x}\cos(\theta) \tag{3.2}$$

where M is the mass of the cart, m is the mass of the pole (which is concentrated at the tip of the pole), l is the length of the pole, g is the gravitational acceleration, b is the friction between the cart and the ground, F is the force exerted by the subjects, k is a factor that scales between the force exerted by the subjects and the force acting on the cart, x is the position of the cart, and θ is the angle between the pole and a vertical line.

Different levels of cognitive workload were achieved in the task using three different task difficulty levels: underchallenging, challenging and overchallenging. These levels of cognitive workload allowed us to test hypothesis H1.1. In the challenging version, the constants in 3.1 and 3.2 were set in such a way as to make balancing the pendulum moderately challenging. The initial value of θ was 5° while initial values of both $\dot{\theta}$ and $\ddot{\theta}$ were zero. In the overchallenging version, a half-second delay was introduced between the time the cart was moved and the time the cart's movement affected the pole, making it more unpredictable. Additionally, the pole was heavier (m was multiplied by 1.5 from the challenging version), the pole was longer (l was multiplied by 1.5) and the friction between the cart and the ground was smaller (b was multiplied by 0.75). This made the task extremely difficult to perform successfully. In the underchallenging version, the pendulum never fell: $\ddot{\theta}$ was kept at zero throughout the time period and the initial value of both θ and $\dot{\theta}$ was zero. The subject was simply asked to move the cart left and right at a moderate speed.

All three difficulty levels were implemented in low physical workload and high physical workload versions. These levels of physical workload allowed us to test hypothesis H1.2. The versions were identical except for one factor: in the high physical workload versions, more physical force was required to move the HapticMaster. The scaling factor *k* in equation 3.2 was divided by five, forcing the subject to apply five times the force that had been applied in the low physical workload versions. This gave us a total of six task conditions: underchallenging with low physical workload, challenging with low physical workload, overchallenging with low physical workload, underchallenging with high physical workload, challenging with high physical workload, and overchallenging with high physical workload. Since the physical and cognitive difficulty can be adjusted independently of each other, hypotheses H1.3 and H1.4 can thus be tested.

A single task performance feature was measured: the *number of times that the pendulum fell* (and was reset). It was measured only in challenging and overchallenging conditions since the pendulum never fell in the underchallenging conditions.

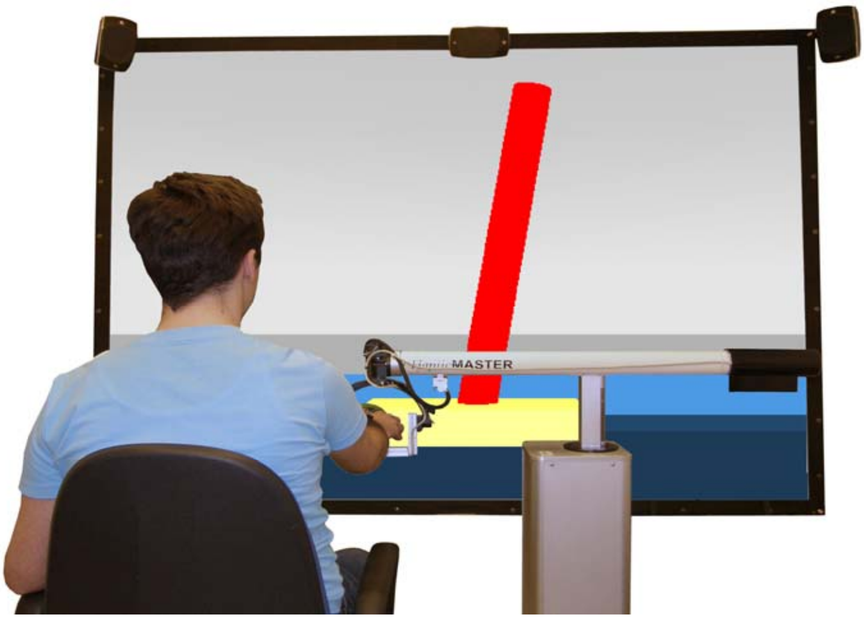


Figure 3.1: A subject performing the inverted pendulum task with the HapticMaster.

3.2.2 Measurement protocol

The experiment was conducted in a quiet area of the laboratory where external stimuli did not disturb the subjects. The temperature and humidity in the laboratory were kept constant. There was never more than one subject and one experiment supervisor inside the laboratory at any time. Each subject performed the experiment in two separate time blocks. Each block consisted of an initial rest period (which served as the baseline) followed by the three different difficulty levels performed in random order. Each condition lasted for five minutes. After each condition, the subject was presented with a self-report questionnaire administered by the experiment supervisor and then the next condition began immediately.

One time block was performed with low physical workload while the other was performed with high physical workload. The order in which the two blocks were presented as well as the order of difficulty levels within each block was randomly chosen before each subject's arrival in the laboratory.

Upon arrival, the task and the experiment procedure were explained to the subject. Then, the challenging difficulty level was presented for the subject to practice using the HapticMaster at the level of physical workload that would be present during the first block. Everyone was required to practice for at least five minutes, and more time was given to anyone who felt that he or she had not yet reached a basic level of proficiency. This practice period (as well as practice periods in all other studies described in the dissertation) was presented in order to reduce the effect of novelty: psychophysiological responses are generally strongest during the first exposure to a new stimulus, then decrease as a result of habituation [91]. For this reason, psychophysiological studies frequently perform a practice session before the actual experiment in order to reduce the effects of novelty during the experiment session [92].

After practice had been completed, the physiological sensors were attached and turned on. Then, the first block of the experiment was performed. After the first block had been completed, a brief informal interview was conducted with the subject. He or she was allowed to rest briefly if desired. Then, he or she was required to practice the task at the level of physical workload that would be present during the second block for at least five minutes. After the practice, the second block of the experiment was performed. After the second block had been completed, the subject was disconnected from the equipment and an informal interview was conducted about the entire experience.

3.2.3 Participants

Thirty students and staff members from various departments of the University of Ljubljana (age range: 19-46 years, mean 26.2, standard deviation 5.8) participated in the study. Twenty-three were male, seven were female. All were healthy, without any major cognitive or physical defects. Each subject signed an informed consent form.

3.2.4 Statistical methods

Results were analyzed with a two-way repeated-measures ANOVA in order to evaluate significance and effect size. One factor was physical workload (two levels: low/high) while the other was cognitive workload. Cognitive load had two levels for task performance (challenging / overchallenging, since the pendulum did not fall during the underchallenging condition), three for psychophysiology and biomechanics (underchallenging / challenging / overchallenging), and four for the SAM (baseline / underchallenging / challenging / overchallenging). Psychophysiological features were normalized while others were not. Effect size was calculated as partial η^2, the proportion of total variability attributable to the factor, excluding other factors from the total nonerror variation [93]. All hypotheses were tested at a 5% significance level. The Sidak correction for multiple comparisons [94] was used for all post-hoc tests. The Huynh-Feldt correction [95] was used in cases of violations of sphericity in ANOVA. The Kolmogorov-Smirnov test with Lilliefors' modification [96] was used to test for normality. If the requirements for regular ANOVA were not met, ANOVA on ranks was used instead.

3.2.5 Results

3.2.5.1 Performance

For low physical workload, the pendulum was reset 3.2 ± 1.3 times per minute during the challenging condition (mean ± standard deviation across all 30 subjects) and 5.6 ± 1.0 times per minute in the overchallenging condition. For high physical workload, the pendulum was reset 2.8 ± 0.9 times per minute in the challenging condition and 5.4 ± 1.2

times per minute in the overchallenging condition. The pendulum did not fall during the two underchallenging conditions. There was a significant main effect of cognitive workload ($p < 0.001$, partial $\eta^2 = 0.87$) as well as a smaller significant main effect of physical workload ($p = 0.042$, partial $\eta^2 = 0.18$). There was no significant interaction effect ($p = 0.79$, partial $\eta^2 = 0.00$).

3.2.5.2 Self-assessment manikin

Table 3.1 shows results from the SAM for baseline and task conditions.

Table 3.1: Results of self-report measures, presented as mean ± standard deviation. High values represent positive valence or high arousal.

		baseline	underchallenging	challenging	overchallenging
low physical workload	valence	5.5 ± 1.1	4.9 ± 1.5	5.4 ± 1.2	4.1 ± 1.6
	arousal	1.3 ± 1.3	1.9 ± 1.8	4.5 ± 1.9	4.3 ± 1.9
high physical workload	valence	5.4 ± 1.5	4.5 ± 1.9	5.5 ± 1.3	4.5 ± 1.8
	arousal	1.3 ± 1.7	2.2 ± 1.7	4.7 ± 1.7	4.5 ± 1.8

There was a significant main effect of cognitive workload on *valence* ($p = 0.002$, partial $\eta^2 = 0.49$). Post-hoc tests found a significant difference between baseline and overchallenging conditions ($p = 0.019$) as well as between challenging and overchallenging conditions ($p = 0.001$). There was also a significant main effect of cognitive workload on *arousal* ($p < 0.001$, partial $\eta^2 = 0.81$). Post-hoc tests found a significant difference between the baseline condition and all other three conditions ($p = 0.029$ for baseline-underchallenging, $p < 0.001$ for the other two) as well as a significant difference between the underchallenging condition and the other two task conditions ($p < 0.001$ in both cases). There were no significant main effects of physical workload (*valence*: $p = 0.75$, partial $\eta^2 = 0.005$; *arousal*: $p = 0.45$, partial $\eta^2 = 0.03$) and no significant interaction effects (*valence*: $p = 0.93$, partial $\eta^2 = 0.02$; *arousal*: $p = 0.65$, partial $\eta^2 = 0.07$).

3.2.5.3 Biomechanical measurements

Table 3.2 shows values of all biomechanical features in all task conditions. There was a significant main effect of physical workload on:

- *mean absolute force* ($p = 0.006$, partial $\eta^2 = 0.51$),
- *mean absolute velocity* ($p = 0.024$, partial $\eta^2 = 0.38$),
- *mean absolute acceleration* ($p = 0.029$, partial $\eta^2 = 0.36$),
- total work ($p = 0.019$, partial $\eta^2 = 0.41$),
- *mean frequency of the position signal* ($p = 0.033$, partial $\eta^2 = 0.35$),
- *mean frequency of the acceleration signal* ($p = 0.013$, partial $\eta^2 = 0.44$),
- *mean frequency of the force signal* ($p = 0.003$, partial $\eta^2 = 0.57$).

Table 3.2: Mean values of biomechanical features in all task conditions. Underch. = underchallenging, ch. = challenging, overch. = overchallenging.

	Low physical workload			High physical workload		
	Underch.	Ch.	Overch.	Underch.	Ch.	Overch.
mean absolute force (N)	3.06	1.04	1.38	17.1	6.80	7.36
mean absolute velocity (m/s)	0.148	0.055	0.059	0.128	0.049	0.054
mean absolute acceleration (m/s^2)	0.44	0.17	0.23	0.30	0.13	0.15
total work (J)	127.4	16.9	24.1	570.4	95.1	119.2
mean f. of position (Hz)	0.251	0.103	0.103	0.200	0.090	0.093
mean f. of velocity (Hz)	1.57	4.48	3.35	2.11	6.24	4.56
mean f. of acceleration (Hz)	4.21	2.76	2.85	4.40	2.74	3.33
mean f. of force (Hz)	6.85	7.94	15.4	0.67	0.59	1.73

There was a significant main effect of cognitive workload ($p = 0.001$, partial $\eta^2 = 0.76$) as well as a significant interaction effect between physical and cognitive workload ($p = 0.041$, partial $\eta^2 = 0.47$) on *mean absolute force*. Post-hoc tests found significant differences between the underchallenging and overchallenging conditions for both levels of physical workload ($p = 0.033$ for low physical workload, $p < 0.001$ for high physical workload). The difference between underchallenging and challenging conditions was significant only for high physical workload ($p < 0.001$), though the difference also approached significance in the case of low physical workload ($p = 0.069$).

There was a significant main effect of cognitive workload ($p < 0.001$, partial $\eta^2 = 0.79$) as well as a significant interaction effect between physical and cognitive workload (p =

0.016, partial $\eta^2 = 0.34$) on *mean absolute velocity*. Post-hoc tests found significant differences between the underchallenging condition and the other two condition for both levels of physical workload ($p < 0.001$ in all four cases), but no significant difference between the challenging and overchallenging conditions.

Significant main effects of cognitive workload without accompanying significant interaction effects were found for:

- *mean absolute acceleration* ($p = 0.001$, partial $\eta^2 = 0.59$),
- *total work* ($p = 0.001$, partial $\eta^2 = 0.75$),
- *mean frequency of the position signal* ($p < 0.001$, partial $\eta^2 = 0.66$),
- *mean frequency of the velocity signal* ($p = 0.003$, partial $\eta^2 = 0.50$),
- *mean frequency of the acceleration signal* ($p = 0.001$, partial $\eta^2 = 0.77$),
- *mean frequency of the force signal* ($p = 0.013$, partial $\eta^2 = 0.35$).

For the *mean frequency of the force signal*, post-hoc tests found a significant difference between the challenging and overchallenging conditions ($p = 0.024$). For the other features, post-hoc tests found a significant difference between the underchallenging condition and the other two conditions ($p < 0.05$ in all cases), but no significant difference between the challenging and underchallenging conditions.

3.2.5.4 Psychophysiological measurements

Table 3.3 shows normalized values of all psychophysiological features in all task conditions. Furthermore, Table 3.4 shows p-values and partial η^2 for main effect of physical workload, main effect of cognitive workload, and interaction effect between physical and cognitive workload. To better illustrate differences between conditions, normalized values of four physiological features are shown as graphs: *mean heart rate* (Fig. 3.2), *SCR frequency* (Fig. 3.3), *respiratory rate variability* (Fig. 3.4) and *final skin temperature* (Fig. 3.5).

Table 3.3: Mean values of normalized psychophysiological features in all task conditions. Statistically significant differences from baseline value are indicated with bolded values and asterisks: * for p < 0.05, ** for p < 0.01 and *** for p < 0.001. Underch. = underchallenging, ch. = challenging, overch. = overchallenging.

	Low physical workload			High physical workload		
	Underch.	Ch.	Overch.	Underch.	Ch.	Overch.
mean heart rate (%)	-1.6	-1.6	**-2.4***	**10.2*****	**3.9****	**4.7*****
SDNN (%)	-3.2	**-12.3****	**-7.4***	**-17.1*****	**-12.3****	**-8.8****
RMSSD (%)	-2.7	2.3	4.0	**-23.9*****	-10.2	-4.1
pNN50 (%)	12.2	29.7	54.1	**-26.6****	27.1	57.1
LF/HF ratio (%)	17.0	14.3	13.6	**42.9***	21.0	29.9
total HF power (%)	**-17.2****	-6.4	-6.1	**-22.2****	-8.5	1.9
total LF power (%)	**-8.7***	-7.4	3.8	3.1	3.7	16.0
mean SCL (μS)	0.0	0.02	0.4	**1.0*****	**0.7****	**1.0*****
SCR frequency (%)	49.2	**133.7***	**175.8****	**236.2*****	**340.7*****	**368.9*****
mean SCR amplitude (%)	**40.0***	23.0	22.8	21.0	20.8	28.1
mean respiratory rate (%)	**13.7*****	**15.8*****	**15.2*****	**17.4*****	**24.0*****	**22.5*****
respiratory rate variability (%)	**-14.3****	**-28.6*****	**-11.1****	**-6.0***	**-14.4****	4.6
final skin temperature (%)	-0.3	-0.1	**-1.3***	-0.7	-0.1	**-1.5***

Table 3.4: p-values and partial η^2 for main effect of physical workload, main effect of cognitive workload, and interaction effect between physical and cognitive workload.

	physical workload		cognitive workload		interaction	
	p	partial η^2	p	partial η^2	p	partial η^2
mean heart rate	**<0.001**	0.64	**0.008**	0.25	**<0.001**	0.38
SDNN	0.15	0.12	0.60	0.03	0.47	0.18
RMSSD	0.08	0.17	**0.031**	0.21	**0.024**	0.20
pNN50	0.22	0.10	**0.013**	0.25	0.80	0.02
LF/HF ratio	0.53	0.03	0.64	0.03	0.34	0.06
total HF power	0.61	0.02	0.064	0.18	0.94	0.00
total LF power	0.53	0.03	**0.043**	0.34	0.44	0.10
mean SCL	**0.008**	0.34	0.38	0.06	0.24	0.08
SCR frequency	0.067	0.18	**0.048**	0.20	0.66	0.05
mean SCR amplitude	0.94	0.00	0.060	0.02	0.39	0.04
mean respiratory rate	0.82	0.00	0.29	0.05	0.38	0.04
respiratory rate variability	0.45	0.03	**0.004**	0.43	0.88	0.01
final skin temperature	0.45	0.04	**0.048**	0.35	0.61	0.02

Post-hoc tests for cognitive workload and interaction effects found the following differences:

- *mean heart rate*: significant difference between the underchallenging condition and the other two conditions in the case of high physical workload only ($p < 0.001$ in both cases);
- *RMSSD*: significant difference between underchallenging and challenging conditions in the case of high physical workload only ($p < 0.001$);
- *pNN50*: significant difference between underchallenging and overchallenging conditions ($p = 0.047$);
- *total LF power*: significant difference between challenging and overchallenging conditions ($p = 0.040$);
- *SCR frequency:* significant difference between the underchallenging condition and the other two conditions ($p < 0.05$ in both cases);
- *respiratory rate variability*: significant difference between challenging and overchallenging conditions ($p = 0.002$);
- *final skin temperature*: significant difference between the overchallenging condition and the other two conditions ($p < 0.05$ in both cases).

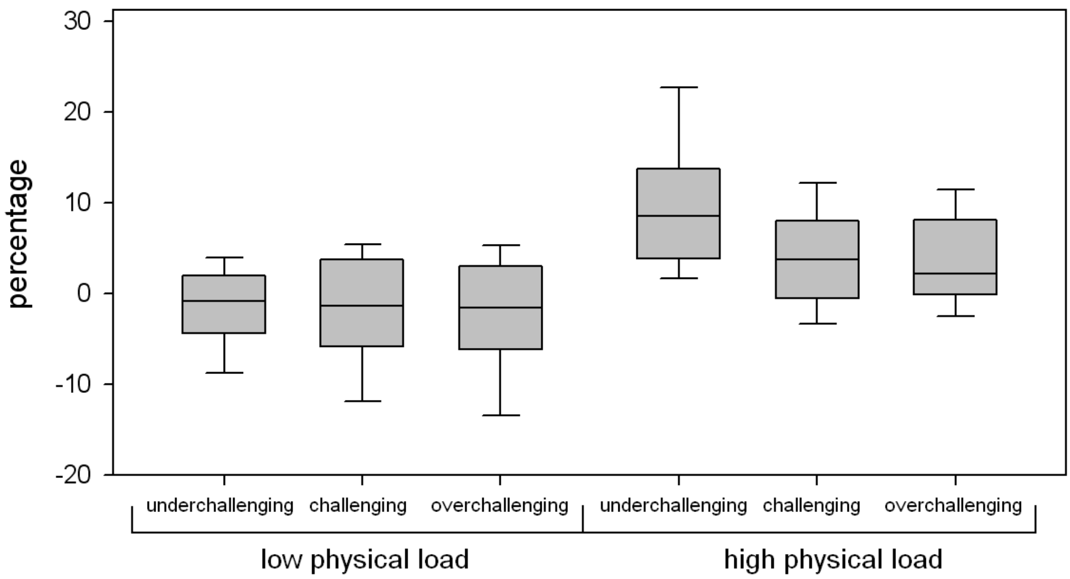


Figure 3.2: Normalized values of *mean heart rate* in different conditions.

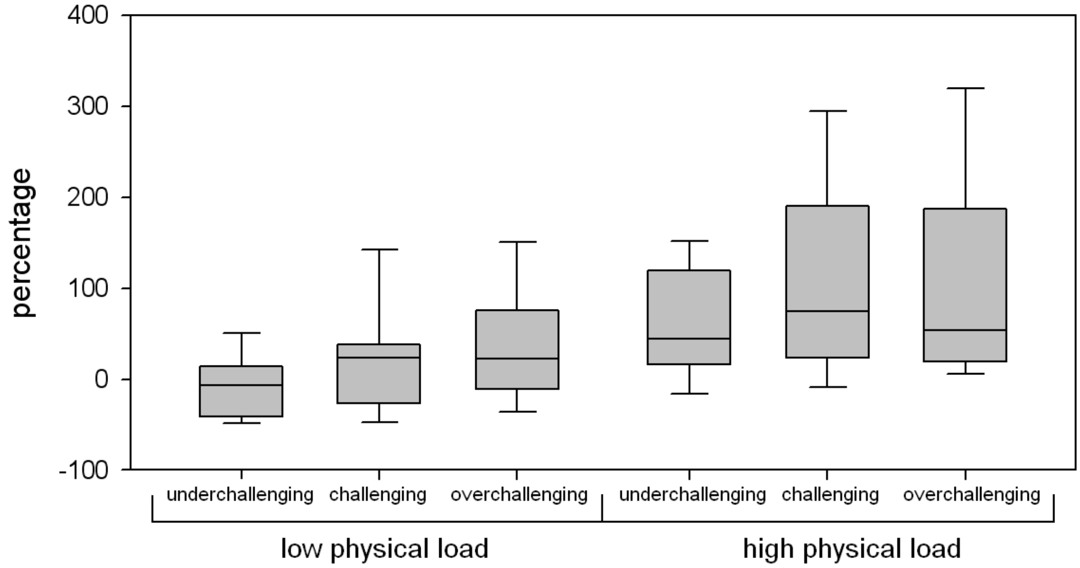


Figure 3.3: Normalized values of *SCR frequency* in different conditions.

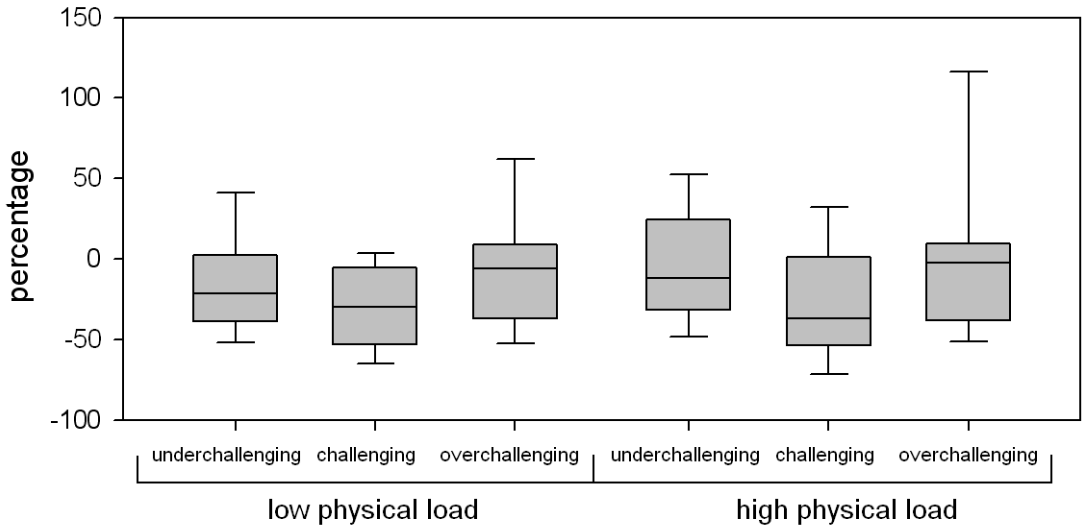


Figure 3.4: Normalized values of *respiratory rate variability* in different conditions.

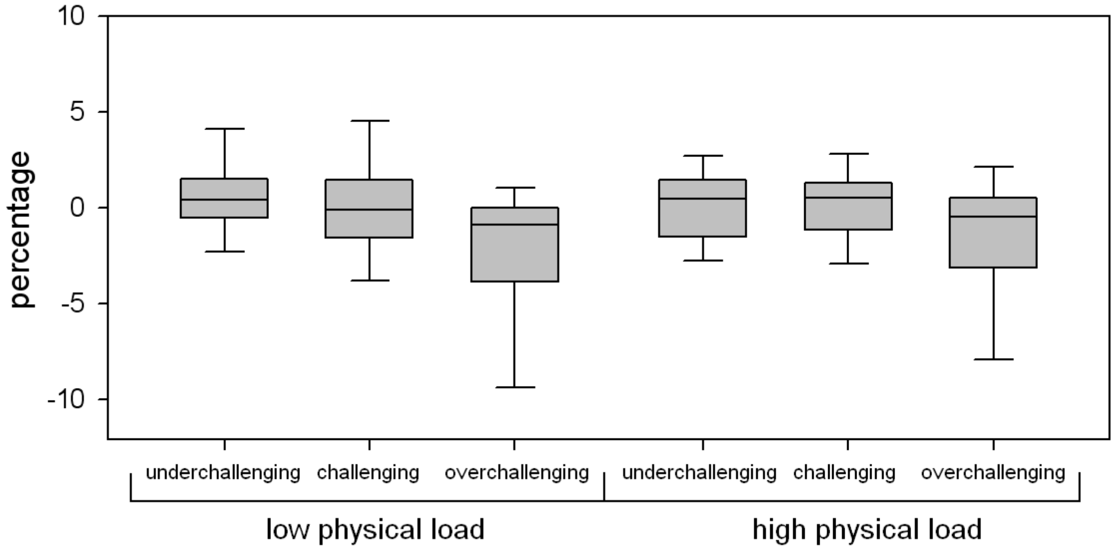


Figure 3.5: Normalized values of *final skin temperature* in different conditions.

3.2.6 Discussion

First of all, the biomechanical measurements show that subjects were not equally physically active in the three cognitive difficulty levels. The highest physical activity was during the underchallenging condition. In a way, this is logical since the measures taken to affect cognitive workload also affect the task kinematics. This in turn affects forces and physical workload. Thus, it is necessary to be cautious when comparing physiological responses to different difficulty levels. Despite a main effect of cognitive workload, the change in a physiological response may actually be caused by the increased physical workload associated with task difficulty. This is most likely the case for *mean heart rate*, *RMSSD*, *pNN50* and the *LF/HF ratio*, which show the highest deviation from baseline during the underchallenging high physical workload condition, where physical workload was by far the highest (as seen, for example, from the mean absolute force measurements).

Both *mean SCL* and *SCR frequency* are also affected by physical workload, as seen in Table 3.4. The main effect of physical workload on *SCR frequency* is not quite significant ($p = 0.067$), but nonetheless it is clear that skin conductance is affected by physical activity. This is not surprising since skin conductance changes as a result of sweating. It is, however, surprising that there was no significant effect of cognitive workload on *mean SCL* even though SCL is a generally well-established indicator of cognitive workload

[12]. One possibility is that the effect of physical workload on SCL masks the effect of cognitive workload. An alternative possibility is that, as a result of habituation, the effect of cognitive workload on SCL fades within the five-minute period while the effect of physical workload does not. There is, however, a significant effect of cognitive workload on *SCR frequency* that does not appear to be due to higher physical workload in the underchallenging condition since *SCR frequency* is higher in the challenging and overchallenging conditions. Thus, *SCR frequency* may have been an effective indicator of cognitive workload if the effect of physical workload had been better-controlled.

There was, however, no significant effect of physical workload on either respiration or skin temperature, suggesting that these are more robust to changes in physical workload. Both *respiratory rate variability* and *final skin temperature* differentiated between challenging and overchallenging conditions, with *final skin temperature* additionally differentiating between underchallenging and overchallenging conditions as well. Skin temperature in particular seems the least affected by physical workload. While both *mean respiratory rate* and *respiratory rate variability* (Figure 3.4) show visible if nonsignificant differences between low and high physical workload, *final skin temperature*, on the other hand, exhibits much smaller differences between low and high physical workload (Figure 3.5). *Respiratory rate variability* is lower than baseline for all task conditions, but is lowest for the challenging condition (Figure 3.4). A possible explanation is that it decreases as cognitive workload increases, but increases again as the challenge becomes too much to handle. Respiratory rate variability is known to decrease as a result of cognitive workload [60], confirming part of this explanation. *Final skin temperature* only significantly decreases from baseline in the overchallenging condition, not in the other conditions (Figure 3.5). Thus, it may be a good indicator of when a subject is overworked. Previous studies have found decreases in skin temperature as a result of tension and anxiety [65], supporting this explanation. However, other studies have found skin temperature to decrease as a result of cognitive workload [64]. If skin temperature decreases due to cognitive workload, it should also decrease during the challenging condition. One possibility is that a certain threshold of cognitive workload must be exceeded before skin temperature decreases.

Though heart rate has been used as a psychophysiological indicator in many studies, results of this study suggest that, in haptic interaction, it is primarily influenced by physical workload. *Mean heart rate* showed a large main effect of physical workload (partial $\eta^2 = 0.64$), and by far the greatest increase was during the underchallenging high physical workload condition where the exerted *mean absolute force* was also the greatest. Similarly, *RMSSD*, *pNN50* and the *LF/HF ratio* showed the largest deviation from baseline in the underchallenging high physical workload condition. The main effect of cognitive workload on *total HF power* was nearly significant ($p = 0.067$) and may have been able to differentiate between the underchallenging and the other two conditions, but this result is fairly unreliable. It appears that the increase in heart rate and HRV due to physical workload can completely overshadow any psychological effects. Since the effects of physical workload on heart rate have been extensively studied, a possible solution in future studies would be to collect information about physical workload from sensors built into the haptic robot. This information could be used in conjunction with a physiological model to provide an estimate of the effects of physical workload on heart rate. Such a model has already been developed for use in robot-assisted lower extremity rehabilitation [97], so it should be possible to develop a similar model for the upper extremities.

To sum up briefly, both heart rate and skin conductance are significantly affected by physical activity. At high levels of physical activity, the effects of cognitive workload on these two responses may be completely obscured. Respiration and especially peripheral skin temperature are less noticeably affected and are likely to be more effective at a higher level of physical activity. Considering that the four signals can be used to discriminate between different levels of cognitive and physical workload, they appear to be sufficient for the goal of the dissertation: to keep the patient moderately challenged during rehabilitation and keep him/her from becoming bored (very low cognitive workload) or frustrated (very high cognitive workload). They are also unobtrusive compared to additional sensors that might provide more detailed information but are unlikely to be accepted by patients in clinical practice (e.g. electroencephalography). However, these findings are valid only on the level of statistical analysis and only for healthy subjects. An analysis of psychophysiological responses in hemiparetic stroke patients is presented in the next section.

3.2.7 Others' contributions

The physical model of the inverted pendulum task and the experiment protocol were designed with the assistance of my advisor, Matjaž Mihelj. The Simulink drivers and blocks for HapticMaster control were developed by a number of other colleagues in the laboratory.

3.3 The effects of stroke

3.3.1 Tasks

Though very useful for the analysis of the effects of physical workload in section 3.2, the inverted pendulum task is not commonly used in rehabilitation. The analysis of the effects of stroke, however, should focus primarily on a task suitable for upper extremity rehabilitation. This task should ideally have at least two difficulty levels so that the differences in psychophysiological responses to different difficulty levels can be observed. Such a task has already been developed at the Laboratory of Robotics and is described in section 3.3.1.1.

The virtual rehabilitation task combines both physical and cognitive workload. While either element can be reduced, it is impossible to remove either physical or cognitive work from the task. In order to evaluate psychophysiological responses to differently demanding tasks, it was thus decided to also include a task with physical but minimal cognitive workload and a task with cognitive but no physical workload. For the first, a simple physical control task was used and described in section 3.3.1.2. For the second, a classical psychological task was used and described in section 3.3.1.3. By analyzing the differences between these tasks, it is possible to get an idea of how subjects who have experienced a stroke respond to cognitive and physical challenges. Hypothesis H2.2 can be tested using the cognitively demanding task, hypothesis H2.3 can be tested using the virtual rehabilitation task, and hypothesis H2.1 can be tested by having both controls and stroke patients perform all of these tasks.

3.3.1.1 Virtual rehabilitation task

Developed by colleagues at the Laboratory of Robotics at the University of Ljubljana, the virtual rehabilitation (VR) task combines reaching and grasping exercise. A photo of a subject performing the task using the HapticMaster robot is shown in Figure 3.6. In the centre of the screen, there is a table sloped toward the subject. At the beginning of the task, a ball appears at the top of the slope and starts rolling downward. The subject's goal is to catch the ball before it reaches the lower end of the table. Once the ball is grasped, a basket appears above the table. The subject must then hold the ball and place it in the basket. Once the ball is dropped into the basket or falls off the table, another ball appears at the top of the table, the basket disappears and the task continues. A quiet applause is also played over the speakers when the ball is successfully placed in the basket. The different steps of the task are shown in Figure 3.7. The robot allows the subject to feel each virtual item.

The robot offers various modes of haptic support. If a subject is unable to perform any or all of the following, the robot will actively guide his or her arm in order to move left or right and reach the ball, squeeze the grasping device in order to grasp the ball, and/or lift the ball into the basket. For reaching support (left-right movements), the robot pulls the subject's hand toward the ball with a maximum force of 10 N. The subject can thus reinforce or resist the robot's guiding force with his or her active arm movement. If the subject does not resist the robot, the reaching support system will reach the ball in a majority of cases (but may miss the ball if the starting position of the subject's hand is sufficiently far away from the ball). For grasping support, the robot automatically grasps the ball as long as the subject's hand is in the correct position, regardless of whether the subject is squeezing the grasping module. For lifting support, the subject's hand is pulled along a predefined trajectory [98] toward the basket. The subject can reinforce or resist the guiding force with his or her active arm movement. If the subject does not resist lifting support or release the ball, the ball will always be successfully placed into the basket by the lifting support system.

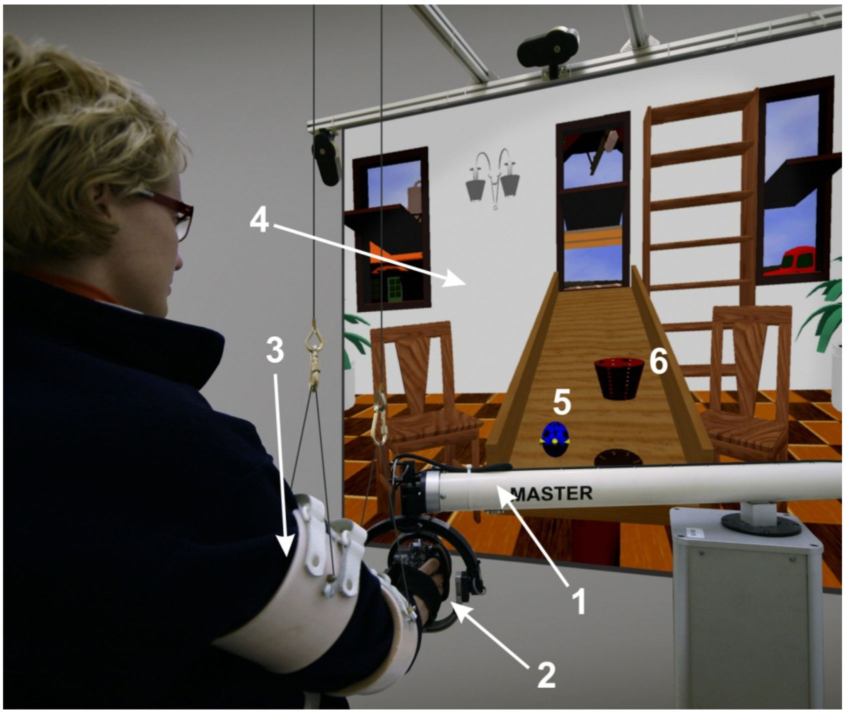


Figure 3.6: A subject performing the virtual rehabilitation task using the HapticMaster (1) and grasping device (2) while his/her arm is supported by cuffs (3). The screen (4) shows a sloped table, a ball (5) and a basket (6).

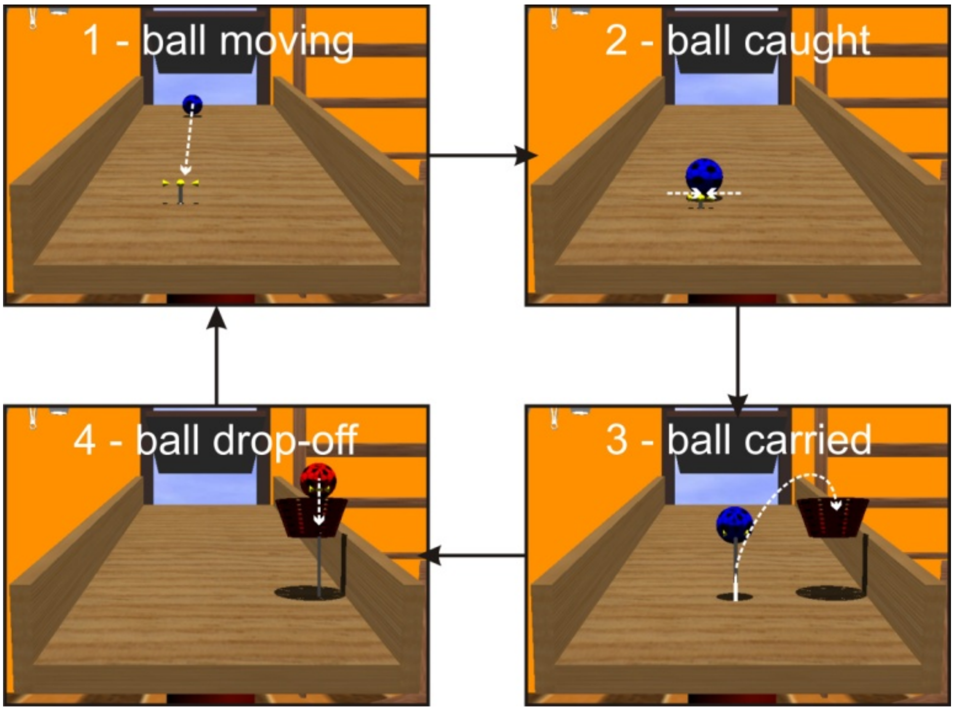


Figure 3.7: The virtual rehabilitation task. A ball appears on the top of a sloped table (1) and begins to roll down. The subject then catches it (2) and carries it toward a basket that appears above the table (3). Once the ball is above the basket (4), the subject drops it into the basket and a new ball appears.

A second, harder version of the task (henceforth referred to as the harder VR task) was also designed. Meant to be more cognitively demanding but equally physically demanding, the harder VR task had inverted left-right controls. If the subject moved his or her arm to the left, the virtual hand on the screen moved right (and vice-versa).

Two performance features were calculated for the VR task. They were the *percentage of caught balls* and *percentage of balls placed into the basket* (calculated as percentage of all balls in the period, not as percentage of caught balls). The *percentage of caught balls* was calculated only for subjects who did not receive catching support from the robot while the *percentage of balls placed into the basket* was calculated only for subjects who received neither catching nor lifting assistance.

3.3.1.2 Physical control task

In addition to the VR task, it was decided to also evaluate psychophysiological responses to a less complex task. To this end, a physical control task was introduced where subjects moved the robot left and right at a moderate speed while nothing was shown on the display and all force feedback was disabled. While both the physical control task and the VR task require reaching and some degree of coordinated movement, the physical control task is less complex and less cognitively demanding since it does not require lifting movements, does not provide visual stimuli and is not timed. Thus, since psychophysiological responses are strongly influenced by cognitive workload, the physical control task should evoke weaker psychophysiological responses.

3.3.1.3 Stroop word-colour interference task

In addition to the VR task, subjects were presented with a task that required only cognitive effort: a variant of the Stroop word-color interference task [25] that has been extensively studied by psychologists. Subjects were shown a word on the screen. The word was either “red”, “blue” or “green”. The color of the letters was also red, blue or green – but the word and the color of the letters did not always match. Subjects were given a keypad with three grey buttons, with the words “red”, “blue” and “green” written above the buttons (Figure 3.8). They were told to ignore the meaning of the word and, as quickly as possible, press the button corresponding to the color with which the word was written. Once a button was pressed, a new word was generated. Occasionally, however, the word was generated in black color. In this case, subjects had to push the button corresponding to the meaning of the word rather than its color. The words were randomly generated with the following probabilities: 40% chance of word with matching color, 40% chance of word with different color, 20% chance of word in black.

Two performance features were calculated for the Stroop task: *percentage of correct answers* (i.e. correctly chosen colors) and the *mean answer time* (the interval from the moment a color was displayed to the moment the subject pressed any button).

Figure 3.8: The keypad used for the Stroop word-colour interference task. The words 'red', 'green' and 'blue' are written above the buttons in Slovenian. During the task, the keypad was strapped to the subject's leg so that it would stay in one place.

3.3.2 Measurement protocol

The experiment was conducted in a dedicated room at the University Rehabilitation Institute of the Republic of Slovenia. The room and equipment are shown in Figure 3.9. Three people were present: the subject, experiment supervisor and occupational therapist. Upon arrival, subjects were informed of the purpose and procedure of the experiment. They signed an informed consent form and filled out the BAS/BIS questionnaire. Then, they were seated in front of the robot. The affected arm was strapped into the cuffs and grasping device, and the physiological sensors were attached. The normal VR task was demonstrated, and subjects were allowed to practice it briefly. Each subject practiced for at least two minutes, and more time was given to any subject who had not yet attained a basic level of proficiency. During practice, the three modes of support were set manually for each subject. Then, subjects went through the following procedure: rest period, physical control task, rest period, normal VR task, harder VR task.

After the harder VR task, the keypad used for the Stroop task was strapped to the subject's upper leg so that it would not fall off (Figure 3.8). The Stroop task was explained and demonstrated, and at least two minutes were once again given to practice. Then, subjects went through a three-minute rest period followed by the Stroop task. They pushed the buttons on the keypad with the thumb of their unaffected hand. The skin conductance and skin temperature sensors were not removed, but remained on the other fingers of the unaffected hand. While this may have affected measurements, it was necessary since many subjects were unable to push buttons with the affected hand. After the Stroop task, the experiment was concluded and a brief informal interview was conducted.

Each task and rest period lasted three minutes, and the SAM was presented on the screen after each period. Subjects verbally made a selection for both arousal and valence scales. Subjects remained quiet during rest, as these periods served as baseline periods for physiological measurements. The periods were shortened from five minutes (section 3.2) to three in order to shorten the overall experiment; despite the frequent rest periods, patients in pretesting found the entire experiment to be too long and tiring with five-minute periods.

While the experiment supervisor and occupational therapist maintained silence in the room during most baseline and task periods, the therapist was permitted to provide verbal guidance and encouragement during the VR task. This was unavoidable, as some subjects in the stroke group required guidance to perform the task without becoming excessively frustrated. To establish similar conditions, the therapist also provided verbal guidance to the control group. If guidance was not necessary, encouraging statements were provided to ensure that all subjects were verbally stimulated.

Figure 3.9: The room where the experiment took place. The task was displayed on a large screen using backprojection (middle left). The subject (center, blurred for anonymity) sat in front of the screen. The robot (not visible) was situated between the subject and screen. A metal frame (top) supported the screen, speakers and gravity compensation motors. The experimenter (lower right) sat to the side and had access to a secondary task display screen (lower center) as well as a separate computer used for physiological recordings (lower left).

3.3.3 Participants

The stroke group consisted of twenty-three subjects (age 51.0 ± 13.3 years, age range 23-69 years, 16 males, 7 females). They were diagnosed with subarachnoid hemorrhage (4 subjects), intracerebral hemorrhage (9 subjects) or cerebral infarction (10 subjects). As a result of the stroke, 13 suffered from hemiparesis of the left side of the body and 10 suffered from hemiparesis of the right side of the body. All were right-handed before the stroke. All were undergoing motor rehabilitation at the University Rehabilitation Institute of the Republic of Slovenia. Time between stroke onset and the experiment session was 154 ± 79 days. A majority of the group had received secondary stroke prevention drugs (including antihypertensives) prior to participation in the study. Three had received insulin due to diabetes (but had no diabetes-related complications), two had received ischemic heart disease treatment drugs, five had received SSRIs, three had received low doses of antiepileptics (preventive doses following aneurysm surgery), three had received short-acting sedatives and one had received a low dose of antipsychotics.

A day before the session, subjects in the stroke group were tested with both the mini-mental state examination (MMSE) [99] and the Functional Independence Measure (FIM) [100]. Score on the MMSE was 27.2 ± 3.6 (out of a possible 30). All but three subjects scored between 26 and 30. Of the remaining three, one scored 24 but was not excluded from the study since he was able to communicate and comprehend the tasks without problems. The other two subjects had lower scores due to dysphasia. These two were interviewed by a clinical expert and approved for participation in the study. Score on the FIM was 101 ± 13 (out of a possible 126).

The control group consisted of twenty-three subjects (age 50.5 ± 12.6 years, age range 24-68 years, 16 males, 7 females) with no major physical or cognitive defects. All were right-handed. To better match the control group and the stroke group, 13 controls performed the tasks with their left hand while 10 performed the tasks with their right hand.

In the VR task, four subjects in the stroke group required reaching support, seven required grasping support, and eight required lifting support (with some subjects requiring multiple types of support). The control group did not receive any support from the robot. The two subjects with dysphasia were excluded from the Stroop task.

3.3.4 Statistical methods

Before describing the methods used to analyze the data, it should be mentioned that the analysis of most biomechanical features will not be presented here. The analysis was chiefly done by J. Ziherl and published as a separate research paper [101]. It is thus not part of this dissertation. Only one biomechanical feature is included here: *total work* (section 2.3.5), which allows the effects of physical activity to be illustrated. Similarly, the analysis of the BAS/BIS scales will not be presented here since it was chiefly performed by M. Milavec and has been submitted to the Slovenian journal Horizons of Psychology as a separate paper (Milavec et al., under review as of mid-2011).

The **first step** of the data analysis was to examine performance and *total work* during the two levels of the VR task and the Stroop task. For the two levels of the VR task, a mixed-design ANOVA with one between-subjects factor (group: stroke or control) and one within-subjects factor (task difficulty: normal or harder VR task) was used for each feature. For the Stroop task, t-tests were used to compare the two performance features between groups. The goal of this step was to determine whether the control group performed better than the stroke group and whether performance during the harder VR task was worse than during the normal VR task.

The **second step** of the data analysis was to compare nonnormalized values of psychophysiological features between baseline and task periods. This was done separately for each task in a mixed-design ANOVA with one between-subjects factor (group: stroke or control) and one within-subject factor (period type: baseline or task). The goal of this step was to determine whether baseline values of physiological features are different between the groups, whether each task causes significant psychophysiological changes and whether these changes are different between the stroke and control groups.

The **third step** of the data analysis was to compare normalized values of psychophysiological features between the different task periods. The comparison of normalized values was done in a mixed-design ANOVA with one between-subjects factor (group: stroke or control) and one within-subjects factor (task type: physical control task, normal VR task, harder VR task and Stroop task). The goal of this step was to determine whether the different task periods cause different psychophysiological responses at an aggregate level. For instance, the harder VR task could evoke larger psychophysiological responses than the normal VR task since most subjects can be expected to find it more cognitively demanding.

The **fourth step** of the data analysis was to correlate normalized values of psychophysiological features with normalized results of the SAM and with performance data. Spearman correlations were used in cases involving results of the SAM (where the data is ordinal). Pearson correlations were used in other cases. *Valence* and *arousal* were normalized by subtracting the baseline value prior to calculating correlations. The goal of this step was to analyze the connections between physiological and non-physiological responses. For instance, due to the large differences between subjects, it is entirely possible that the ANOVAs performed in the previous step would show no significant difference between the normal and harder VR task. However, a correlation would show that the psychophysiological response to a task does depend on the subject's arousal, valence or performance during the task.

It should be noted that four subjects from the stroke group (including the two with dysphasia) and two from the control group reported no changes in valence or arousal during the experiment. Such a lack of changes in self-reported arousal is likely to be the result of a misunderstanding of the SAM. In fact, during the final informal interview, three of these six subjects mentioned how active they were during the tasks compared to the baseline period. Thus, these subjects' SAM results were considered unreliable and discarded.

The threshold for significance was set at $p = 0.05$. Due to space constraints, most results with $p > 0.05$ are not reported. For significant effects in ANOVA, effect size for a factor is also reported as partial η^2 (the proportion of total variability attributable to the factor, excluding other factors from the total nonerror variation [93]). The Sidak correction [94] for multiple comparisons was used for all post-hoc tests. The Huynh-Feldt correction [95] was used in cases of violations of sphericity in ANOVA. The Kolmogorov-Smirnov test with Lilliefors' modification [96] was used to test for normality.

3.3.5 Results

3.3.5.1 Performance and work

In the normal VR task, the stroke group caught 63% of all balls placed and placed 51% of all balls into the basket while the control group caught 84% of balls and placed 72% of balls into the basket. In the harder VR task, the stroke group caught 48% of balls and placed 36% of balls into the basket while the control group caught 51% of balls and placed 44% of balls into the basket. As previously stated, these percentages were calculated only for subjects who did not receive the relevant haptic support.

In the two levels of the VR task, there was a main effect of task difficulty (normal vs. harder task) on:

- *percentage of caught balls* ($p < 0.001$, partial $\eta^2 = 0.66$),
- *percentage of balls placed into the basket* ($p < 0.001$, partial $\eta^2 = 0.61$),
- *total work* ($p = 0.002$, partial $\eta^2 = 0.20$).

There was also a main effect of group (stroke vs. control) on:

- *percentage of caught balls* ($p = 0.018$, partial $\eta^2 = 0.14$),
- *percentage of balls placed into the basket* ($p = 0.005$, partial $\eta^2 = 0.20$),
- *total work* ($p < 0.001$, partial $\eta^2 = 0.36$).

There was an effect of interaction between task difficulty and group for:

- *percentage of caught balls* ($p = 0.003$, partial $\eta^2 = 0.22$),
- *percentage of balls placed into the basket* ($p = 0.004$, partial $\eta^2 = 0.22$),
- *total work* ($p = 0.001$, partial $\eta^2 = 0.23$).

Post-hoc tests showed that both groups caught fewer balls (stroke: $p = 0.002$; control: $p < 0.001$) and placed fewer balls into the basket (stroke: $p = 0.02$; control: $p < 0.001$) in the harder VR task. The control group performed less *total work* in the harder VR task than in the normal VR task (46.9 ± 17.4 J vs. 38.2 ± 13.7 J, $p < 0.001$) while the difference was not significant for the stroke group (30.1 ± 9.2 J vs. 30.5 ± 8.2 J).

In the Stroop task, the control group had a higher *percentage of correct answers* (stroke: 93.9 ± 5.6%; control: 97.9 ± 4.9%; $p = 0.007$) and lower *mean answer time* (stroke: 2.8 ± 1.6 s; control: 2.0 ± 1.2 s; $p = 0.016$).

3.3.5.2 Temporal changes of psychophysiological signals

The usefulness of psychophysiological signals strongly depends on how quickly and how strongly the signals react to stimuli. The physiological signals of a typical control subject during baseline (rest), the physical control task and the normal VR task are shown in Figures 3.10, 3.11 and 3.12 in order to illustrate how quickly and how much the signals change. For skin conductance (Figure 3.10), an increase can be seen during tasks (with the signal responding to the beginning of the task within seconds), and this general increase is gauged by *mean SCL*. Additionally, a greater number of brief increases in the skin conductance signal appear during the two task periods – there is a higher *SCR* *frequency*. For skin temperature (Figure 3.11), there is a slight decrease during the physical control task followed by a return to baseline as well as a larger decrease during the VR task (although temperature only begins decreasing approximately half a minute after the task begins). Heart rate (Figure 3.12) shows high variability during both baseline and task periods, but increases during tasks (though not as quickly as skin conductance). Respiratory rate is not shown since most subjects' differences between baseline and task periods were obscured by the high variability of the signal.

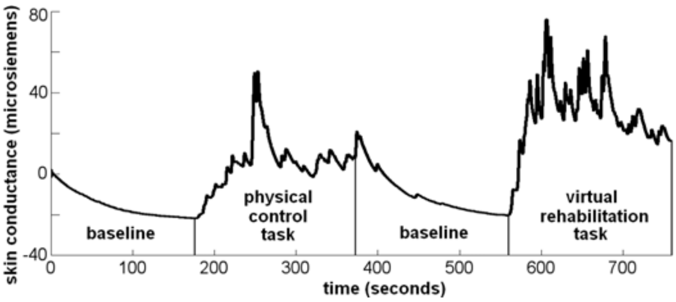


Figure 3.10: A typical subject's skin conductance as a function of time during two baseline periods, the physical control task and the virtual rehabilitation task. The initial value was defined as zero.

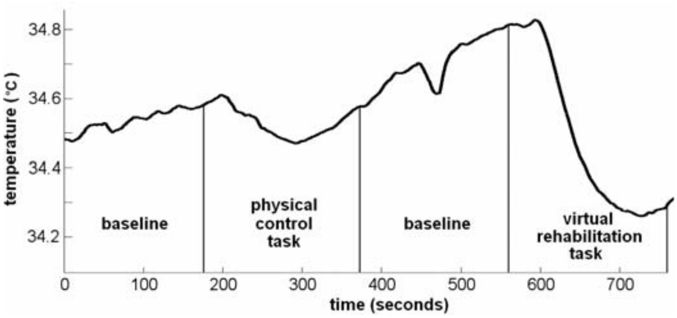


Figure 3.11: A typical subject's skin temperature as a function of time during two baseline periods, the physical control task and the virtual rehabilitation task.

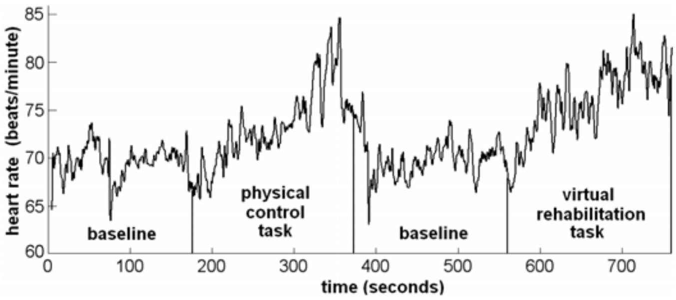


Figure 3.12: A typical subject's heart rate as a function of time during two baseline periods, the physical control task and the virtual rehabilitation task.

3.3.5.3 Baseline-task comparisons

Comparison of absolute values of psychophysiological features between baseline and task periods was done separately for each task in a mixed-design ANOVA. Table 3.5 shows differences between baseline and task for all four task periods and for both groups, as well as the main effect size of time period (baseline vs. task).

Table 3.5: Differences between baseline and task (mean ± standard deviation) and main effect size of baseline-task.

	physical control task			normal VR task			harder VR task			Stroop task		
	stroke	control	p. η^2	stroke	control	p. η^2	stroke	control	p. η^2	stroke	control	p. η^2
valence (scale 1-9)	-0.4±1.0	0.1±0.6	0.00	-0.2±1.2	0.1±0.9	0.00	-0.5±1.6	-0.1±1.4	0.04	-0.2±1.4	0.1±0.7	0.00
arousal (scale 1-9)	1.5±1.8	1.0±1.6	**0.35****	1.5±1.9	2.4±1.4	**0.59****	1.9±1.7	3.5±1.7	**0.72****	1.0±2.1	1.6±1.9	**0.31****
mean heart rate (bpm)	2.7±2.4	4.9±2.7	**0.70****	4.2±3.8	6.6±5.3	**0.59****	4.0±4.2	7.9±5.1	**0.63****	2.6±2.2	7.2±7.9	**0.43****
SDNN (% of baseline)	12±35	25±49	**0.13***	12±34	15±41	0.04	8±34	27±58	0.08	-19±21	-7±26	**0.27****
RMSSD (%)	10±46	-7±31	0.06	26±85	7±35	0.03	12±53	11±38	0.01	-3±24	-8±34	0.09
mean resp. rate (bpm)	1.7±2.8	3.0±2.3	**0.47****	2.1±2.9	2.4±2.8	**0.39****	1.7±2.7	2.2±3.4	**0.30****	2.6±2.0	3.6±2.3	**0.69****
resp. rate var. (%)	76±115	20±73	0.05	45±117	64±98	**0.11***	52±107	89±99	**0.24****	26±92	15±79	0.01
mean SCL (µS)	8±19	12±13	**0.31****	25±32	24±21	**0.47****	31±39	33±28	**0.48****	22±31	22±21	**0.43****
SCR frequency (%)	41±68	128±325	**0.34****	91±136	82±148	**0.57****	95±144	91±165	**0.41****	172±456	164±351	**0.65****
mean SCR amp. (µS)	-0.6±4.9	1.0±1.3	0.00	0.7±1.5	2.4±2.4	**0.38****	0.8±1.4	1.7±1.9	**0.37****	0.3±1.4	1.6±2.2	**0.22****
final temperature (K)	0.2±0.6	-0.2±0.4	0.01	-0.3±0.6	-0.4±0.6	**0.24****	0.2±0.4	-0.8±0.9	**0.25****	-0.1±0.8	-0.9±1.0	**0.24****

bpm = beats per minute or breaths per minute. % = percentage of baseline value. µS = microsiemens. K = kelvin.

p. η^2 = partial eta-squared value for main effect of time period (baseline vs. task). Bolded values and asterisks indicate significance of effect at $p < 0.05$ (*) or $p < 0.01$ (**).

Main effect size of group (stroke vs. control) and interaction effects are listed separately in text.

Significant effects of group (stroke vs. control) were found for:

- *arousal* (partial $\eta^2 = 0.12$ in physical control task and 0.11 in normal VR task; higher *arousal* in stroke group in both cases),
- *mean heart rate* (partial $\eta^2 = 0.23$ in physical control task, 0.18 in normal VR task and 0.15 in harder VR task; higher *mean heart rate* in stroke group in all cases),
- *SDNN* (partial $\eta^2 = 0.26$ in physical control task, 0.21 in normal VR task, 0.27 in harder VR task and 0.21 in Stroop task; lower *SDNN* in stroke group in all cases).

Significant effects of interaction between group and time period were found for:

- *arousal* (partial $\eta^2 = 0.18$ in harder VR task; the control group showed a larger increase in *arousal*),
- *mean heart rate* (partial $\eta^2 = 0.16$ in physical control task, 0.15 in harder VR task and 0.14 in Stroop task; the control group showed a larger increase in heart rate in all cases),
- *RMSSD* (partial $\eta^2 = 0.17$ in physical control task and 0.19 in normal VR task; the control group showed a smaller increase in *RMSSD* in both cases),
- *mean SCR amplitude* (partial $\eta^2 = 0.12$ in normal VR task and 0.15 in Stroop task; the control group showed a larger increase in SCR amplitude in both cases),
- *final skin temperature* (partial $\eta^2 = 0.14$ in harder VR task and 0.11 in Stroop task; the control group showed a larger decrease in skin temperature in both cases).

In order to better illustrate some of the findings, Figures 3.13, 3.14, 3.15 and 3.16 show box plots of differences between baseline and task for four physiological features: *mean SCL, final skin temperature, mean heart rate and mean SCR amplitude*. A positive value represents an increase from baseline. On the box plots, the middle line represents the median, the upper and lower edges of the rectangle represent the 25th and 75th percentiles, and the whiskers represent the 10th and 90th percentiles.

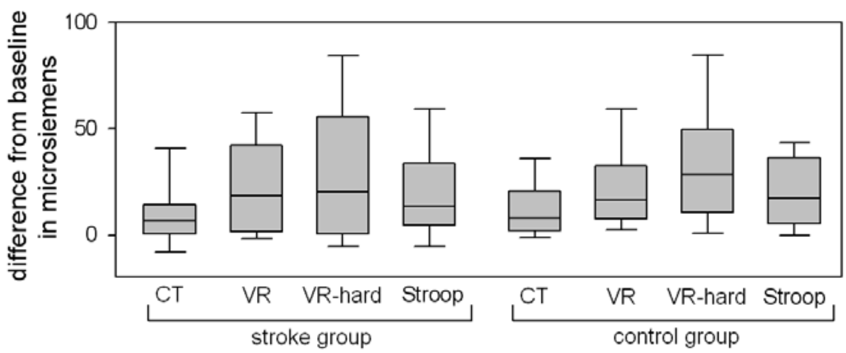


Figure 3.13: Changes in *mean SCL* as a response to different tasks. CT = physical control task, VR = normal virtual rehabilitation task, VR-hard = harder virtual rehabilitation task.

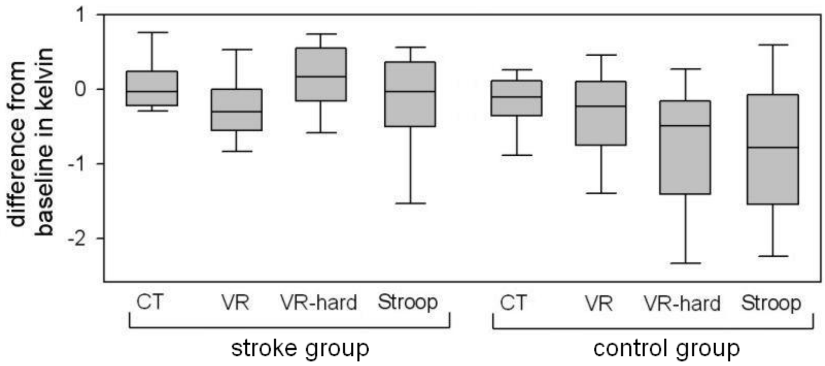


Figure 3.14: Changes in *final skin temperature* as a response to different tasks. CT = physical control task, VR = normal virtual rehabilitation task, VR-hard = harder virtual rehabilitation task.

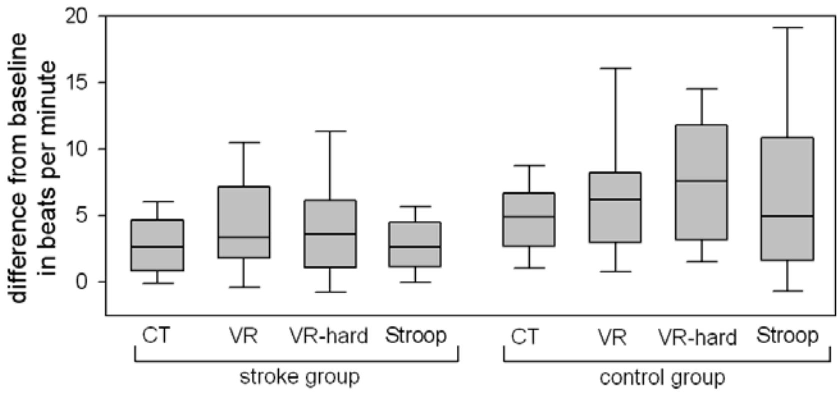


Figure 3.15: Changes in *mean heart rate* as a response to different tasks. CT = physical control task, VR = normal virtual rehabilitation task, VR-hard = harder virtual rehabilitation task.

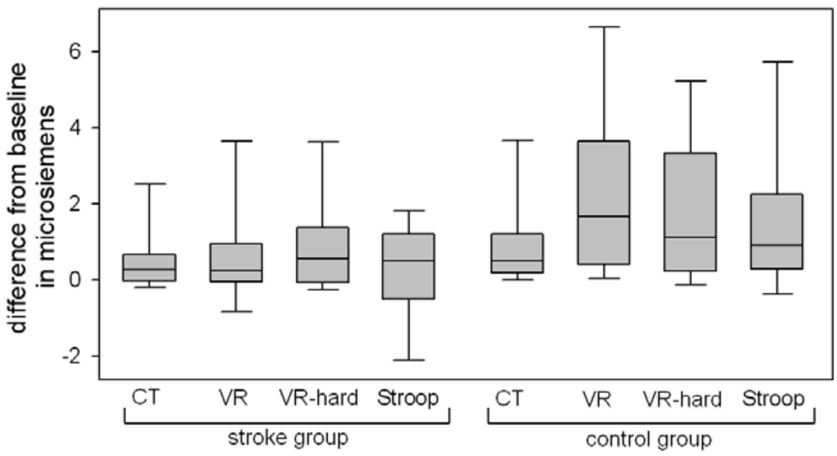


Figure 3.16: Changes in *mean SCR amplitude* as a response to different tasks. CT = physical control task, VR = normal virtual rehabilitation task, VR-hard = harder virtual rehabilitation task.

3.3.5.4 Comparison of normalized values between tasks

The comparison of normalized values was done in a mixed-design ANOVA with one between-subjects factor (group: stroke or control) and one within-subjects factor (task type: physical control task, normal VR task, harder VR task and Stroop task). As the focus was primarily on differences between the physical control task and the two versions of the VR task, differences between the Stroop task and the other three task periods are not reported.

Analysis of self-reported *arousal* found an effect of task type ($p < 0.001$, partial $\eta^2 = 0.24$). Post-hoc tests found higher *arousal* in the harder VR task than in the physical control task ($p < 0.001$) and the normal VR task ($p < 0.001$). There was also an effect of interaction between task type and group ($p < 0.001$, partial $\eta^2 = 0.18$). Post-hoc tests found that the control group showed significant differences between the physical control task, the normal VR task and the harder VR task ($p < 0.05$ for all three pairwise comparisons), but that the stroke group showed no significant differences in *arousal* between these three tasks.

Analysis of *SDNN* found an effect of task type ($p < 0.001$, partial $\eta^2 = 0.15$). However, post-hoc tests found that the only significant differences were between the Stroop task and the other tasks. Similarly, analysis of *RMSSD* found an effect of task type ($p = 0.031$, partial $\eta^2 = 0.08$), but post-hoc tests found no significant differences between the tasks. Analysis of *respiratory rate variability* found an effect of task type ($p = 0.039$, partial $\eta^2 = 0.07$), but post-hoc tests found no significant differences between the tasks. There was also an effect of interaction between task type and group ($p = 0.024$, partial $\eta^2 = 0.08$). Post-hoc tests showed a difference between the physical control task and the harder rehabilitation task in the control group, but no difference in the stroke group.

Analysis of *mean SCL* (Figure 3.13) found an effect of task type ($p < 0.001$, partial $\eta^2 = 0.18$). Post-hoc tests showed significant differences between all three tasks ($p < 0.05$ for all three pairwise comparisons).

Analysis of *mean SCR amplitude* (Figure 3.16) found an effect of task type ($p = 0.049$, partial $\eta^2 = 0.08$), but post-hoc tests found no significant differences between the tasks. There was also an effect of group ($p = 0.009$, partial $\eta^2 = 0.17$).

Analysis of *final skin temperature* (Figure 3.14) found an effect of task type ($p = 0.010$, partial $\eta^2 = 0.10$). Post-hoc tests found that temperature was lower in the normal VR task than in the physical control task ($p = 0.010$). There was also an effect of group ($p = 0.001$, partial $\eta^2 = 0.24$) and an effect of interaction between task type and group ($p = 0.005$, partial $\eta^2 = 0.11$). Post-hoc tests found that, in the stroke group, temperature in the normal VR task was lower than in both the physical control task and the harder VR task. In the control group, temperature in the harder VR task was lower than in the physical control task and the normal VR task.

3.3.5.5 Correlations

Table 3.6 shows significant correlations between different features for the control group while Table 3.7 shows significant correlations for the stroke group. It should be restated that all SAM results and all psychophysiological features were normalized. Furthermore, the number of subjects is listed next to correlations since some subjects' SAM results were excluded as described in section 3.3.4.

Table 3.6: Significant correlations between different features for the control group. CT = physical control task, VR = virtual rehabilitation task. Correlation coefficients are Spearman's (ρ) for all correlations involving the SAM (valence or arousal) and Pearson's (r) for all others.

task	feature 1	feature 2	p	ρ or r	N
CT	total work	SCR frequency	< 0.001	0.66	23
VR	arousal	SCR frequency	0.004	0.60	21
VR	valence	SCR frequency	0.046	0.44	21
VR	valence	mean respiratory rate	0.051	0.43	21
VR	% of caught balls	RMSSD	0.023	-0.49	23
Stroop	arousal	SCR frequency	< 0.001	0.81	21
Stroop	valence	% of correct answers	0.043	0.46	21
Stroop	mean answer time	respiratory rate variability	0.031	0.47	23

Table 3.7: Significant correlations between different features for the stroke group. CT = physical control task, VR = virtual rehabilitation task. Correlation coefficients are Spearman's (ρ) for all correlations involving the SAM (valence or arousal) and Pearson's (r) for all others.

task	feature 1	feature 2	p	r or r	N
CT	arousal	mean SCL	0.045	0.51	19
VR	arousal	percentage of balls placed into the basket	0.029	0.66	19
VR	arousal	SCR frequency	0.019	0.59	19
VR	valence	percentage of caught balls	0.045	0.56	19
VR	valence	percentage of balls placed into the basket	0.047	0.61	19
VR	valence	final skin temperature	0.017	0.62	19
VR	percentage of balls placed into the basket	mean SCL	0.040	0.49	23
VR	total work	mean SCL	0.021	0.50	23
Stroop	arousal	final skin temperature	0.042	-0.54	19
Stroop	valence	percentage of correct answers	0.016	0.65	19
Stroop	valence	mean respiratory rate	0.026	0.59	19
Stroop	valence	respiratory rate variability	0.045	-0.58	19
Stroop	mean answer time	RMSSD	0.025	0.51	21
Stroop	mean answer time	respiratory rate variability	0.044	0.47	21

3.3.6 Discussion

3.3.6.1 Stroop task

In the Stroop task, where no physical activity was required, the expected responses to a cognitive task were noted in both the stroke and control groups: increased *SCL* (Figure 3.13) [12], increased *SCR frequency* [57, 58], decreased *final skin temperature* (Figure 3.14) [64], increased *mean heart rate* (Figure 3.15) [12, 13], decreased HRV [13, 14] and increased *mean respiratory rate* [13, 59] relative to baseline.

There was a significant main effect of group on physiological responses. The stroke group showed higher *mean heart rate* and lower *SDNN* than the control group during both baseline and task periods, confirming the results of previous studies that have found increased heart rate [87] and decreased HRV [23] after stroke.

In the comparison of physiological features between baseline and task, there was also a significant interaction effect. No decrease in *final skin temperature* from baseline to task was noted in the stroke group (Figure 3.14), and the control group showed a larger increase in *mean heart rate* than the stroke group (Fig. 3.15). This shows that the stroke group exhibits different physiological responses even to a cognitive task with no physical activity.

3.3.6.2 Comparison of normalized values between tasks

Mean SCL was the only feature that showed a significant difference between the physical control task, the normal VR task, and the harder VR task (Figure 3.13). Additionally, it showed no significant effect of group or group-task interaction. Since skin conductance is a well-documented indicator of arousal [12, 57, 58] and also showed large, rapid changes in this study (Figures 3.10 and 3.13), it seems to be the most effective indicator of physiological differences between difficulty levels in this task. However, another question needs to be answered to ensure reliability. Since the harder VR task was always performed after the normal VR task, was there an influence of task order? In other words, would skin conductance have kept increasing even if the normal VR task had been followed by an easier task? At the end of a task and beginning of a baseline period, skin conductance decreases again (Figure 3.10). Additionally, a qualitative examination of the recorded signals showed that most subjects' (N = 20 for patients, N = 21 for controls) skin conductance reached a plateau within approximately a minute and then stayed at that plateau or even decreased slowly. Such a plateau can be seen for both task periods in Figure 3.10, and a slow drift can be observed for the VR task in the same figure. However, some subjects (N = 3 for patients, N = 2 for controls) did show a constant rise in skin conductance throughout the VR task, so the influence of task order cannot be ruled out.

Another interesting feature is *final skin temperature*, which showed significant task-group interaction effects. In the stroke group, it decreased from baseline during the normal VR task but was actually higher than baseline in the harder VR task (Figure 3.14). In the control group, however, the greatest decrease was in the harder VR task. This could be explained by a transient nature of skin temperature changes [63]. In several subjects, skin temperature began to return to baseline levels toward the end of a task period, thus suggesting that skin temperature will return to baseline after several minutes' exposure to a constant stimulus. This can be seen in the physical control task in Figure 3.11. The control group likely does not find the normal VR task to be difficult (in the final informal interview, many stated that they found it too easy) and thus shows a larger decrease in temperature during the harder VR task where subjects need to focus more. The stroke group, however, is already challenged by the normal VR task. If the harder VR task is not much more challenging, temperature may return to baseline. An alternative explanation could be that the relationship between workload and skin temperature may not be monotonic. While the inverted pendulum study in section 3.2 found that skin temperature decreases as workload increases, it is possible that temperature increases again at a very high level of workload, much as respiratory rate variability did in section 3.2. A final, third possible explanation is that the harder VR task may be so difficult that some subjects in the stroke group simply give up and no longer try hard. This was also noted in the informal interviews.

One factor that may have blurred differences between the two VR tasks was the verbal assistance of the occupational therapist. In the harder VR task, the stroke group may have relied on the therapist more than the control group. Nonetheless, it was impossible to carry out the study without the therapist's verbal advice since several subjects in the stroke group needed guidance to perform the task properly.

3.3.6.3 Correlations – arousal, valence, performance and work

In both groups, *valence* was correlated with task success (*percentage of correct answers*) in the Stroop task. However, only the stroke group showed a correlation between *valence* and task success in the VR task (*percentage of balls placed into the basket*). A possible explanation is that the stroke group finds the task challenging and is thus pleased by success while the control group does not find the task difficult and is thus less concerned about performance.

In the VR task, the stroke group's *arousal* was correlated with the *percentage of balls placed into the basket*. The lack of a correlation between *arousal* and performance in the control group could once again be explained by the fact that control subjects likely did not find the task to be difficult.

3.3.6.4 Correlations – arousal and physiological features

Looking first at the Stroop task, which requires no physical effort, there was a significant correlation between self-reported *arousal* and *SCR frequency* in the control group. This is in agreement with previous studies that have found SCR frequency to be a good indicator of arousal [57, 58]. Surprisingly, there was no significant correlation between *arousal* and *SCR frequency* in the stroke group. There was, however, a correlation between *arousal* and *final skin temperature*. While this is also in agreement with studies that found connections between skin temperature and cognitive workload [64], it is interesting that neither group shows both correlations.

In the VR task, both groups showed a correlation between *arousal* and *SCR frequency*. This raises the question of why the stroke group showed a correlation between *arousal* and *SCR frequency* in the VR task, but not the Stroop task. Additionally, in the physical control task, neither group showed a significant correlation between *SCR frequency* and *arousal*. It is possible that, during a task that requires only physical workload, *SCR frequency* is not a good measure of *arousal*. Still, of all tested features, *SCR frequency* appears to be the most reliable indicator of self-reported arousal.

Mean SCL, the only physiological feature that showed a significant difference between the different tasks, was only correlated with *arousal* in the physical control task (and only for the stroke group).

3.3.6.5 Correlations – valence and physiological features

The stroke group's *valence* was correlated with *mean respiratory rate* and *respiratory rate variability* in the Stroop task and with *final skin temperature* in the VR task. Evidence does exist for connections between respiratory variability and anxiety [62] as well as between skin temperature and anxiety [65]. However, respiratory rate has mainly been associated with arousal and cognitive workload. Similarly, the control group's *valence* was correlated with *SCR frequency*, which is a documented indicator of arousal rather than valence.

Responses of the autonomic nervous system appear to be better at indicating arousal than valence. This is to be expected. Skin conductance is regulated exclusively by the sympathetic branch of the autonomic nervous system and is thus poor at distinguishing different levels of valence [41]. Connections between heart rate and valence are, at the moment, controversial (see Peter and Herbon [41] for examples). Similarly, while some studies have reported a connection between skin temperature and tension/anxiety, others have found that skin temperature is also primarily regulated by the sympathetic branch of the autonomic nervous system [63] and a better indicator of arousal. Respiratory variability may be an indicator of valence, but this is a complex issue since studies have found different respiratory variability responses to different negative emotions [62].

3.3.6.6 Correlations – performance, work and physiological features

For the control group, there was a correlation between HRV and *percentage of caught balls* in the VR task. The lower the *RMSSD*, the more balls the subject caught. Since decreases in HRV have been linked to increased cognitive workload [13, 14], it is probable that subjects who concentrated harder also performed better.

Similarly, a correlation was found between the control group's *mean answer time* and *respiratory rate variability* in the Stroop task. Since decreases in respiratory variability have been linked to increased cognitive workload [60], it is again likely that subjects who focused harder were able to answer faster.

In the physical control task, the control group showed a correlation between *SCR frequency* and *total work*. This may indicate an influence of physical activity on skin conductance. Such an influence is expected (exertion would cause sweating), but would likely make it more difficult to separate the physiological effects of cognitive and physical workload.

For the stroke group, *mean answer time* in the Stroop task was also correlated with HRV (*RMSSD*) and *respiratory rate variability*. The reasoning for this is identical to the reasoning above for the control group. In the VR task, *mean SCL* was correlated with the *percentage of balls placed into the basket*. Although no significant correlation was found between *arousal* and *mean SCL*, skin conductance is a known indicator of arousal, and *arousal* is also significantly correlated with the *percentage of balls placed into the basket*. *Mean SCL* was also correlated with *total work*. This may also indicate an influence of physical activity on skin conductance.

3.3.6.7 Study limitations

Three limitations of the study should be highlighted: limitations of self-report measures, lack of task order randomization, and potential physiological effects of drugs.

First, the study found some connections between self-report measures and physiological responses that were at odds with previous research. For instance, correlations were found between valence and skin conductance even though skin conductance has been shown to be almost exclusively affected by arousal. Additionally, there were some discrepancies between physiological and self-report measures. However, these issues are not as problematic as they may appear. Several other studies have found only weak associations between physiological responses and self-reported emotions [82]. Some studies even found that subjects are sometimes not aware of their own emotions or are simply unwilling to report them [83]. This was noted in this study as well. Several subjects commented that they found certain tasks to be highly engaging, but did not report any increase in *arousal* on the SAM. Others reported the same *valence* for the entire session despite obvious frustration and annoyance (expressed by, for example, cursing and annoyed facial expressions). Thus, imperfect connections between self-report measures and psychophysiological responses should be taken not only as a limitation of this specific study, but also as an unavoidable facet of research in human-robot interaction.

Second, the order in which tasks were performed was not randomized. Thus, there may have been an influence of task order on features such as *mean SCL* and *final skin temperature*. While it is true that a fully randomized task order would have been preferable from a methodological viewpoint, it was also known in advance that the number of subjects would be limited. As stroke patients are already likely to exhibit large intersubject variability due to impairments of cognitive and motor ability, it was decided in the planning phase to avoid task order randomization since it would further increase variability and potentially obscure important results. The decision was thus made for purely practical reasons.

Finally, several subjects had received drugs that may have affected psychophysiological responses. For instance, sedatives have been shown to affect skin conductance [102]. Antiepileptics and antipsychotics may have also had an effect. There may have been an influence of secondary stroke prevention drugs, but since these are commonly used in stroke rehabilitation, their effects cannot be avoided and could be taken as inherent in that subset of the population.

3.3.7 Others' contributions

Matjaž Mihelj and Marko Munih helped design the experiment protocol and oversaw the study. The ball-catching scenario and the HapticMaster control algorithms were programmed by Jaka Ziherl and Andrej Olenšek. Maja Milavec helped select and prepare the questionnaires. Nika Goljar, MD, of the University Rehabilitation Institute oversaw the study at the Institute and selected suitable patients. Metka Javh, Janja Poje and Julija Ocepek were the occupational therapists who guided the patients during the experiment sessions and ensured their safety.

3.4 Summary of findings

The two studies described in this section separately analyzed the effects of physical activity on psychophysiological responses (by having subjects perform the same task at different levels of physical and cognitive workload) and the effects of stroke on psychophysiological responses (by having stroke and control groups perform the same tasks). However, the two studies did not identify a clearly preferable psychophysiological signal.

The first study found that respiration and skin temperature were fairly robust with regard to physical activity. The second study, on the other hand, found few significant results for respiration and showed that changes in skin temperature are much smaller after a stroke. Furthermore, while the second study found skin conductance to be the most informative (with regard to both correlations and differences between different conditions), the first study noted an important influence of physical activity on *mean SCL* and, to a lesser degree, on *SCR frequency*. Thus, it appears that the two factors influence different psychophysiological responses. Both studies, however, agree that only limited psychophysiological information can be gleaned from heart rate during motor rehabilitation since it is significantly influenced by both physical activity and stroke. Nonetheless, heart rate can be useful in rehabilitation since the patient also should not be physically overworked.

Having found that all of the used psychophysiological responses are, to some degree, affected by either physical activity or stroke, it was decided to nonetheless continue with the second part of the dissertation: data fusion and biocooperative control. However, it was important to keep in mind that data fusion would likely be more difficult in stroke patients than in healthy controls due to the effects of stroke. Furthermore, the effects of physical activity would have an effect on data fusion in all subjects. After these two studies, however, it was not yet possible to know whether the physical activity would make data fusion more difficult by masking responses to psychological stimuli. Another possibility was that, since the overall goal was to determine how suitable the task is for the patient, the effects of physical activity would contribute additional information regarding task suitability (a heart rate of 120 beats per minute, for instance, should be avoided regardless of its cause).

4 Data fusion and biocooperative control

4.1 Introduction

Having performed the statistical analysis of psychophysiological responses in section 3, the next step was to implement data fusion and biocooperative control. Specifically, the overall goal was to determine how suitable the rehabilitation task currently is for the patient and then adjust task difficulty accordingly in order to make the experience better for the patient. This is, however, not a trivial goal. There has been a long history of researchers using psychophysiological measurements to try to identify psychological states. However, despite a vast body of literature available on the subject, there is still no universally accepted set of rules that would translate physiological data to psychological states. Aside from theoretical limitations to inferring significance from psychophysiological data [35], there are entirely practical disagreements among psychologists – for instance, whether the subject's psychological state can be classified into one of several basic emotions (anger, sadness, fear, surprise, happiness…) [36] or whether it is defined with multiple independent variables such as arousal and valence [37]. Furthermore, while valence and arousal can be defined as continuous variables, another possibility is to define arousal-valence quadrants: low arousal/positive valence, low arousal/negative valence, high arousal/positive valence and high arousal/negative valence.

Though no universal set of rules for fusion of psychophysiological measurements currently exists, it is possible to look at all the work that has been done and identify some of the most promising strategies. Thus, it was decided to review psychophysiological studies performed in the last ten years, examine all those dealing with the fusion of autonomic nervous system responses, and implement some of the most promising data fusion methods. The review also included those studies that use psychophysiological measurements as a form of biofeedback: to make changes to the environment around the subject in response to psychophysiological changes. This is important since biocooperative control is, in essence, psychophysiological biofeedback applied to rehabilitation robotics.

Figure 4.1 shows the general process of measuring, interpreting and using autonomic nervous system responses in psychophysiology. Three of these steps (signal recording, feature extraction and normalization) were already described in section 2 and utilized in section 3. A review of the remaining three steps (dimension reduction, classification/estimation and biofeedback) is presented in the next three subsections (4.1.1 – 4.1.3). Finally, section 4.1.4 specifies the methods implemented for the dissertation. When choosing the methods to implement, the overall goal of the dissertation was also taken into account. While many psychophysiological studies try to distinguish many psychological states or emotions, the goal in motor rehabilitation is simply to keep the subject from becoming bored or stressed. Thus, a relatively simple psychological model should be sufficient.

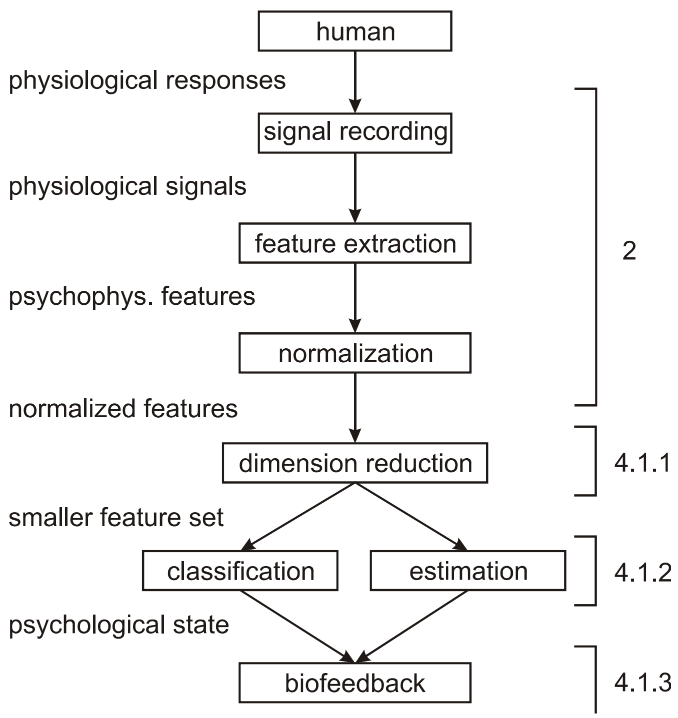


Figure 4.1: The general process of measuring, interpreting and using autonomic nervous system responses in psychophysiology. The blocks contain the human (top) and different steps to be performed. The data used at each stage is written on the left, while the numbers on the right show which section of this dissertation describes the different steps.

4.1.1 Dimension reduction

When multiple psychophysiological signals are measured, it is possible that a very large number (20+) of features will be extracted from them. In data fusion, this can lead to the 'curse of dimensionality': with a large number of features, an extremely large training data set is required to build an accurate classifier or estimator. If the training data set is too small, overfitting can occur – data fusion rules trained on a small data set may not generalize well to new data. Thus, it can be beneficial to reduce the number of features prior to data fusion. Some data fusion methods already incorporate dimension reduction (e.g. classification tree pruning, random forests – section 4.1.2.1.5), but several general dimension reduction techniques also exist.

The techniques commonly used in psychophysiology can be roughly divided into three types: selection of individual features that ignores correlations between different features (section 4.1.1.1), projection of the feature space onto a lower-dimensional space (section 4.1.1.2) and selection of individual features that takes correlations between different features into account (section 4.1.1.3). While the first and third type are mutually exclusive (with the third type having been found superior to the first as described in section 4.1.1.3), the second can be used together with either of the other two (as described in section 4.1.1.3). Other dimension reduction techniques that have seen limited use in psychophysiology are described in section 4.1.1.4. All these techniques can be used online. For online use, either the best features are selected in advance or the projection of the feature space is calculated in advance. Online data fusion then either uses the selected features or transforms the features using the precalculated projection rule.

4.1.1.1 Individually best features

A simple way to select the most appropriate features for data fusion is to rank the features according to a criterion of how much information each individual feature provides. Then, the best features (either a preselected number of features or all those who exceed a certain predefined threshold) are selected for data fusion. In psychophysiology, the most common way to rank individual features has been through ANOVA, correlations and chi-square tests: statistical methods that find differences between different conditions (e.g.

between 'sad' and 'angry' emotions) or connections between different variables (e.g. between arousal and heart rate). Only features that show significant differences between conditions or significant connections between different variables can then be used in data fusion.

ANOVA was used for feature selection by Wagner et al. [103], where psychophysiological features were ranked according to their p-value and a preselected number of most significant features were selected. ANOVA was also used by van den Broek et al. [75], where features with a p-value below 0.001 were selected. The chi-square test was used for feature selection by Pour et al. [104], where the ten most significant features were chosen. Correlations with self-reported psychological variables were used for feature selection by Liu et al. [45] and Rani et al. [105], where only psychophysiological features that had an absolute correlation coefficient of at least 0.3 were chosen. Correlations were also used to identify the most relevant psychophysiological features by Bailenson et al. [106], though this information was not later used in data fusion.

The weakness of this approach is that it ignores correlations between different psychophysiological features. For example, if two features correlate highly with self-reported arousal, they may also correlate highly with each other. In this case, it may make sense to only include one of the two features in data fusion since the other one would not provide enough additional information to justify its inclusion.

4.1.1.2 Principal component analysis and Fisher's projection

Principal component analysis (PCA) is a method that transforms a large number of features into a smaller number of uncorrelated features (called principal components) that explain as much of the variability in the data as possible. Since it ensures that the principal components are uncorrelated with each other, PCA has an advantage over methods from the previous section which ignore correlations between different features. It has been used for dimension reduction in several psychophysiological studies [75, 103, 107, 108]. However, it does have one important weakness: while the principal components explain as much of the variability in the data as possible, there is no guarantee that they are better-correlated with psychological states than the original features. If we have a training data set where each data point is labeled with a specific class (e.g. anger, fear…), it would be useful to take the labels into account during feature selection in order to ensure that the selected features discriminate between different classes. PCA, however, ignores any data labels.

The above weakness of PCA is addressed by Fisher's projection, which can be thought of as a supervised alternative to PCA. While PCA projects the original features onto a lower-dimensional space in such a way as to explain as much of the variability in the data as possible, Fisher's projection projects the original features onto a lower-dimensional space where between-class scatter is maximized and within-class scatter is minimized. In other words, it projects the original data into a lower-dimensional space where different classes (e.g. anger, fear…) are easier to linearly separate. Fisher's projection is essentially a version of linear discriminant analysis (section 4.1.2.1.3), except used for dimension reduction rather than classification. It has been used in several psychophysiological studies [46, 109, 110]. One weakness of Fisher's projection should be noted, however: since it transforms the original feature space into a space where different classes are linearly separable, it is less suitable for use with nonlinear data fusion methods such as neural networks.

4.1.1.3 Sequential feature selection

Unlike PCA and Fisher's projection, which linearly transform the feature space, sequential feature selection methods (also known as stepwise methods) are methods that sequentially select individual features from the feature space. Unlike the approaches presented in section 4.1.1.1, however, sequential feature selection methods do not ignore connections between different features.

Perhaps the most common sequential feature selection method is sequential forward selection, which works as follows. In the first step of the sequence, no features are included in the selection. The method evaluates all features to determine which one best discriminates between classes in the training data set (using criteria such as the F-value of each feature). That feature is included in the selection. In the next steps, all remaining features are evaluated to determine which one best discriminates between classes after the contributions of all previously selected features have already been taken into account. This process continues until no remaining feature contributes enough additional information to warrant its inclusion (for instance, the *F*-value of all remaining features is lower than a certain value). Sequential forward selection has been used in several psychophysiological studies [78, 103, 111-113]. It has been shown to outperform the approach of selecting the best individual features (section 4.1.1.1) [78, 112].

A very similar method to sequential forward selection is sequential backward selection. The difference is that, while forward selection begins with no features in the selection and sequentially adds features, backward selection begins with all features in the selection and sequentially removes features according to which one contributes the least to discrimination between classes. The process continues until the contribution of all remaining features exceeds a given threshold (for instance, the *F*-value of all remaining features is higher than a certain value). Sequential backward selection has been used by Kim et al. [114, 115], who reported that it outperformed sequential forward selection (though quantitative results were not reported for forward selection).

A combination of the above two methods is sequential floating forward selection (SFFS), sometimes called sequential forward-backward selection. Starting with no features included in the selection, it sequentially adds features like sequential forward selection, but at each step it also evaluates whether any of currently included features can be removed. The most common criterions for inclusion or exclusion are *F*-value thresholds: a higher one for inclusion and a lower one for exclusion. SFFS has been used in several psychophysiological studies, including Picard et al. [46], Gu et al. [110] as well as Wilson and Russell [116], where it was used in conjunction with discriminant analysis (section 4.1.2.1.3) and thus called stepwise discriminant analysis.

It should finally be noted that Fisher's projection and sequential feature selection are not mutually exclusive. Instead of providing Fisher's projection with all possible features, it is possible to first select a subset of features using sequential feature selection and use Fisher's projection on this subset. This was first done in a study by Picard et al. [46], where the combination of the two approaches outperformed both of the two approaches used individually. Wagner et al. [103] also found improved performance when using both approaches, though not for all classification methods. Finally, a combination of the two approaches was used by Gu et al. [110], though it was not compared with using either approach individually. In principle, PCA could also be used with sequential feature selection, and both Fisher's projection and PCA could be used with selection of individually best features (section 4.1.1.1). However, this has not been done in psychophysiology and there is no strong rationale for it since sequential feature selection has been shown to outperform individually best feature selection and since Fisher's projection takes class labels into account while PCA does not.

4.1.1.4 Other

The aforementioned feature selection methods are of course not the only ones; they are simply the most prevalent in psychophysiology. Other methods include, for instance, Davies-Bouldin clustering [137, 138], BestFirst [122] and genetic algorithms [113]. However, these methods have not yet seen much use in psychophysiology, and further studies will be required before their strengths and weaknesses can be properly assessed.

4.1.2 Classification and estimation

This section describes several possible methods for psychophysiological data fusion - a process which takes a psychophysiological feature vector (consisting of several features extracted from multiple physiological signals) as input and assigns a psychological label to it. This psychological label can be categorical, in which case the feature vector is assigned to one of possible classes (e.g. 'angry', 'sad', 'low stress', 'high stress') and the process is called classification (section 4.1.2.1). Alternatively, the label can be a continuous value (e.g. an arousal of 9.2 on a scale between 0 and 10), in which case the process is called estimation (section 4.1.2.2). Since different methods are generally used for the two approaches, they are described in different sections. They are not, however, equally popular; in psychophysiology, classification has been used far more than estimation. For each specific method, a brief description is provided followed by examples of use with autonomic nervous system responses in psychophysiology. Detailed descriptions of each method are available in pattern recognition textbooks such as Bishop [117].

Sections 4.1.2.1 and 4.1.2.2 describe different classification and estimation methods, and a comparison is then made in section 4.1.2.3. All of these methods can be used both offline and online (real-time), though some are more computationally intensive and thus perhaps less suitable for online use (as discussed in section 4.1.2.3).

4.1.2.1 Classification

4.1.2.1.1 Nearest neighbors

The *k*-nearest neighbor (kNN) algorithm is one of the simplest classification algorithms. When a new data point needs to be classified, the algorithm computes the (usually Euclidean or Mahalanobis) distance to each data point in the training data set. The training points are then ranked according to their distance to the new sample, and the k (where $k \geq 1$) nearest training points (neighbors) are used to classify the new data point using a majority vote: the sample is assigned to the class that is most common among the k nearest neighbors. The simplest version of this is the 1-nearest neighbor rule, where a new sample is assigned to the same class as the nearest point in the training data set. Before calculating distances, it is usually necessary to scale the different data features (e.g. normalizing each feature to [0 1]) so that all features contribute equally to the distance calculation. Dimension reduction is also usually necessary, since the algorithm otherwise weighs all features equally even though some may not be relevant.

Despite its simplicity, the kNN algorithm has become popular in psychophysiology. It has been used for:

- classification of basic emotions [46, 71, 72, 103, 108, 118, 119];
- classification of arousal-valence quadrants [75, 110, 120];
- classification of the level of frustration [121] or enjoyment [113].

An algorithm extremely similar to the kNN algorithm is the nearest class center. It differs in that, instead of distances being calculated for all data points, each class is represented by the center of the data points for that class (e.g. the mean and covariance of the data). A new data point is then simply assigned to the class with the nearest center, which is less computationally intensive than computing distances to every point. This approach was used in Setz et al. [74] and Frantzidis et al. [122].

4.1.2.1.2 Naïve Bayes classifier and Bayesian networks

A relatively simple classifier, the naïve Bayes classifier is based on Bayes' theorem and assumes that all features are independent of each other. Given the training data, it creates a probability model which estimates the probability that a data point belongs to a certain class. It then uses a decision rule to assign a class to the data point based on the probability model. Perhaps the most common rule is the 'maximum a posteriori' rule, which classifies the point as coming from the class with the highest posterior probability.

Like the kNN algorithm, the naïve Bayes classifier has proven surprisingly effective despite its simplicity [123]. One important advantage is that, by assuming independence between features, it requires a much smaller training data set than other, more complex methods. Since the sample size in applied psychophysiological studies is often limited, the naïve Bayes classifier could be an attractive option. Nonetheless, most studies have preferred more complex methods, so the naïve Bayes classifier has only been used in a few studies: for classification of basic emotions [46, 124], classification of arousal-valence quadrants [125] and classification of the stress level [69].

The naïve Bayes classifier is actually a very simple form of Bayesian network, a probabilistic model of random variables and their conditional dependencies. More advanced networks that do not assume that all features are independent of each other have also been used in psychophysiology. They were used for classification of basic emotions [124], type and intensity of basic emotions [119], arousal-valence quadrants [126] and frustration [121]. A Bayesian network was also used to select appropriate songs based on physiological features [127], though the user's mood was not explicitly classified.

4.1.2.1.3 Discriminant analysis

Discriminant analysis is a well-known classification method which finds a linear (linear discriminant analysis - LDA, also known as Fisher's linear discriminant) or quadratic (quadratic discriminant analysis - QDA) combination of input features which best separate data points into two or more classes. This combination of input features is essentially a hyperplane in n-dimensional space (where n is the number of input features) that separates data points of different classes. For a two-class problem, a linear discriminant function thus takes the form

$$D(\boldsymbol{x}) = \boldsymbol{w}\boldsymbol{x} + b \tag{4.1}$$

where $D(\boldsymbol{x})$ is the discriminant function, $\boldsymbol{x}$ is the vector of input features, $\boldsymbol{w}$ are the weights of the function and b is the intercept. $\boldsymbol{x}$ is then assigned to one class if $D(\boldsymbol{x})$ is positive and to the other class if $D(\boldsymbol{x})$ is negative. Both $\boldsymbol{w}$ and b are computed from training data as follows:

$$b = -\boldsymbol{w}^{\mathrm{T}} \cdot \frac{1}{2} \cdot (\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \tag{4.2}$$

$$\boldsymbol{w} = (\boldsymbol{S_1} + \boldsymbol{S_2})^{-1} \cdot (\boldsymbol{\mu_2} - \boldsymbol{\mu_1}) \tag{4.3}$$

where $\boldsymbol{S_i}$ is the covariance matrix for class i and $\boldsymbol{\mu_i}$ is the vector of mean feature values for class i.

Since each input feature has its own weight assigned to it, it is easy to determine how important it is to discrimination between classes. Though originally used for two-class problems, it can also be extended to multiclass situations. The greatest limitation of discriminant analysis is that it only allows linear or quadratic psychophysiological relations; if strongly nonlinear relations are expected in the data, other methods may be preferable.

Because it is easy to use and transparently shows the contribution of each feature to discrimination between classes, discriminant analysis has been a popular data fusion method in psychophysiology. LDA has been used for:

- classification of basic emotions [53, 66, 71, 72, 103, 107, 118];
- classification of arousal/valence quadrants [114];
- classification of the level of workload [116] or stress [109];
- discrimination of stress and cognitive workload [74];
- separation of phobic / non-phobic [111] or anxious / non-anxious subjects [128].

QDA, which allows some nonlinearity in classification, has been used by Chanel et al. [129] and Setz et al. [130]. Another variety of discriminant analysis is worth mentioning despite not having seen use in psychophysiology: diagonal discriminant analysis, which ignores correlations between different variables and is thus actually a type of naïve Bayes classifier. Specifically, diagonal QDA assumes that all classes have diagonal covariance matrices while diagonal LDA assumes that all classes have the same diagonal covariance matrix. Despite its simplicity, diagonal discriminant analysis has proven to be very effective for classification of nonpsychophysiological data [131].

4.1.2.1.4 Support vector machines

Similarly to discriminant analysis, support vector machines (SVMs) are a method of generating hyperplanes in n-dimensional space (where n is the number of input features) that separate data points of different classes. The principal difference between the two is the criterion used to calculate these hyperplanes. While LDA maximizes a discriminative projection, SVMs are a maximum margin classifier: they create the hyperplane so that the distance (margin) between the hyperplane and the closest data points on each side is maximized.

Basic SVMs thus have similar advantages and disadvantages as discriminant analysis. They are transparent and it is easy to determine the contribution of each input feature; but on the other hand, they are a linear classifier. To avoid the limitation of linearity, SVMs are commonly expanded using so-called kernels. A good explanation of kernels is provided in Schölkopf et al. [132], but in essence the training data is transformed into a higher-dimensional space and a hyperplane is generated in this space. While the hyperplane is linear in the new transformed space, it may be nonlinear in the original feature space, resulting in a nonlinear classifier.

The good performance and nonlinearity of SVMs has led to their frequent use in applied psychophysiology. They have been used for:

- classification of basic emotions [67, 104, 106, 119, 124, 133-135];
- classification of both the type and intensity of basic emotions [18];
- classification of arousal-valence quadrants [75, 120, 125, 129];
- classification of the level of stress [69], frustration [121] or enjoyment [18];

- discrimination of natural and unnatural behavior [73];
- discrimination of stress and cognitive workload [74].

4.1.2.1.5 Classification trees

Classification trees assign a class to a data point by progressing through several branching IF-THEN logical rules. This branching structure is the reason why they are called trees. An example of a psychophysiological classification tree would be “If SCR frequency is below five per minute, the subject is bored. Otherwise, if skin temperature is below 33 degrees Celsius, the subject is frustrated. Otherwise, the subject is entertained.” While not an accurate set of rules, this serves as a simple illustration of a classification tree. The rules are not defined manually; several different algorithms exist to learn the rules from training data. At each new node of the tree, these algorithms select the feature that best discriminates between classes after all the previous decisions made in the tree have been taken into account. Features are selected using criteria such as information gain.

Classification trees offer a very transparent way of classifying psychophysiological data. The decision process can be easily followed by researchers and can be visualized graphically, making the trees a very ‘white-box’ approach. The tree building process acts as a form of dimension reduction, and many tree-building algorithms also incorporate tree pruning, which prevents the tree from becoming too complex and overfitting the data. In psychophysiology, classification trees have been used for:

- classification of basic emotions [124];
- classification of both type and intensity of basic emotions [119];
- classification of arousal-valence quadrants [122, 125];
- classification of the level of stress [69] or anxiety [45, 105];
- distinguishing natural and unnatural behavior [73].

Advanced variants of classification trees have also been used in psychophysiology. One example are fuzzy trees [136], which combine the hierarchical structure of classification trees with fuzzy logic (described in section 4.1.2.2.2). A second example are ensemble methods such as random forests [108], and boosted decision stumps [106] - sets of many trees whose outputs are combined to produce the final classification.

4.1.2.1.6 Artificial neural networks

Inspired by biological systems, artificial neural networks (ANNs) consist of a large number of simple, interconnected components (‘neurons’) operating in parallel. Each neuron receives a number of inputs and uses them to calculate the ‘activation’ of the neuron. Perhaps the simplest way to calculate this activation is to calculate a weighted sum of the inputs, then set the output as 1 if the weighted sum exceeds a certain threshold and 0 if the weighted sum does not exceed the threshold. This output is then fed to the next layer of neurons and so on until the final output is determined. Such a layered network with weighted sums and threshold is called a multilayer perceptron. Multilayer perceptrons can model functions of very high complexity if enough layers and neurons are used. However, other types of ANNs that incorporate more complex elements also exist (e.g. radial basis function networks). Complexity can be especially increased by allowing outputs of one layer of neurons to be used as inputs to both preceding and succeeding layers. This type of network is called a feedback network.

ANNs are taught to perform a particular function using a training data set by adjusting the weights of the connections between different neurons. They are nonlinear tools capable of modeling very complex relationships between variables, which can be very useful in psychophysiology. To train an ANN as a classifier, it simply needs to be provided with a training data set where the inputs are physiological features and the outputs are numbers corresponding to different classes (e.g. 1 – ‘angry’, 2 – ‘sad’). However, ANNs have one important disadvantage. Once trained, it is difficult to determine how different input variables contribute to the output. ANNs thus provide users with little information about the underlying system. Despite this lack of transparency, they have been frequently used with psychophysiological data, specifically for:

- classification of basic emotions [71, 72, 103, 124, 137];
- classification of arousal-valence quadrants [70, 75, 76, 125, 138];
- classification of the level of workload [116, 139, 140];
- classification of entertainment value and preferences [78, 112].

4.1.2.1.7 Other

Though the previous subsections describe the classification methods commonly used in applied psychophysiology, there are also several other, less-often used methods that bear mentioning. Some of these are:

- Fuzzy logic: More properly an estimation technique, it has also been used for psychophysiological classification by simply assigning classes to different values of the output variable [105, 134]. Notable for not requiring a training data set, it is described in more detail in section 4.1.2.2.2.
- Hidden Markov models: Actually a type of dynamic Bayesian network (section 4.1.2.1.2), hidden Markov models are notable because they allow the classification of temporal sequences. Though popular in research fields such as speech recognition and activity recognition, they have seen little use in psychophysiology where the preferred approach is to calculate features from a temporal sequence and then classify those features instead. Two examples of their use in psychophysiological data fusion are Kulić and Croft [77] and Scheirer et al. [141].
- Relevance vector machines: Functionally similar to SVMs (section 4.1.2.1.4), relevance vector machines are embedded in a Bayesian framework. They have been shown to provide results similar to SVMs, but with sparser solutions. They were used for psychophysiological data fusion by Chanel et al. [129].
- Large margin algorithm: A simpler version of SVMs (section 4.1.2.1.4), the large margin algorithm makes certain assumptions about the data in order to reduce computational complexity. It was used by Yannakakis and Hallam [112], but is unlikely to see wider use in psychophysiology where computational complexity is generally not a problem.

4.1.2.1.8 Classifier fusion

Of course, it is not necessary to use only a single classifier to perform data fusion. It is also possible to combine several classifiers (of the same type or different types) either in series or in parallel and thus hopefully obtain a better result. Though this approach is not especially widespread in psychophysiology, the structure of the input data or the psychological model used may lend themselves naturally to classifier fusion. For instance, if other data modalities are used in addition to psychophysiological data, it is possible to obtain a classification result using data from each modality separately and then combine these unimodal results to obtain a final result. This is commonly called decision-level fusion and was performed using speech and psychophysiology [114] as well as with both central and autonomic nervous system responses [129]. On the level of sensors rather than modalities, Setz et al. [130] used the same approach to obtain a separate classification result from each physiological sensor (one result for all features extracted from the ECG, one result for all features extracted from skin conductance etc.) and then fuse them together.

The above approach features several classifiers working in parallel. An alternate option is to have several classifiers in series: the first classifier performs a rough separation into two nonspecific classes (e.g. low arousal/high arousal) while the following classifier classifies the data point into a specific subclass within the previous class. Two examples of this exist in psychophysiology, both involving the arousal-valence space. In both cases, the first classifier classifies the data point into one half of the arousal-valence space while the second classifier classifies the data point into one of the two remaining possible quadrants [115, 122].

4.1.2.2 Estimation

4.1.2.2.1 Linear sums and linear regression

Perhaps the simplest way to estimate a psychological quantity from psychophysiological features is to define it as a weighted sum of (usually normalized) psychophysiological features:

$$y(\boldsymbol{x}) = \boldsymbol{w}\boldsymbol{x} + b \tag{4.4}$$

where y is a psychological quantity (e.g. arousal), $\boldsymbol{x}$ are psychophysiological features (e.g. mean heart rate, SCR frequency), $\boldsymbol{w}$ are the weights assigned to the different features and b is the intercept. $\boldsymbol{w}$ and b can be defined manually (e.g. [142, 143]), but a more optimal approach is to perform the technique of linear regression on the training data set. Given a data set with known y and $\boldsymbol{x}$, linear regression usually estimates $\boldsymbol{w}$ and b using the least squares method, though other methods are also possible. It has been used for estimation of distress, worry and task engagement [144], estimation of amusement and sadness [106] and estimation of arousal [145].

4.1.2.2.2 Fuzzy logic

Fuzzy logic is an estimation procedure that makes use of easy-to-understand IF-THEN logical rules. The difference between fuzzy logic and classical logic is that statements in fuzzy logic do not have to be absolutely true or false, but have “degrees” of truth. There are thus also no hard boundaries between categories or exclusive memberships. Perhaps the most famous example of fuzzy logic involves temperature control, described with the statements: “If the room is cold, the heating should be set to maximum. If the room is hot, the heating should be off.” In fuzzy logic, the room can be both cold and hot to some degree (e.g. 0.8 cold, 0.2 hot), and the heating is thus also set to some intermediate value. An example from psychophysiology would be “if heart rate is high and skin conductance is high, arousal is high”. Ranges for each variable are defined using membership functions and can overlap.

Fuzzy logic is appropriate for situations where a precise mathematical model does not exist, but experts can identify general rules underlying the system – as in psychophysiology. It is also appropriate for systems with a high level of noise, which is also common in psychophysiology due to the intra- and intersubject variability. Expert-defined fuzzy rules have been used to estimate stress and anxiety [17, 146] as well as arousal and valence [16]. Expert-defined fuzzy rules are especially noteworthy because, unlike most of the methods described in this paper, they do not explicitly require training data.

If the underlying behavior of the system cannot be described by experts, machine learning approaches also exist to identify the parameters of a fuzzy logic system using training data. Examples of fuzzy system identification for the purpose of user state assessment from psychophysiological data are presented in Katsis et al. [134], Kumar et al. [147] and Ting et al. [148].

4.1.2.2.3 Artificial neural networks

Previously described in section 4.1.2.1.6, artificial neural networks (ANNs) consist of a large number of simple, interconnected components ('neurons') operating in parallel. They are taught to perform a particular function (which can be simple or very complex) using a training data set by adjusting the weights of the connections between different neurons. While mostly used in psychophysiology for classification, they do not necessarily have to output a categorical value (e.g. 1 – 'angry', 2 – 'sad'); they can easily be trained to output continuous values and thus estimate the level of a particular psychological variable. However, this approach has seen far less use than classification (section 4.1.2.1.6). One recent example is the work by Bailenson et al. [106], where ANNs are used to estimate the level of amusement and sadness.

4.1.2.3 Recommendations for use in rehabilitation

Having reviewed several data fusion methods, it is only natural to ask ourselves "which method is the best for a biocooperative rehabilitation system?" The answer, of course, is not simple and depends critically on the purpose of the system and the properties of the data.

4.1.2.3.1 Classification or estimation?

A choice must first be made between classification and estimation. As previously mentioned, classification has been used in psychophysiology far more often than estimation. This can partially be attributed to the psychological models used. If an experiment is built around basic emotions (anger, sadness, fear, surprise, happiness…) [36], the psychological state can be described as one of several discrete classes, naturally creating a classification problem. If the psychological state is described in terms of arousal and valence, estimation is more useful since arousal and valence are both continuous quantities [37]. Similarly, if the goal is to determine the level of a particular psychological variable (e.g. stress, anxiety), estimation can be the appropriate choice. However, since inducing a great number of arousal, valence, stress or anxiety levels can be difficult, researchers often settle for splitting psychological states into arousal-valence quadrants [122, 129] or discrete levels of a psychological variable [69, 121, 139], creating a classification problem again.

In rehabilitation, it is not necessary to identify a large number of emotional states; it is sufficient to identify when a user is bored or frustrated and take steps to remedy this. Classification with a small number of classes is thus sufficient for the purposes of this dissertation. However, there is still a large number of different classifiers that are available and widely used in psychophysiology

4.1.2.3.2 Selecting a classifier

Choosing an appropriate classifier depends on a number of factors. Perhaps the most important one is accuracy – how well it can classify data points. To evaluate accuracy, we can first turn to large-scale classifier comparisons from other fields. One comprehensive nonpsychophysiological comparison of classifiers on different real-world data sets was made in the 1990s by King et al. [149]. Other classifier comparisons with nonpsychophysiological data include Harper [150] (medical data), Hua et al. [151] (with a special focus on classifier accuracy as a function of sample size), Caruana and Niculescu-Mizil [152], and Caruana et al. [153]. Some general conclusions can be drawn from these comparisons that may also apply to psychophysiological data. For instance, classification trees seem to outperform discriminant analysis when used on data with high skewness and kurtosis [149].

A number of studies have compared different classifiers specifically on psychophysiological data. These are briefly summarized in Table 4.1. Unfortunately, it is again difficult to identify an optimal classifier, as different studies report results that may at first glance be contradictory. For instance, Nasoz et al. [71] find ANNs to perform much better than kNN, but van den Broek et al. [75] report higher classification accuracy with kNN than with ANNs. Similarly, Zhai and Barreto [69] find SVMs to be much more accurate than the naïve Bayes classifier, but Müller [125] reports similar accuracy for both methods.

Furthermore, classifier accuracy depends on many different factors such as the input features and the possible classes used. For instance, several studies report that certain classifiers are better at recognizing certain emotions [71, 103, 119], though it is unknown whether or not this is a sampling fluke. Rani et al. [105] compare classification trees and fuzzy logic on data sets of different qualities and find that while trees generally result in higher accuracy, fuzzy logic is more accurate if the data quality is low. Thus, direct comparison of classification accuracies between studies is difficult.

Table 4.1: Psychophysiological studies that compare different classifiers.

Study	Classifiers compared	Classification of?	# classes
Zhai and Barreto [69]	naïve Bayes, SVM, trees	stress or no stress	2
Nasoz et al. [71]	kNN, LDA, ANN	basic emotions	6
Lisetti and Nasoz [72]	kNN, LDA, ANN	basic emotions	6
Mohammad et al. [73]	SVM, trees	natural behavior	2
Setz et al. [74]	LDA, SVM, nearest center	stress, cognitive load	2
van den Broek et al. [75]	kNN, SVM, ANN	emotion types	4
Wagner et al. [103]	kNN, LDA, ANN	basic emotions	4
Rani et al. [105]	trees, fuzzy logic	anxiety level	3
Rigas et al. [108]	kNN, trees (random forest)	basic emotions	3
Yannakakis et al. [112]	ANN, large margin algorithm	game preferences	2
Kim and Andre [115]	LDA, classifier fusion	arousal/valence	4
Wilson and Russell [116]	LDA, ANN	low/high workload	2
Rani et al. [119]	kNN, SVM, trees, Bayesian network	basic emotions	5
Shen et al. [120]	kNN, SVM	arousal/valence	4
Kapoor et al. [121]	kNN, SVM, Bayesian network	frustration level	2
Calvo et al. [124]	numerous	basic emotions	8
Müller [125]	naïve Bayes, SVM, ANN, trees	arousal/valence	4
Chanel et al. [129]	LDA, QDA, SVM, RVM	arousal/valence	3
Katsis et al. [134]	SVM, fuzzy logic	basic emotions	4
Pastor-Sanz et al. [135]	kNN, naïve Bayes, SVM, trees	basic emotions	6

Since classification of psychophysiological features has not yet been used in rehabilitation, it is difficult to say which classifier would be the most accurate for a particular number of classes or input signals. However, it is not necessary to select only one. A common approach in psychophysiology is to record a larger data set in advance, then use the technique of crossvalidation to evaluate classifiers. The prerecorded data set is first divided into multiple parts. Then, the rules for classification are constructed using data from all but one part and tested on the remaining part. This process is repeated as many times as there are parts, with each part serving as the 'test' part exactly once. Crossvalidation provides an estimate of a classifier's accuracy when used on previously unseen data. A few examples of crossvalidation in psychophysiological data fusion can be found in Rani et al. [119], Zhai and Barreto [69] and Chanel et al. [129]. It is thus possible to select a larger number of potential dimension reduction and classification methods, then test them on an already recorded data set and identify the most accurate method. Given the lack of prior work with psychophysiological measurements in motor rehabilitation, this is recommended as the best approach at this time. This was also done in this dissertation, as described in section 4.1.4.

In addition to accuracy, a number of factors could be considered when selecting the best classifier to use. For instance, one might look at the speed and computational cost of a classifier. While all classifiers described so far can be used both offline and online, some are less suited for online use. Calvo et al. [124], for example, found ANNs to be much slower than SVMs and thus less suitable for online use. Here, it is important to differentiate between the time needed to train the classifier (which can be done in advance) and the time needed to apply the classifier to a new data point (which often needs to be done online). Discriminant analysis, for instance, is simple to both train and apply. SVMs and ANNs can be time-consuming to train, but can be applied to new data much faster. On the other hand, a nearest-neighbor classifier requires no advance training, but can be computationally intensive to apply to a new data point online since the distance to each data point in the training data set must be calculated in many dimensions. However, this is unlikely to be a problem in rehabilitation. If the user is bored or stressed, an action needs to be taken to correct this. If such actions are taken too often, though, they may upset the patient even further since he/she would not have time to adjust to the new rehabilitation task. Furthermore, given that many psychophysiological responses react slowly to stimuli, it is unnecessary to perform data fusion very often.

Another, generally less crucial factor is the transparency of the classifier. Rather than the most accurate classifier, we might choose a slightly less accurate classifier whose classification procedure can be easily understood by humans. In this case, classification trees provide a very transparent method since their if-then reasoning can be easily followed. Discriminant analysis is also fairly simple to understand while nonlinear methods such as neural networks are often looked down on despite attempts to do away with their reputation as a 'black box' [155]. Here, a consideration must be made whether the potential decrease in accuracy from using a transparent classifier is an acceptable sacrifice for increased transparency. Such a decision is fairly subjective and thus generally left to the researcher's preference.

4.1.3 Biofeedback in non-rehabilitation settings

While data fusion primarily is the process of interpreting psychophysiological responses, biofeedback is the act of acting on psychophysiological responses in order to make changes to the environment. These changes then again affect psychophysiological responses. Psychophysiological biofeedback is not a new idea by any means; for instance, a 1996 review [40] discusses several attempts to use psychophysiological measurements to control the level of automation in a task. Work on psychophysiological feedback has primarily focused on three separate purposes. The first (described in section 4.1.3.1) is adaptive automation: making a task easier for the user by providing automated assistance when necessary. The second (described in section 4.1.3.2) is game difficulty adjustment: making a game easier or harder for the user in order to provide an appropriate challenge. The third (described in section 4.1.3.3) is the adjustment of the audio or visual properties of an application that the user is interacting with in order to make it more pleasant and attractive for the user or to evoke a certain other mood. Of course, psychophysiological biofeedback has also been used for other purposes, and these are described in section 4.1.3.4.

Complex data fusion is not strictly necessary for biofeedback. In several studies described in this section, biofeedback is based on simple rules such as “make task easier if skin conductance is above threshold” [156, 157]. In those studies, data fusion is implicitly included in the design of the feedback rules; a person who manually sets such rules should already know how autonomic nervous system responses are influenced by different psychological states.

4.1.3.1 Adaptive automation

Though the majority of work on adaptive automation through psychophysiology has focused on electroencephalography, some studies have also incorporated autonomic nervous system responses, either by themselves or in combination with other measurements. Four examples of adaptive automation using autonomic nervous system responses are Wilson and Russell [140], Ting et al. [148], Prinzel et al. [158] and Liao et al. [159].

The first three take a similar approach: automated assistance is enabled when the user's level of stress or workload is high and disabled otherwise. A HRV threshold is used to enable and disable assistance in Prinzel et al. [158]. A Bayesian network is used for fusion of autonomic nervous system responses and video in Liao et al. [159] while ANNs are used with electroencephalography, electrooculography and autonomic nervous system responses by Wilson and Russell [140]. The fourth study, by Ting et al. [148], differs slightly in that different automation levels are available; i.e. the control for automation is not only an on/off switch. The level of automation is determined by fusing features derived from the ECG and electroencephalogram using fuzzy logic.

4.1.3.2 Game difficulty adjustment

Multiple studies have used autonomic nervous system responses to adjust the parameters of a computer game in order to make it easier or harder for the subject. The level of data fusion in these games differs strongly, from no data fusion to complex data fusion.

Looking first at examples of game difficulty adjustment based on only one physiological measurement, Bersak et al. [160] created a racing computer game where the speed of the car is inversely proportional to the value of the user's skin conductance: the lower the skin conductance, the faster the car. Nenonen et al. [161] used heart rate to affect the difficulty of a biathlon computer game, though it is questionable whether changes in heart rate are caused by psychological factors. In their game, high heart rate results in fast skiing, but inaccurate shooting, and vice-versa.

Moving on to studies combining multiple physiological measurements, Toups et al. [142] used skin conductance and electromyography to increase or decrease the activity level of enemies in a computer game, though data fusion was simply performed as a linear sum of individual normalized features. Dekker and Champion [156] changed the player's movement speed, visibility to enemies and the damage of his/her weapons in a first-person shooter game based on both heart rate and skin conductance. Similarly, Kuikkaniemi et al. [162] controlled the player's movement speed and firing accuracy in a first-person shooter using heart rate and skin conductance, though no data fusion was performed. Haarmann et al. [57] combined heart rate and skin conductance in a flight simulator. Manually set thresholds were used on the features to determine how aroused the subject was, and turbulence was turned on and off in the flight simulator depending on the level of arousal. Liu et al. [45] used a classification tree on multiple physiological signals to estimate the level of anxiety and then used both task performance and anxiety to control the difficulty of a game of Pong.

A final interesting example that is not strictly a computer game, but rather a human-robot interaction system, is a study where children need to throw baskets through a basketball hoop controlled by a robotic arm [18]. The hoop is constantly moved in different directions, with the speed and direction of movement changed to maximize the child's enjoyment of the game. The child's level of enjoyment during the game is determined by using SVMs to fuse multiple psychophysiological features. Furthermore, the robotic arm gradually adapts to the current subject by changing the biofeedback rules. Since there is no guarantee that two users will respond to a particular action in the same way, the arm learns the subject's preferences through reinforcement learning, which learns by trying certain actions and noting the subject's response. Given enough time to try different actions, the system learns what action is likely to lead to a certain response for that subject.

4.1.3.3 Adjustment of audiovisual features

Unlike adaptive automation, which has been extensively applied to critical situations such as flight, adjustment of audiovisual features has primarily been explored within the context of multimedia applications, computer games and virtual reality. Here, the purpose is to evoke a certain mood in the user using a feature of the environment or to have the environment reflect the user's current mood.

An example of environments that try to match the subject's mood is described by Wang et al. [163]: an online chatting interface where the colour and shape of the text changes to match the user's skin conductance. Dekker and Champion [156] directly map the ambient volume and shading of the environment in a first-person shooter to heart rate and skin conductance, without any data fusion. In [70], ANNs are used to classify the subject's mood, and appropriate wallpaper is displayed on the computer background. Groenegress et al. [157] describe a virtual environment with an avatar whose respiratory frequency matches the user's, and who taps his foot at a frequency proportional to the user's skin conductance.

One way of trying to guide a person into a desired mood is by playing music that evokes specific emotions. Oliver and Kreger-Stickles [164] proposed a music player that combines both physiological features and body movement to suggest songs from a playlist, though this does not necessarily include psychological factors since autonomic nervous system responses in their study are strongly affected by physical activity. Janssen et al. [127] also suggested a music player that combines skin conductance and skin temperature using a Bayesian network in order to suggest songs. Liu et al. [165] attempt to control heart rate around a certain threshold by playing appropriate music, though their heart rate sensor is embedded in a seat beneath the subject and is thus fairly nonstandard.

A similar approach to the music recommendation systems outlined above is a content delivery system which classifies the user's autonomic and central nervous system responses using kNN and SVMs [120]. It then suggests content (different documents) that would be appropriate in that mood.

Finally, a system by Grigore et al. [143] tries to help the subject relax by adjusting the level of ambient light in a room. A simple weighted sum of different heart rate and skin conductance features is used to estimate the subject's current state.

4.1.3.4 Other

Three studies should be mentioned which do not quite fit into any of the previous three subsections. The first is a study where a mobile robot performs tasks in the environment while monitoring a human's level of anxiety [17]. The level of anxiety is calculated from heart rate, skin conductance and the electromyogram using fuzzy logic. If anxiety exceeds a certain threshold, the robot ceases its normal operations and queries the human whether he or she requires assistance.

The second is an interesting application (though of questionable practical value) where a computer monitors the user's engagement level through a combination of skin conductance and nonphysiological signals [166]. If the user is not focused on working with the computer, the computer decreases the microprocessor speed in order to save energy.

The third is a game that does not monitor autonomic nervous system responses to stimuli, but rather requires the player to consciously control their skin conductance, temperature and pulse in order to dodge obstacles in the game [167]. No data fusion is performed, and it is uncertain whether the game actually falls within the domain of psychophysiology.

4.1.4 Our approach

Having performed a thorough review of the psychophysiological literature, a number of well-established dimension reduction and classification methods were first selected for implementation in the dissertation. These methods are described in section 4.1.4.1. Then, another method was implemented: adaptive discriminant analysis, which was previously unknown outside electroencephalography. Since it can learn online, adaptive discriminant analysis has an advantage over established methods. It is described in section 4.1.4.2. Finally, a brief overview of the process of implementing data fusion and biocooperative control is given in section 4.1.4.3.

4.1.4.1 Established methods

Having examined the different algorithms and methods for dimension reduction, classification, estimation and biofeedback, a few methods can be selected for implementation. As was discussed in section 4.1.2.3, classification is the best-established data fusion approach in psychophysiological literature and is also easier to validate using questionnaires or independent observers than estimation. Since most classification methods are not difficult to implement, it is possible to implement several different ones and test them on the training data set using cross-validation. The following standard classifiers were implemented:

- LDA,
- QDA,
- diagonal LDA (a type of naïve Bayes classifier),
- diagonal QDA (also a type of naïve Bayes classifier),
- kNN based on Euclidean distance,
- kNN classification based on Mahalanobis distance,
- classification tree,
- SVM with a radial basis function kernel.

In addition to classification, dimension reduction is also recommended due to the large number of features. It can also be easily tested in crossvalidation together with different classification algorithms. For this dissertation, both PCA and SFFS are used for dimension reduction. Fisher's projection is not used since, for reasons described later, there are only two possible task suitability states (too easy / too hard) in both sections 4.2 and 4.3. In a two-state problem, Fisher's projection can only reduce the number of dimensions to one, rendering further data fusion unnecessary.

All classifiers were thus tested without dimension reduction, with PCA, or with SFFS. Since some of these methods have parameters that need to be set, the optimal values were determined in crossvalidation as follows: a set of possible values was tested on the training data set, and the value that yielded the best classifier accuracy was used on the test data set. The values that were set thusly were as follows:

- PCA: number of principal components (possible values: 2, 3, 5, 7, 9, 11),
- kNN (both Euclidean and Mahalanobis distance): number of considered neighbors (possible values: 1, 3, 5),
- classification tree: the minimum number of data points at a node for that node to be split into two branches (possible values: 5, 10, 25, 50).

For SFFS, the statistical F-value was used as the criterion to add or remove a feature. The threshold F-value to add a feature was 3.5 while the threshold F-value to remove a feature was 3. An exception was made if no features exceeded the threshold F-value to add a feature. This often occurred when only psychophysiological data was entered into the stepwise procedure. In this case, both thresholds were lowered in steps of 0.5 until at least one feature's F-value exceeded the threshold.

In addition to these classifiers, which are all well-established in psychophysiology, a new type of classification that has previously not been used in psychophysiology was also applied: adaptive discriminant analysis, described in the next section.

4.1.4.2 Adaptive discriminant analysis

Since motor rehabilitation is a long-term process, it would be useful for the biocooperative feedback loop to gradually adapt to a particular subject's preferences and thus become more useful. One example of adaptation in psychophysiology is through reinforcement learning [18], which gradually tries different actions and records the subject's (psychological or psychophysiological) response in order to determine the best actions to take within the feedback loop. Another partially adaptive approach is used by Gu et al. [110] and Ting et al. [148], who create a separate set of data fusion rules for each subject using only data previously obtained for that subject.

However, the weakness of these approaches is that they require a long training period before they become useful. Reinforcement learning additionally requires the biocooperative feedback loop to take actions that affect the subject, and these actions are not necessarily beneficial. Thus, it was felt that a better adaptive approach would be to start with a set of general data fusion rules, then make adjustments to these rules as the system obtains information about the new subject. The initial rules would be trained using a prerecorded training data set from multiple subjects. Though psychophysiological responses exhibit great intersubject variability, rules derived from multiple subjects should nonetheless be accurate enough to serve as a starting point.

The process of gradually adapting the data fusion rules, however, is not trivial. In the course of the research covered by this dissertation, Matjaž Mihelj had the idea of improving psychophysiological data fusion by combining methods already established in psychophysiology with Kalman filtering, a well-known technique for state prediction and estimation in data fusion [168]. Kalman filtering is used when the state of a dynamic system is difficult to estimate due to noisy sensor data, approximations in the model describing the system, and external factors that have not been accounted for. The filter uses a model of the system, known inputs to that system, and known measurements of the system to form an estimate of the system's state that is better than the estimate obtained by using any one measurement alone. Given an approximate psychophysiological model, it could thus potentially improve accuracy of psychophysiological data fusion by treating errors in the model as unaccounted-for external factors and treating intersubject variability as noise. For the model, it should be possible to use one of the data fusion methods described in section 4.1.2, as these methods are essentially models that translate physiological measurements to psychological states.

Having decided to combine Kalman filtering with data fusion methods already established in psychophysiology, a literature search was conducted to find any papers from nonpsychophysiological studies that combine Kalman filtering with any of the data fusion methods described in section 4.1.2. This yielded a paper by Vidaurre et al. [26] which combines Kalman filtering with LDA and QDA (section 4.1.2.1.3) in order to adaptively classify electroencephalographic data. In that paper, classifier accuracy was significantly improved by online adaptation. Based on the encouraging results and the fact that the method had been developed for use with physiological measurements, the two adaptive discriminant analysis methods described in the paper (Kalman adaptive linear discriminant analysis, an adaptive LDA, and the Adaptive information matrix, an adaptive QDA) were thus selected for implementation in the dissertation’s biocooperative feedback loop. The use of these methods also affected the study designs; since adaptive discriminant analysis requires multiple data points from each subject within a single longer session, the studies described in section 4.2 and 4.3 also require subjects to perform a task for several short time periods.

The mathematical foundations of Kalman adaptive linear discriminant analysis (KALDA) and the Adaptive information matrix (ADIM) are described in the following two subsections (4.1.4.2.1 and 4.1.4.2.2). These mathematical foundations were established by Vidaurre et al. [26] and are mostly repeated or paraphrased. However, as described in the previous paragraph, a limitation of adaptive discriminant analysis is that it is supervised: the subject's opinion is required to perform the update process. Such a supervised learning approach is obviously inappropriate in practice. If the subject’s opinion regarding task difficulty is available, no automated feedback loop is necessary – the subject's opinion can be taken into account instead. Thus, section 4.1.4.2.3 describes a modification that makes adaptive discriminant analysis unsupervised and more useful in practice.

The adaptive discriminant analysis methods were tested in non-diagonal and diagonal variants as well as in supervised and unsupervised variants for a total of 8 (2x2x2) adaptive classifiers. These 8 adaptive classifiers were tested without dimension reduction, with PCA and with SFFS.

4.1.4.2.1 Kalman adaptive linear discriminant analysis

KALDA is an adaptive version of LDA in which the weights of the discriminant function (b, $\boldsymbol{w}^T$) are recursively updated online using a Kalman filter as new data becomes available. The Kalman gain varies the update coefficient and changes the adaptation speed depending on the properties of the data. As previously mentioned in section 4.1.2.1.3, the LDA equations are as follows:

$$D(\boldsymbol{x}) = b + \boldsymbol{w}^T \cdot \boldsymbol{x} \tag{4.5}$$

$$b = -\boldsymbol{w}^T \cdot \frac{1}{2} \cdot (\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \tag{4.6}$$

$$\boldsymbol{w} = (\boldsymbol{S_1} + \boldsymbol{S_2})^{-1} \cdot (\boldsymbol{\mu_2} - \boldsymbol{\mu_1}) \tag{4.7}$$

$$C(\boldsymbol{x}) = \begin{cases} 1; D(\boldsymbol{x}) < 0 \\ 2; D(\boldsymbol{x}) \geq 0 \end{cases} \tag{4.8}$$

where $\boldsymbol{x}$ is the vector of input features, $D(\boldsymbol{x})$ is the discriminant function, b and $\boldsymbol{w}$ are the weights of $D(\boldsymbol{x})$, $\boldsymbol{S_k}$ is the covariance matrix for class k, $\boldsymbol{\mu_k}$ is the vector of mean feature values for class k, and $C(\boldsymbol{x})$ is the class to which $\boldsymbol{x}$ is assigned.

For KALDA, the above equations are expanded with:

$$\boldsymbol{H}_k = [1, \boldsymbol{x}_k^{\mathrm{T}}] \tag{4.9}$$

$$e_k = y_k - \boldsymbol{H}_k \cdot \widehat{\boldsymbol{w}}_{k-1} \tag{4.10}$$

$$v_k = 1 - UC \tag{4.11}$$

$$Q_k = \boldsymbol{H}_k \cdot \boldsymbol{A}_{k-1} \cdot \boldsymbol{H}_k^{\mathrm{T}} + v_k \tag{4.12}$$

$$\boldsymbol{k}_k = \frac{\boldsymbol{A}_{k-1} \cdot \boldsymbol{H}_k^{\mathrm{T}}}{Q_k} \tag{4.13}$$

$$\widehat{\boldsymbol{w}}_k = \widehat{\boldsymbol{w}}_{k-1} + \boldsymbol{k}_k \cdot e_k \tag{4.14}$$

$$\widetilde{\boldsymbol{A}}_k = \boldsymbol{A}_{k-1} - \boldsymbol{k}_k \cdot \boldsymbol{H}_k \cdot \boldsymbol{A}_{k-1} \tag{4.15}$$

$$\boldsymbol{A}_k = \frac{\mathrm{trace}(\widetilde{\boldsymbol{A}}_k) \cdot UC}{p} + \widetilde{\boldsymbol{A}}_k \tag{4.16}$$

where e_k is the one-step prediction error, y_k is the current class label, $\boldsymbol{x}_k$ is the current input vector, $\widehat{\boldsymbol{w}}_k$ is the state vector ($\widehat{\boldsymbol{w}}_k$ = $[b,\ \boldsymbol{w}^T]$, the estimated weights for the LDA), Q_k is the estimated prediction variance, $\boldsymbol{A}_k$ is the a priori state error correlation matrix, $\widetilde{\boldsymbol{A}}_k$ is an intermediate value needed to compute $\boldsymbol{A}_k$, v_k is the variance of the innovation process, $\boldsymbol{k}_k$ is the Kalman gain, UC is the update coefficient and p is the number of elements of $\widehat{\boldsymbol{w}}_k$. The starting values of $\boldsymbol{A}_0$ and $\widehat{\boldsymbol{w}}_0$ as well as the optimal value of UC are computed from the training data set.

Kalman adaptive discriminant analysis can also be used with diagonal LDA, which assumes that all classes have the same diagonal covariance matrix [131]. For diagonal KALDA, the initial weights of the discriminant function are calculated using a diagonal covariance matrix. The update process is then performed normally.

4.1.4.2.2 Adaptive information matrix

ADIM is an adaptive version of QDA in which the covariance matrices and mean values of the different classes are recursively updated online using a Kalman filter as new data becomes available. The Kalman gain varies the update coefficient and changes the adaptation speed depending on the properties of the data.

Basic QDA can be computed using the Mahalanobis distance (d_i) of the feature vector $\boldsymbol{x}$ to each class *i*:

$$d_i(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu_i})^{\mathrm{T}} \cdot \boldsymbol{S_i}^{-1} \cdot (\boldsymbol{x} - \boldsymbol{\mu_i}) \tag{4.17}$$

$$D(\boldsymbol{x}) = \sqrt{d_1(\boldsymbol{x})} - \sqrt{d_2(\boldsymbol{x})} \tag{4.18}$$

$$C(\boldsymbol{x}) = \begin{cases} 1; D(\boldsymbol{x}) < 0 \\ 2; D(\boldsymbol{x}) \geq 0 \end{cases} \tag{4.19}$$

where $\boldsymbol{S_i}$ is the covariance matrix for class i, $\boldsymbol{\mu_i}$ is the mean value for class i, and $C(\boldsymbol{x})$ is the class to which $\boldsymbol{x}$ is assigned. For ADIM, these three equations are expanded with the following recursive equation:

$$\boldsymbol{S}_{k,i}^{-1} = (1 + UC) \cdot \boldsymbol{S}_{k-1,i}^{-1} - \frac{1+UC}{1-UC+\boldsymbol{x}_{k,i} \cdot \boldsymbol{v}} \cdot \boldsymbol{v} \cdot \boldsymbol{v}^{\mathrm{T}} \tag{4.20}$$

where $\boldsymbol{v} = \boldsymbol{S}_{k-1,i}^{-1} \cdot \boldsymbol{x}_{k,i}$, UC is the update coefficient for the adaptation, $\boldsymbol{x}_{k,i}$ is the current class *i* feature vector, *k* is the current sample and *i* is the current class. The mean vector $\boldsymbol{\mu_i}$ is also needed for the computation of $D(\boldsymbol{x})$ and needs to be estimated. This mean vector was incorporated as additional row and column data to the matrix $\boldsymbol{S}_i$ for its automatic estimation and to avoid an extra algorithm, resulting in an "extended" covariance matrix. The starting matrices $\boldsymbol{S}_{0,i}$ were computed from the training data set.

ADIM can also be used with diagonal QDA, which assumes that all classes have diagonal covariance matrices [131]. For diagonal ADIM, $\boldsymbol{S}_{0,i}$ are calculated as diagonal covariance matrices. The update process is then performed normally.

4.1.4.2.3 Unsupervised adaptive discriminant analysis

From Equations 4.10 and 4.20, it is evident that the recursive Kalman update equations require information about the current class ('too easy' or 'too hard') in order to update the classifier. However, in a real-world application, this information would not be available and the update process would need to be unsupervised – to run without requiring the subject's opinion about the current class. It would thus need to use its own, internal estimate of the current class. For KALDA, this would be done by replacing y_k (the actual current class label) in Equation 4.10 with the estimated current class label. For ADIM, this would be done by replacing i (the actual current class) in Equation 4.20 with the estimated current class.

However, such an approach can also amplify classification errors. If incorrect class estimates are used to recursively update the classifier, the classifier will gradually become worse and more prone to errors. One possible way to address this would be to also generate a measure of how ‘reliable’ the estimate is. The system would then only update the classifier if the estimate was sufficiently reliable.

In LDA, if $D(\boldsymbol{x})$ is greater than zero, the input is classified as class 2, and if the output is equal to or less than zero, the input is classified as class 1. If the absolute value of $D(\boldsymbol{x})$ is very close to zero, it can be assumed that the estimate is unreliable and not perform the recursive classifier update. Similarly, in QDA it can be assumed that the estimate is unreliable if the absolute value of $D(\boldsymbol{x})$ is very close to zero. To examine the possibility of unsupervised adaptation, KALDA and ADIM were modified so that the estimated class label was used in equations 4.10 and 4.20. The classifiers were, however, not recursively updated in each classification step. Rather, the classifiers were only updated in a particular step if the absolute value of $D(\boldsymbol{x})$ (in Equations 4.5 and 4.18) was larger than a certain threshold. This threshold was computed from the training data set. Otherwise, the classifiers were not updated in that step.

With such an unsupervised adaptive fusion method, an expanded biocooperative feedback loop is obtained. It is shown in Figure 4.2.

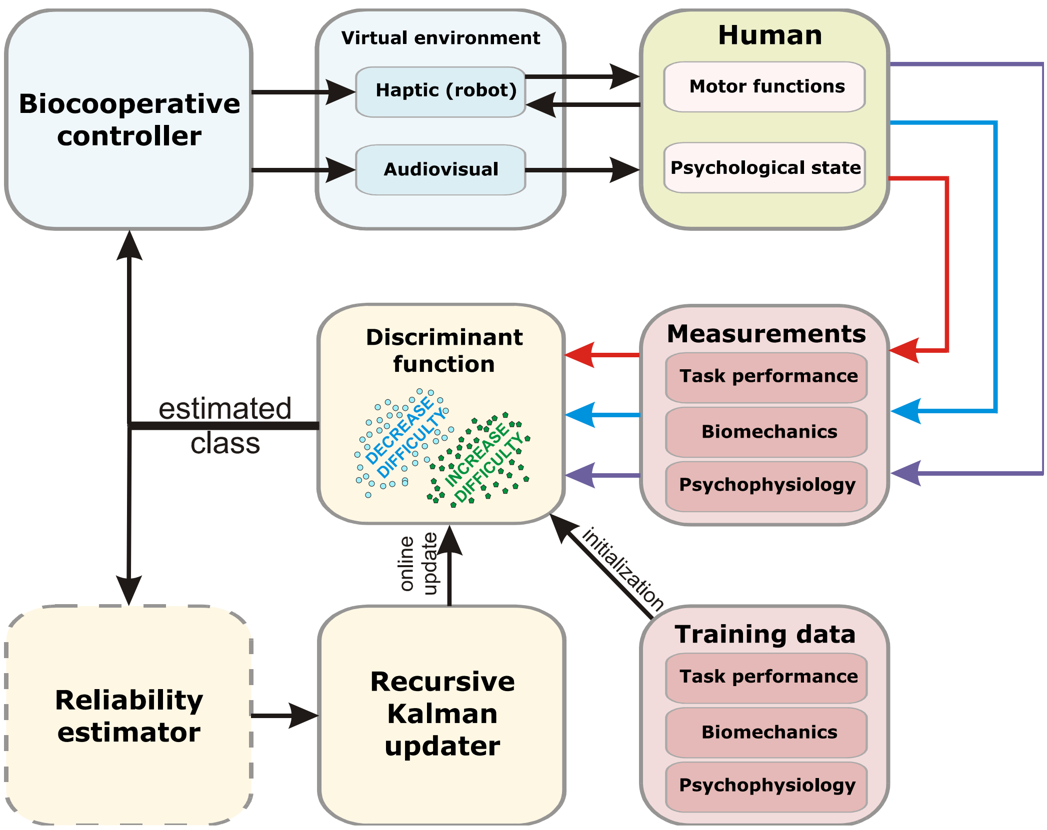


Figure 4.2: A biocooperative feedback loop incorporating adaptive discriminant analysis.

4.1.4.3 Implementation process

The previously established data fusion methods (section 4.1.4.1) as well as adaptive discriminant analysis (section 4.1.4.2) were first tested with healthy subjects in a relatively simple, controlled task with no haptic robot. This allowed us to gauge the effectiveness of different methods in the absence of the effects of physical activity and pathological conditions. This first data fusion step is described in section 4.2. Then, the same data fusion methods were used in the ball-catching task (previously used in section 3.3) with both healthy subjects and hemiparetic patients. They were first used in an open-loop setting (to only estimate task suitability without changing it), then in a biocooperative closed-loop setting (to automatically change task suitability). Data fusion and biocooperative control in the ball-catching task are described in section 4.3.

In both cases, the approach was similar. The subject performed the task for six two-minute periods. After each period, he/she was asked whether he or she would prefer the task to be easier or harder. At the same time, the classifier also estimated whether the task should be easier or harder. The task difficulty was then adjusted according to the subject's opinion (open-loop setting) or the data fusion method (closed-loop setting). Adaptive discriminant analysis expanded this with the Kalman update process. The discriminant function was first initialized with data from the training set. Then, whenever the subject provided his or her opinion, the Kalman filter updated the discriminant function based on the difference between the data fusion estimate and the subject's response. In this way, an adaptive feedback loop which gradually adapts to the current subject is obtained.

Psychophysiological (and other) features were thus mapped directly to the change that needs to be made to task difficulty (make task easier or harder). This is a common approach in psychophysiological feedback [57, 140], but is relatively “black-box” - it does not tell us much about the subject’s actual psychological state. Early on, an alternative implementation was considered: first mapping the psychophysiological features to arousal-valence quadrants. However, this idea was later abandoned since autonomic nervous system measurements are poor at distinguishing different levels of valence (as mentioned, for example, in section 3.3.6.5). Furthermore, in motor rehabilitation the goal does not need to be to distinguish many emotional states; it is sufficient to optimally control the level of engagement and concentration in order to keep the user from becoming either bored or stressed. Thus, it was felt that a relatively simple two-class approach would be better for a first implementation of psychophysiological data fusion.

4.2 Data fusion in a non-rehabilitation setting

While the ultimate goal of the dissertation was to implement data fusion in a biocooperative feedback loop for rehabilitation, the data fusion methods were first tested with healthy subjects in a relatively simple, controlled task with no haptic robot. This allowed us to validate the effectiveness of different methods in the absence of the effects of physical activity and pathological conditions.

4.2.1 Task

The Corsi block-tapping task, a classic psychological experiment (see Berch et al. [27] for a review), was used in this study. Originally, this task consisted of nine blocks laid on a table. The experiment supervisor would tap on the blocks one by one. After the supervisor had finished, the subject had to repeat the sequence. A computerized version of the Corsi task was implemented with nine white blocks laid out on the screen in approximately the same configuration used in the original experiment (Figure 4.3). The blocks briefly darken one by one (the next block darkens half a second after the previous one returns to white). At the end of the sequence, all blocks are white and the subject has to try to repeat the sequence by clicking on the blocks with the mouse. Once the subject has finished clicking (i.e. has clicked as many blocks as the length of the sequence), a “CORRECT” or “FALSE” sign is briefly shown among the blocks before the next sequence begins.

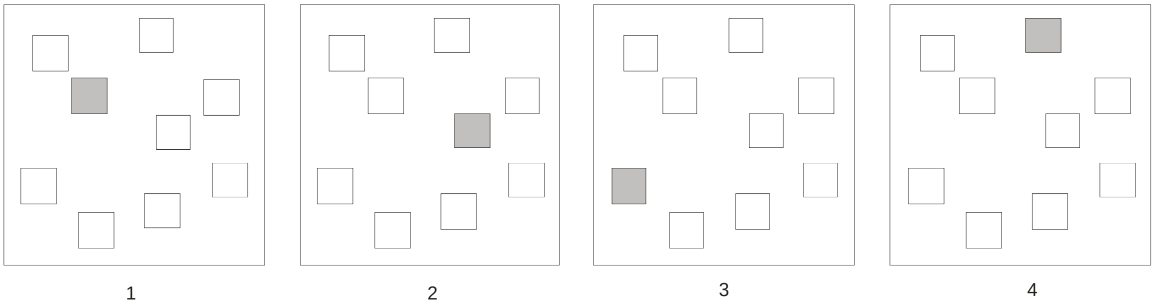


Figure 4.3: A sequence of four blocks in the Corsi task.

The advantage of the Corsi task is that difficulty can be easily varied by changing the length of the sequence that needs to be repeated. Eight possible difficulty levels were implemented, with the lowest difficulty level featuring sequences of two blocks and the highest difficulty level featuring sequences of nine blocks. While five or six blocks present a moderate challenge, a sequence of nine blocks is extremely difficult and a sequence of two is extremely easy.

4.2.2 Measurement protocol

Upon arrival, the task was demonstrated to the subject and the procedure was described. The subjects signed an informed consent form and filled out the BAS/BIS questionnaire. Then, the physiological measurement equipment was attached and turned on. The subject first rested for two minutes, then performed the Corsi task for six two-minute periods (12 minutes total). Within each period, the subject repeated several sequences of blocks in the task at a constant difficulty. At the end of a period, the subject was asked whether he or she would prefer the difficulty of the task to increase or decrease. The 9-point arousal and valence scales of the SAM were also presented. The difficulty of the task then changed randomly by one or two levels in the selected direction. Thus, subjects were able to control whether the difficulty will increase or decrease, but not by how much. Subjects were not given the option to stay at the same difficulty level.

The randomness in the difficulty change was introduced in order to expose subjects to a wider range of difficulty levels. In pretesting, subjects usually stayed within a narrow range of difficulty levels (5-7 blocks per sequence) if there was no random element in the change of difficulty. Additionally, with no randomness, subjects would often alternate between increasing and decreasing difficulty, making classification very easy (a rule saying ‘pick the opposite of what you did last time’ would be sufficient).

After the experiment had been completed and the subject had left the laboratory, the collected psychophysiological and performance data were processed and normalized features were extracted for each time period, including the initial rest period. Since each subject had performed the task for six two-minute periods, six different data points were thus obtained for each subject.

4.2.3 Participants

Twenty undergraduate students from the Faculty of Electrical Engineering at the University of Ljubljana participated in the study. Seventeen were male, three were female. Age range was 21-25, mean 22.6 years, standard deviation 1.3 years.

4.2.4 Fusion methods

In addition to the normalized psychophysiological features described in section 2, four task performance features were extracted for each time period:

- *difficulty level* (2-9),
- *time period* (1 – first time period, 6 – last time period),
- *percentage of correctly repeated sequences*,
- *mean time needed to repeat a sequence* (whether correctly or incorrectly).

Thus, there were three possible data sets to classify: only psychophysiological features, only performance features and both types of features. Each data set consisted of six data points from each subject (one for each two-minute time period) for a total of 120 data points.

Prior to performing classification on this data, correlations were calculated between the SAM and performance/psychophysiology, correlations between performance and psychophysiology, and correlations between the BAS/BIS scales and performance/psychophysiology. This allowed us to validate that the different difficulty levels actually induced different psychological states as well as to determine whether task performance and psychophysiological responses are affected by the subject's innate motivational systems. Spearman correlations were used for all correlations involving the SAM and BAS/BIS (since the data is ordinal) while Pearson correlations were used in other cases. *Valence* and *arousal* were normalized by subtracting the baseline value prior to calculating correlations. Here, it should be especially emphasized that there are six data points for each subject, resulting in correlations where the different data points are not completely independent of each other. The correlation significance and coefficient may thus be higher than is realistic since the number of data points is not 20 (number of subjects), but 120 (number of subjects x 6 periods per subject).

After correlation analysis, binary logistic regression was performed and the Nagelkerke R^2 coefficient [169] was calculated with different types of input data (performance, psychophysiology, BIS/BAS, all) and with the subject's preference (easier/harder) as the binary output. The classic R^2 coefficient describes the proportion of the variability of a data set that is accounted for by a statistical model. The Nagelkerke R^2 coefficient is a pseudo-R^2 coefficient which can be thought of as a generalized equivalent of the classic R^2 coefficient (which is not defined for logistic regression). In this way, it is possible to statistically estimate how well the different types of input data can predict the subject's preference before performing classification.

After these two initial statistical analyses, the data was classified using the classifiers described in section 4.1.4, and the technique of leave-one-out cross-validation was used to evaluate the classifier accuracy. The entire data set was split into the test data (all six data points from one subject) and the training data (all other data points from all other subjects). The classifiers were built using the training data, then validated using the test data. For instance, in the case of LDA, the training data was used to calculate $\boldsymbol{w}$ and b using Equations 4.2 and 4.3. Then, the six data points in the test data set were classified using Equation 4.1 and the calculated $\boldsymbol{w}$ and b. This procedure was repeated as many times as there were subjects, with each subject’s data used as the test data exactly once. The classes assigned to the data points from the different test phases were then used to calculate the accuracy rate.

The final accuracy rate of a classifier was calculated as the number of correctly classified data points divided by the number of all data points across all subjects. For purposes of calculating accuracy rate, all data points are considered to be independent even though there are six from each subject. A data point was considered to be correctly classified if the class assigned to the data point by the classifier (task is too easy or too hard) for that data point was the same as the choice that the subject had made. Given that there are two possible classes, a 50% accuracy rate would correspond to chance (random classification) while 100% would correspond to perfect classification. A 75% accuracy rate, for instance, would mean that 90 out of the total 120 data points (20 subjects with 6 data points per subject) were classified correctly.

Since the data was available offline, the adaptive discriminant analysis methods were tested as follows. The first data point from each subject (i.e. from the first time period of a session) was classified using the initial classifier obtained from the training data (Equations 4.6-4.8). Then, the classifier was recursively updated using this data point and (in the supervised implementations) the choice that the subject had made according to equations 4.9-4.16. The updated classifier was tested on the second data point from each subject, once again updated and so on.

Additionally, it would be useful to know which specific combination of features would be most informative. After classification, SFFS was used to rank the different features. For this purpose, the F-to-enter threshold was lowered to 1.0 and the F-to-remove threshold was lowered to 0.8. While these thresholds are too low for accurate classification, they can still be used to rank features. It should be emphasized again that SFFS does not rank the features independently of each other; in each step, the selected feature is the one that provides the most additional information for classification, taking the contributions of the already selected features into account.

4.2.5 Results

4.2.5.1 Correlation and regression analysis

4.2.5.1.1 Correlations: SAM and performance

Significant Spearman correlations between *valence* and *arousal* on one hand and the different performance features on the other are listed in Table 4.2.

Table 4.2: Significant correlations between SAM and performance.

feature 1	feature 2	p	ρ
valence	arousal	0.016	-0.22
	difficulty level	< 0.001	-0.50
	percentage of correctly repeated sequences	< 0.001	0.58
	mean time needed to repeat a sequence	< 0.001	-0.47
arousal	difficulty level	< 0.001	-0.46
	percentage of correctly repeated sequences	< 0.001	0.36
	mean time needed to repeat a sequence	< 0.001	-0.42

Difficulty level and the *percentage of correctly repeated sequences* were significantly correlated with each other ($r = -0.73$, $p < 0.001$), as were *difficulty level* and *mean time needed to repeat a sequence* ($r = 0.74$, $p < 0.001$).

4.2.5.1.2 Correlations: SAM and psychophysiology

Significant Spearman correlations between *valence* and *arousal* on one hand and the different psychophysiological features on the other are listed in Table 4.3.

Table 4.3: Significant correlations between SAM and psychophysiology.

feature 1	feature 2	p	ρ
valence	SDNN	0.046	-0.19
	LF/HF ratio	0.001	-0.32
	total LF power	0.002	-0.29
	respiratory rate variability	0.003	-0.27
arousal	mean heart rate	0.01	-0.24
	SDNN	0.004	0.26
	RMSSD	< 0.001	0.33
	pNN50	0.014	0.23
	total LF power	0.05	0.18
	mean respiratory rate	< 0.001	-0.33
	mean SCL	0.022	-0.21
	SCR frequency	0.004	0.27
	final skin temperature	< 0.001	0.49

4.2.5.1.3 Correlations: BAS/BIS scales and performance

Significant Spearman correlations were found between the different BAS subscales and task performance features. *BAS Fun Seeking* was significantly correlated with *mean time needed to repeat a sequence* ($\rho = -0.26$, $p = 0.006$) while *BAS Reward Responsiveness* was significantly correlated with difficulty level ($\rho = -0.19$, $p = 0.048$) and *mean time needed to repeat a sequence* ($\rho = -0.28$, $p = 0.002$). There was no correlation between the *BIS* scale and task performance.

There were also significant correlations between the different BAS/BIS scales themselves. *BAS Drive* was significantly correlated with *BAS Fun Seeking* ($\rho = -0.30$, $p = 0.001$), *BAS Reward Responsiveness* ($\rho = -0.21$, $p = 0.025$) and *BIS* ($\rho = -0.37$, $p < 0.001$).

4.2.5.1.4 Correlations: Performance and psychophysiology

Significant Pearson correlations between task performance features and psychophysiological features are listed in Table 4.4.

Table 4.4: Significant correlations between performance and psychophysiology.

feature 1	feature 2	p	r
difficulty level	SDNN	0.003	0.28
	total LF power	< 0.001	0.36
	respiratory rate variability	0.022	0.22
	mean SCL	0.049	0.18
percentage of correctly repeated sequences	SDNN	0.03	-0.20
	LF/HF ratio	0.039	-0.19
	total LF power	0.009	-0.25
	respiratory rate variability	0.047	-0.19
	mean SCL	0.017	-0.22
mean time needed to repeat a sequence	SDNN	< 0.001	0.38
	RMSSD	0.016	0.23
	total HF power	0.02	0.22
	total LF power	< 0.001	0.50
	respiratory rate variability	0.002	0.29
time period	mean heart rate	0.02	-0.22
	pNN50	0.049	0.19
	total HF power	0.039	-0.19

4.2.5.1.5 Correlations: BAS/BIS scales and psychophysiology

Significant Spearman correlations between the different BAS/BIS subscales and psychophysiological features are listed in Table 4.5.

Table 4.5: Significant correlations between BAS/BIS scales and psychophysiology.

feature 1	feature 2	p	ρ
BAS Drive	mean heart rate	0.004	-0.27
	SDNN	< 0.001	0.44
	RMSSD	< 0.001	0.42
	pNN50	< 0.001	0.37
	total HF power	0.001	0.31
	total LF power	0.045	0.19
BAS Fun Seeking	SDNN	< 0.001	-0.34
	RMSSD	0.027	-0.21
	pNN50	0.003	-0.27
	LF/HF ratio	< 0.001	-0.31
	total LF power	0.017	-0.22
	mean SCR amplitude	0.001	0.30
	final skin temperature	0.008	0.25
BAS Reward Responsiveness	mean heart rate	0.004	-0.27
	mean respiratory rate	< 0.001	-0.36
	mean SCL	0.001	-0.31
	SCR frequency	0.02	-0.22
BIS	SDNN	0.014	0.23
	RMSSD	0.001	0.31
	pNN50	< 0.001	0.35
	total HF power	0.012	0.23
	total LF power	< 0.001	0.36
	mean respiratory rate	0.001	-0.31
	respiratory rate variability	0.041	0.19
	mean SCL	0.014	-0.23
	SCR frequency	< 0.001	-0.37

4.2.5.1.6 Logistic regression

Table 4.6 shows the Nagelkerke R^2 coefficient for logistic regression using different types of input data (performance, psychophysiology, BAS/BIS and various combinations) and the subject's preference (easier/harder task) as the binary output.

Table 4.6: Nagelkerke R^2 coefficient for logistic regression using different types of input data and the subject's preference as the binary output.

input data	Nagelkerke R^2
performance	0.409
psychophysiology	0.370
BAS/BIS	0.008
performance + BAS/BIS	0.432
performance + psychophysiology	0.644
performance + psychophysiology + BAS/BIS	0.677

4.2.5.2 Classification

Table 4.7 shows classification results for established classification methods while Table 4.8 shows classification results for different types of adaptive discriminant analysis. The best value is bolded and underlined for each input data type.

Table 4.7: Classification results in a non-rehabilitation setting for methods already established in psychophysiology. Results are shown for different input data types (performance, psychophysiology, both) and for different methods of dimension reduction (none, PCA, SFFS).

	performance			psychophysiology			both		
dimension reduction	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS
LDA	77.5	76.7	76.7	69.2	**<u>75.0</u>**	70.8	68.3	73.3	74.2
QDA	75.8	76.7	77.5	70.0	64.2	**<u>75.0</u>**	70.0	67.5	70.8
diagonal LDA	78.3	76.7	76.7	68.3	**<u>75.0</u>**	67.5	75.0	73.3	77.5
diagonal QDA	76.7	76.7	77.5	61.7	69.2	66.7	70.0	69.2	70.8
kNN (Euclidean)	77.5	77.5	77.5	65.0	65.8	67.5	70.8	65.8	73.3
kNN (Mahalanobis)	74.62	76.7	75.0	70.0	69.2	67.5	68.3	68.3	70.0
tree	**<u>80.8</u>**	67.5	70.0	**<u>75.0</u>**	73.3	71.7	**<u>85.0</u>**	72.5	66.7
SVM	69.2	75.0	77.5	71.7	68.3	70.0	70.8	79.2	79.2

Table 4.8: Classification results in a non-rehabilitation setting for adaptive discriminant analysis. Results are shown separately for supervised and unsupervised methods, for different input data types (performance, psychophysiology, both) and for different methods of dimension reduction (none, PCA, SFFS).

		performance			psychophysiology			both		
	dimension reduction	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS
supervised	KALDA	77.5	76.7	77.5	79.2	66.7	79.2	71.9	69.3	77.2
	ADIM	75.8	76.7	79.2	80.0	78.3	**86.7**	81.6	80.7	78.9
	diagonal KALDA	77.5	76.7	77.5	80.0	77.5	77.5	**86.0**	79.0	81.6
	diagonal ADIM	**80.8**	78.3	77.5	84.2	77.5	**82.5**	84.2	78.1	78.1
unsupervised	KALDA	77.5	77.5	77.5	77.5	73.3	70.0	70.8	76.7	75.0
	ADIM	**78.3**	77.5	**78.3**	80.0	68.3	72.5	80.8	76.7	77.5
	diagonal KALDA	**78.3**	76.7	76.7	79.2	75.0	75.0	**84.2**	77.5	81.7
	diagonal ADIM	76.7	77.5	73.3	**84.2**	74.2	81.7	82.5	75.0	75.0

When ranking features from healthy subjects with a F-to-enter threshold of 1.0 and a F-to-remove threshold of 0.8, SFFS took the following steps:

1. Entered: *percentage of correct answers* (*F* to enter = 45.40),
2. Entered: *respiratory rate variability* (*F* to enter = 13.59),
3. Entered: *mean heart rate* (*F* to enter = 3.31),
4. Entered: *total power in the HF heart rate band* (*F* to enter = 2.00),
5. Entered: *SCR frequency* (*F* to enter = 1.54).
6. Entered: *difficulty level* (*F* to enter = 1.43).

SFFS did not remove any features during the sequence.

If task performance measures were excluded, SFFS took the following steps:

1. Entered: *respiratory rate variability* (*F* to enter = 11.22),
2. Entered: *mean heart rate* (*F* to enter = 12.27),
3. Entered: *mean respiratory rate* (*F* to enter = 7.13),
4. Entered: *total power in the low-frequency heart rate band* (*F* to enter = 2.06),
5. Entered: *SCR frequency* (*F* to enter = 1.71).

SFFS did not remove any features during the sequence.

4.2.6 Discussion

4.2.6.1 Correlation and regression analysis

Results of the SAM confirmed that the different difficulty levels induced different psychological states. As expected, self-reported *valence* decreased as the difficulty increased. Surprisingly, however, self-reported *arousal* also decreased as the difficulty increased. It had been expected to increase since more cognitive effort would be required to remember the longer block sequences. Correlations between the SAM and psychophysiology also produce unexpected results: *mean respiratory rate*, *mean SCL*, and *SCR frequency* all decrease as *arousal* increases while *final skin temperature* increases together with *arousal*. All of these correlations are at odds with existing literature as well as the research performed in section 3.

Correlations between performance and psychophysiology, on the other hand, are as expected. For instance, increasing *difficulty level* also increases *mean SCL* as predicted by the literature and research in section 3. The *percentage of correctly repeated sequences* is negatively correlated with HRV and respiratory rate variability, which is also expected - as cognitive requirements become too high and performance begins to suffer, both HRV and respiratory rate variability increase. This was previously noted for respiratory rate variability in the inverted pendulum task (section 3.2) and for HRV by other authors [170].

Thus, there seems to be a disagreement between the SAM and performance measures. Though this is not supported by quantitative observations, one possibility is that the disagreement was caused by the subjects' response to high difficulty levels. In many cases, the experimenter observed that, when faced with a high difficulty level, the subjects simply chose blocks randomly and reported a low arousal due to (in their own words) "not really trying". Though this is a valid course of action, it emphasized the need for an alternative method of monitoring the subject. The experimenter's opinion was thus 'officially' later used for data fusion in rehabilitation (section 4.3). Another possibility would be to include additional performance features that could provide information about the subject’s behavior. For instance, to determine whether subjects choose blocks

randomly and whether this affects their arousal, *mean answer time* could be split into two features: the mean answer time for correctly answered questions and the mean answer time for incorrectly answered questions.
Results of the BAS/BIS showed that the innate motivational systems indeed affect performance in the task, as subjects with higher scores on the BAS subscales repeated block sequences more quickly. Interestingly, *BAS Reward Responsiveness* was also negatively correlated with *difficulty level*, suggesting that subjects with higher reward responsiveness prefer to stay at lower difficulties. Though there are few rewards in the task (only "CORRECT" or "FALSE" signs) and the correlation is not especially strong, it makes sense; those who focus on reward are likely to prefer lower difficulties where questions are easier to answer correctly.

There were also several correlations between the BAS/BIS and psychophysiology, showing that a subject's innate psychological properties affect his or her psychophysiological response to a specific task. All four of the psychophysiological signals are affected, and different BAS subscales even show notably different correlations (e.g. *BAS Drive* has a +0.44 Spearman correlation coefficient with *SDNN* while *BAS Fun Seeking* has a -0.34 correlation coefficient with *SDNN*). These results emphasize the effects of intersubject variability in both psychological and psychophysiological responses to tasks.

Finally, results of logistic regression analysis suggest that performance and psychophysiology provide a similar amount of information about the subject's preference (Nagelkerke's R^2 was 0.409 for performance and 0.370 for psychophysiology) while a combination of the two provides a larger amount of information (Nagelkerke's R^2 was 0.644 for a combination of performance and psychophysiology). This is an encouraging result that suggests that classification with both data sources should be more accurate than with only one data source.

4.2.6.2 Classification

From Table 4.7, it is evident that all established classification methods give similar results when used on the same type of data. Most methods produce a classification accuracy of

75-80% with performance features and 65-75% with psychophysiological features. In the case of psychophysiology, the methods that perform best all contain some form of dimension reduction: either PCA, SFFS, or the dimension reduction inherently present in the creation of a classification tree. Interestingly, despite dimension reduction, many established methods produce a worse result when all features are used than when only performance features are used. This is most likely because the data dimensionality is large (4 performance features + 13 psychophysiological features) and the available data set is small (20 subjects with 6 feature vectors each), so it is difficult to create a robust classifier with limited data. Dimension reduction does improve results in most cases when both performance and psychophysiological features are included, and the most accurate classifier is again the classification tree (which already includes dimension reduction). Thus, it appears that, at least in this case, the choice of classification method is not quite as important as selecting the most informative features.

Perhaps the exception here is adaptive discriminant analysis. Though it offers no improvement for performance data, adaptive discriminant analysis noticeably improves classification accuracy for psychophysiological data (best accuracy 75.0% for nonadaptive methods and 86.7% for supervised adaptive methods). Supervised adaptive discriminant analysis generally outperforms unsupervised analysis when psychophysiological features or both types of features are used. This is not surprising, as methods that do not use the subject's actual input must generate an estimate of that input, which can hardly be more accurate than the subject's actual input. The difference in accuracy between supervised and supervised methods ranges between very small and quite sizeable, with unsupervised methods even performing slightly better in a few cases. However, it is difficult to say whether this is a case of 'statistical noise' or actual systematic differences between methods that allow better or worse unsupervised adaptation.

Supervised adaptive discriminant analysis outperforms established classification methods when used with psychophysiological features or both types of features, and unsupervised adaptive discriminant analysis still outperforms established methods when used with only psychophysiological features. This confirms its usefulness in psychophysiological data fusion.

The results show that the implemented data fusion methods are capable of estimating the suitability of the task for the current subject from both performance and psychophysiological data when used in a controlled laboratory setting on a relatively homogenous population (undergraduate engineering students). However, the question remains whether they are equally effective in rehabilitation, especially due to the effects of physical activity and pathological states. This is explored in the next section.

4.2.7 Others' contributions

Matjaž Mihelj helped design the experiment protocol while Maja Milavec helped prepare the computerized version of the Corsi task. Matjaž Mihelj also had the idea to combine Kalman filtering with classification, resulting in the use of adaptive discriminant analysis.

4.3 Data fusion and biocooperative control in rehabilitation

Having tested a multitude of classification methods in a controlled setting and obtained good results, it was decided to continue by implementing them in a physically demanding rehabilitation task and use them in a biocooperative feedback loop with both hemiparetic patients and healthy controls. A lower classification accuracy was expected due to the effects of physical activity and pathological conditions, though it was uncertain just how strongly the accuracy would be affected.

4.3.1 Task

Since the ball-catching task had already been used to study stroke patients' psychophysiological responses in section 3.3, it was reused for purposes of data fusion and biocooperative control. This allowed us to build on already obtained knowledge. The basic premise of the task remains the same: In the centre of the screen, there is a table sloped toward the subject. At the beginning of the task, a ball appears at the top of the slope and starts rolling downward. The subject's goal is to catch the ball before it reaches the lower end of the table. Once the ball is grasped, a basket appears above the table. The subject must then place the ball into the basket. Once the ball is dropped into the basket or falls off the table, another ball appears at the top of the table, the basket disappears and the task continues.

For purposes of data fusion, seven different difficulty levels were implemented, with higher levels featuring progressively smaller and faster balls. While the first level is very easy (the ball is very large and requires approximately fifteen seconds to cross the table), the seventh is almost impossible (the ball crosses the table in less than three seconds and has a radius of 1/5 the radius from the first level). The third level is the one that was used in the section 3.3. The ultimate goal of the biocooperative feedback loop was to change the difficulty level so that the subject is optimally challenged.

Though various modes of active robotic support had been offered in section 3.2, only one was used here. If a subject is unable to open or close his or her hand, the robot can automatically grasp the ball as long as the subject's hand is in the correct position. A requirement for this study was that the subjects be able to catch and carry the ball themselves. This was done to minimize intersubject variability caused by different levels of motor ability.

4.3.2 Measurement protocol

The study was divided into two phases: the open-loop phase (where task difficulty is adjusted manually by the subject and experiment supervisor) and the closed-loop phase (where task difficulty is adjusted by the biocooperative controller). The open-loop phase was conducted first, with the goal of obtaining a larger set of data that could be used to train the classifiers needed for a biocooperative controller. It was performed first with healthy subjects, then with hemiparetic patients. After training the biocooperative controller using the open-loop data, the controller was tested in the closed-loop phase with a smaller number of both healthy subjects and hemiparetic patients.

The experiment procedure for both phases was similar. The experiment was conducted in a dedicated room at the University Rehabilitation Institute of the Republic of Slovenia. Three people were present: the subject, experiment supervisor and occupational therapist. Upon arrival, subjects were informed of the purpose and procedure of the experiment.

The description of the open-loop and closed-loop procedures differed as follows: subjects in the open-loop phase were told that task difficulty would be changed according to their wishes while subjects in the closed-loop phase were told that task difficulty would be changed according to a computer program and may not perfectly agree with their wishes. After being informed of the purpose and procedure of the experiment, subjects signed an informed consent form and were seated in front of the robot. One arm (the paretic arm for patients, the right arm for healthy subjects) was strapped into the cuffs and grasping device, and the physiological sensors were attached. The third level of the task was demonstrated, and subjects were allowed to practice it briefly.

After practice, the subject rested for two minutes while baseline physiological measurements were recorded. Then, the subject began performing the task at level 3, 4 or 5 (randomly chosen). After two minutes of performing the task at that difficulty level, the task was paused briefly and the subject was asked whether he or she would prefer the difficulty of the task to increase or decrease. Subjects were not given the option to stay at the same difficulty level. Obviously, it is possible that a subject finds the current difficulty to be ‘just right’ and does not wish to change it. However, only two choices were offered for two reasons. First, this simplifies data fusion by reducing the problem to two choices rather than three. Second, it was found in pretesting that subjects tended to disproportionately keep difficulty at the same level if offered the option, even if visibly frustrated or bored and even if encouraged by the experimenter to change the difficulty. This was likely due to a desire to please the experimenter and therapist by not reporting any dissatisfaction with the system.

Before asking the subject about his or her preference, the experimenter also noted his own opinion of whether difficulty should increase or decrease. A second, more objective opinion of what difficulty would be appropriate for the subject was thus obtained. The issue of the reliability of self-report measures has been previously raised in psychophysiology, and experience from the analysis of the effects of stroke (section 3.3) and data fusion in a non-rehabilitation setting (section 4.2) suggested that the subject’s opinion can be unreliable or influenced by unexpected factors. The opinions of an observer have been suggested as an alternative or validation measure [83]. The experimenter’s opinion was, of course, also subjective to a degree and was based on factors such as the subject's task performance, level of physical exertion, verbal comments and facial expressions.

In the open-loop phase, once the subject had stated his or her preference, the difficulty changed by one or two levels in the direction chosen by the subject. This randomness was introduced in order to expose subjects to a wider range of difficulty levels and create a more robust training data set. If difficulty had always changed by one level, the system would have most likely quickly reached a 'steady state' where difficulty alternated between increasing and decreasing. In the closed-loop phase, the difficulty changed in the direction chosen by the biocooperative controller, and the subject was informed about the actual change in difficulty.

After task difficulty was changed, the task began again at the new difficulty. In total, the subject went through six two-minute periods, with the subject's preference noted and the difficulty changing after each one. After the final task period, the experiment was concluded.

4.3.3 Participants

Twenty-four healthy subjects (20 males, 4 females, age 31.1 ± 10.9 years, age range 21-61) and eleven hemiparetic patients (8 males, 3 females, age 43.2 ± 13.5 years, age range 22-69) participated in the open-loop phase of the study. Ten healthy subjects (9 males, 1 female, age 33.9 ± 12.6 years, age range 22-62) and six hemiparetic patients (4 male, 2 female, age 58.3 ± 6.3 years, age range 54-67) participated in the closed-loop phase of the study. No subject participated in both phases. All patients were undergoing motor rehabilitation at the University Rehabilitation Institute of the Republic of Slovenia and were tested with the FIM [100] and MMSE [99] within a week of the experiment session. All patients scored at least 26 out of a possible 30 on the MMSE and can thus be considered cognitively intact. None of the patients had been diagnosed with visual neglect.

The patients in the open-loop group were hemiparetic as a result of intracerebral hemorrhage (3 subjects), cerebral infarction (4 subjects), or surgery of a neoplasm of the brain (4 subjects). Time since stroke onset or surgery was 216 ± 228 days (minimum 14, maximum 749). Score on the FIM was 103 ± 14 (out of a possible 126). Six suffered from hemiparesis of the left side of the body and five suffered from hemiparesis of the right side of the body.

The patients in the closed-loop group were hemiparetic as a result of subarachnoid hemorrhage (1 subject), intracerebral hemorrhage (2 subjects), cerebral infarction (2 subjects), or surgery of a neoplasm of the brain (1 subject). Time since stroke onset or surgery was 166 ± 34 days (minimum 110, maximum 202). Score on the FIM was 108 ± 5. Three suffered from hemiparesis of the left side of the body and three suffered from hemiparesis of the right side of the body.

A majority of the patients had received secondary stroke prevention drugs (including antihypertensives) prior to participation in the study. Seven patients in the open-loop group and one patient in the closed-loop group had received low doses of psychotropics that had no noticeable side-effects.

With 24 healthy subjects and 11 patients in the open-loop phase, there were thus 144 data points for healthy subjects and 66 data points for patients in the open-loop phase. With 10 healthy subjects and 6 patients in the closed-loop phase, there were thus 60 data points for healthy subjects and 36 data points for patients in the closed-loop phase.

4.3.4 Fusion and control methods

In addition to the normalized psychophysiological features and biomechanical features described in section 2, four task performance features were extracted for each time period:

- *difficulty level* (1-7),
- *time period* (1 – first time period, 6 – last time period),
- *percentage of caught balls*,
- *percentage of balls placed into the basket*.

4.3.4.1 Open-loop cross-validation

Before performing classification, binary logistic regression was performed and the Nagelkerke R^2 coefficient [169] was calculated with different types of input data from the open-loop phase (performance, psychophysiology, biomechanics, all) and with the subject's preference (easier/harder) as the binary output. The classic R^2 coefficient describes the proportion of the variability of a data set that is accounted for by a statistical model. The Nagelkerke R^2 coefficient is a pseudo-R^2 coefficient which can be thought of as a generalized equivalent of the classic R^2 coefficient (which is not defined for logistic regression). In this way, it is possible to statistically estimate how well the different types of input data can predict the subject's preference before performing classification.

After logistic regression, classification was performed with four possible data sets: only performance features, only biomechanical features, only psychophysiological features and all types of features. They were used with the classifiers described in section 4.1.4, and the technique of leave-one-out cross-validation was used to evaluate the classifier accuracy. The entire data set was split into the test data (all six data points from one subject) and the training data (all other data points from all other subjects). The classifiers were built using the training data, then validated using the test data. For instance, in the case of LDA, the training data was used to calculate $\boldsymbol{w}$ and b using Equations 4.2 and 4.3. Then, the six data points in the test data set were classified using Equation 4.1 and the calculated $\boldsymbol{w}$ and b. This procedure was repeated as many times as there were subjects, with each subject's data used as the test data exactly once. The classes assigned to the data points from the different test phases were then used to calculate the accuracy rate.

The final accuracy rate of a classifier was calculated as the number of correctly classified data points divided by the number of all data points across all subjects. For purposes of calculating accuracy rate, all data points are considered to be independent even though there are six from each subject. A data point was considered to be correctly classified if the class assigned to the data point by the classifier (task is too easy or too hard) for that data point was the same as the choice that the subject had made. Given that there are two possible classes, a 50% accuracy rate would correspond to chance (random classification) while 100% would correspond to perfect classification. In the case of healthy subjects, for instance, a 75% accuracy rate would mean that 108 out of the total 144 data points (24 subjects with 6 data points per subject) had been correctly classified.

Since the data was available offline, the adaptive discriminant analysis methods were tested as follows. The first data point from each subject (i.e. from the first time period of a session) was classified using the initial classifier obtained from the training data (Equations 4.6-4.8). Then, the classifier was recursively updated using this data point and (in the supervised implementations) the choice that the subject had made according to equations 4.9-4.16. The updated classifier was tested on the second data point from each subject, once again updated and so on.

Classifiers were first built and cross-validated with data from only healthy subjects, then separately built and cross-validated with data from only hemiparetic subjects. Finally, the classifiers were also built using data from all healthy subjects and tested them on data from hemiparetic subjects. This allowed us to see whether information obtained from healthy subjects can be applied to patients. From previous experience obtained in section 3.3, it was expected that, at the very least, classifiers incorporating psychophysiological features could not be directly transferred from healthy subjects to patients.

Additionally, it would be useful to know which specific combination of features would be most informative. After classification, SFFS was used to rank the different features. For this purpose, the F-to-enter threshold was lowered to 1.0 and the F-to-remove threshold was lowered to 0.8. While these thresholds are too low for accurate classification, they can still be used to rank features. It should be emphasized again that SFFS does not rank the features independent of each other; in each step, the selected feature is the one that provides the most additional information for classification, taking the contributions of the already selected features into account.

4.3.4.2 Closed-loop validation

The classifier that yielded the highest accuracy rate in open-loop cross-validation was selected for implementation in a closed-loop biocooperative controller. Due to expected differences between healthy subjects and patients, two classifiers were trained: one for healthy subjects and one for patients. They were trained using data from the open-loop cross-validation phase.

As mentioned in section 4.3.2, the closed-loop measurement protocol was similar to the open-loop protocol. At the end of each period, the biocooperative controller output whether the task difficulty should be increased or decreased. The output was shown on the screen to the experimenter, but not to the subject. The subject was asked about his or her preference, but task difficulty was changed according to the output of the controller. Accuracy rate was again calculated as the number of matches divided by the number of all estimates made.

The goal of closed-loop validation was not to compare different methods, classifiers or features; this was done with the larger set of data from the open-loop phase. Instead, the goal was to demonstrate online task difficulty adaptation in a biocooperative feedback loop.

4.3.5 Results

4.3.5.1 Open-loop cross-validation – healthy subjects

Table 4.9 shows the Nagelkerke R^2 coefficient for logistic regression using different types of input data (performance, biomechanics, psychophysiology, and various combinations) and the subject's preference (easier/harder task) as the binary output. Results for established classification methods are shown in Table 4.10 while results for adaptive discriminant analysis are shown in Table 4.11. The best value is bolded and underlined for each input data type. The experimenter and subject agreed on whether difficulty should be increased or decreased in 87.6% of all cases.

Table 4.9: Nagelkerke R^2 coefficient for logistic regression using different types of input data and the subject's preference as the binary output. Calculated for healthy subjects in the open-loop phase.

input data	Nagelkerke R^2
performance	0.574
biomechanics	0.500
psychophysiology	0.340
performance + biomechanics	0.615
performance + psychophysiology	0.744
performance + biomechanics + psychophysiology	0.768

Table 4.10: Classification results for healthy subjects in the open-loop phase for methods already established in psychophysiology. Results are shown for different input data types (performance, biomechanics, psychophysiology, all) and for different methods of dimension reduction (none, PCA, SFFS).

	performance			biomechanics			psychophysiology			all		
dimension reduction	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS
LDA	81.9	81.9	81.9	75.0	73.6	73.6	56.9	58.3	56.9	75.7	68.1	**84.7**
QDA	81.9	81.9	81.9	**76.4**	73.0	68.8	56.9	57.6	56.3	68.1	65.3	81.9
diagonal LDA	80.6	81.3	81.9	74.3	75.0	70.9	60.4	58.3	56.3	77.8	68.8	81.9
diagonal QDA	79.9	**82.6**	81.9	75.0	70.1	68.8	60.4	57.6	53.5	79.9	66.0	81.3
kNN (Euclidean)	77.1	77.1	81.3	73.6	73.6	56.9	60.4	60.4	55.6	62.5	60.4	80.6
kNN (Mahalanobis)	81.9	81.9	81.3	70.8	68.8	66.7	56.9	55.6	52.0	66.7	66.7	81.3
tree	79.9	81.9	78.5	70.1	75.0	67.4	**62.5**	54.9	53.5	77.1	63.9	78.5
SVM	81.3	78.5	81.3	73.6	75.0	72.2	56.9	58.3	56.9	71.5	68.1	82.6

Table 4.11: Classification results for healthy subjects in the open-loop phase for adaptive discriminant analysis. Results are shown separately for supervised and unsupervised methods, for different input data types (performance, biomechanics, psychophysiology, all) and for different methods of dimension reduction (none, PCA, SFFS).

		performance			biomechanics			psychophysiology			all		
	dimension reduction	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS
supervised	KALDA	**82.6**	**82.6**	**82.6**	75.7	68.8	73.6	71.5	71.5	60.4	75.7	72.9	**84.7**
supervised	ADIM	81.9	80.6	81.9	77.8	76.4	68.8	66.0	66.0	60.4	80.6	68.8	81.9
supervised	diagonal KALDA	**82.6**	79.2	**82.6**	80.6	73.6	**82.6**	76.4	**77.1**	76.4	83.3	76.4	**84.7**
supervised	diagonal ADIM	79.9	75.7	79.2	75.0	74.3	72.2	66.0	67.4	65.3	79.9	70.2	79.9
unsupervised	KALDA	**82.6**	81.9	81.9	75.0	66.7	73.6	68.8	68.1	56.3	75.7	70.2	**84.7**
unsupervised	ADIM	81.9	79.9	81.9	77.1	71.5	68.8	63.2	62.5	56.3	73.6	65.3	81.9
unsupervised	diagonal KALDA	80.6	74.3	**82.6**	**78.5**	73.6	**78.5**	70.8	**71.5**	70.1	79.9	66.7	83.3
unsupervised	diagonal ADIM	79.9	74.3	79.9	75.0	65.3	71.5	64.6	63.2	62.5	79.9	68.8	81.3

When ranking features from healthy subjects with a F-to-enter threshold of 1.0 and a F-to-remove threshold of 0.8, SFFS took the following steps:

1. Entered: *percentage of caught balls* (*F* to enter = 129.87),
2. Entered: *mean SCR amplitude* (*F* to enter = 3.94),
3. Entered: *pNN50* (*F* to enter = 4.19),
4. Entered: *total power in the LF heart rate band* (*F* to enter = 2.25),
5. Entered: *SCR frequency* (*F* to enter = 1.87),
6. Entered: *mean absolute acceleration* (*F* to enter = 2.82),
7. Entered: *mean frequency of the acceleration signal* (*F* to enter = 2.04),
8. Entered: *respiratory rate variability* (*F* to enter = 1.97),
9. Entered: *mean SCL* (*F* to enter = 1.63),
10. Entered: *LF/HF ratio* (*F* to enter = 3.40),
11. Entered: *percentage of balls placed into the basket* (*F* to enter = 1.36),
12. Entered: *final skin temperature* (*F* to enter = 1.07).

SFFS did not remove any entered features.

As an illustration of how classification accuracy changes with time, Figure 4.4 shows the accuracy rate of three established classification methods as a function of *time period* when used on psychophysiological data or performance data from healthy subjects. Furthermore, as an illustration of how adaptive methods improve accuracy, Figure 4.5 shows a comparison of nonadaptive and supervised adaptive diagonal LDA as a function of *time period* when used on psychophysiological data from healthy subjects. Although both nonadaptive and adaptive diagonal LDA yield the same accuracy rate during the first task period, accuracy is higher for the adaptive approach afterwards. Finally, as an illustration of how the size of the training set improves classification accuracy, Figure 4.6 shows the accuracy rate of the best nonadaptive method as a function of training set size for different types of data from healthy subjects.

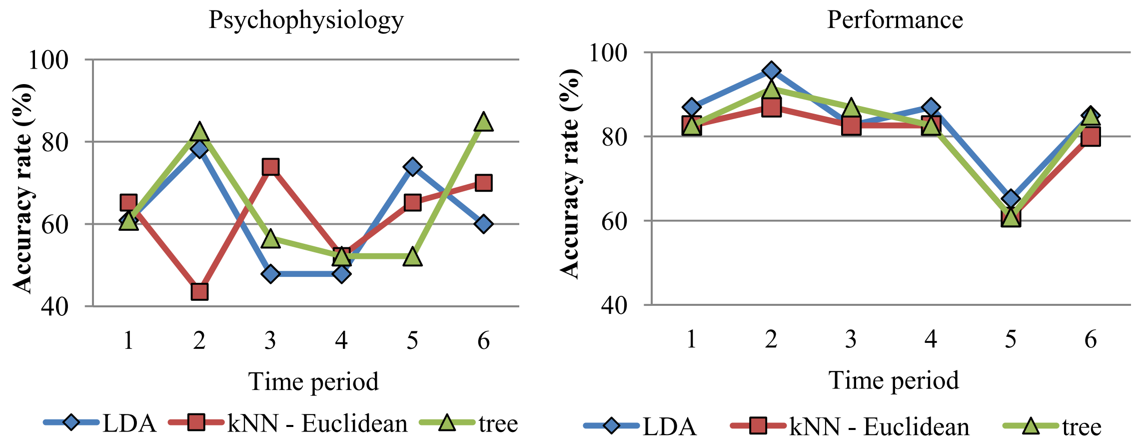


Figure 4.4: Accuracy rate as a function of *time period* for open-loop cross-validation of three established classification methods: LDA, *k*-nearest neighbours with Euclidean distance, and a classification tree. The inputs are psychophysiological (left) or performance (right) features from healthy subjects.

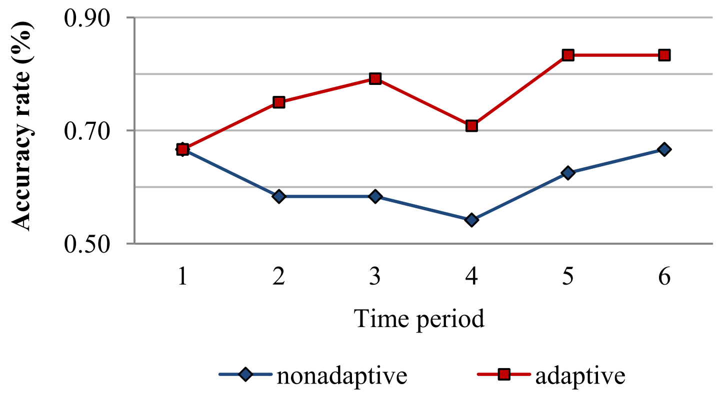


Figure 4.5: Accuracy rate as a function of *time period* for open-loop cross-validation of nonadaptive and supervised adaptive diagonal LDA. The inputs are psychophysiological features from healthy subjects.

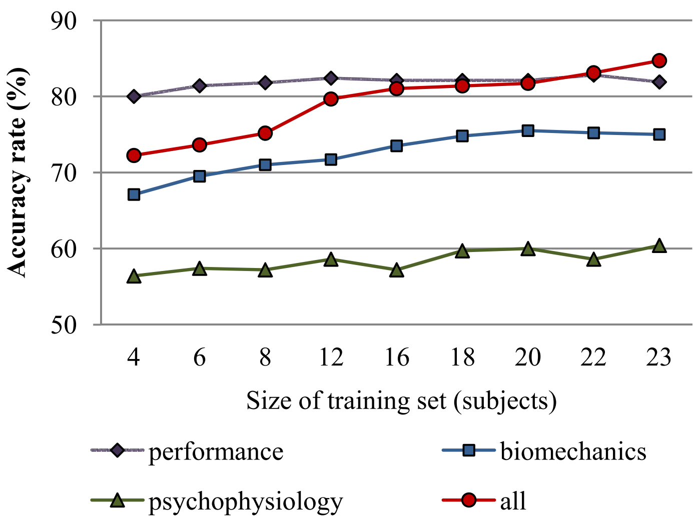


Figure 4.6: Accuracy rate as a function of training set size for different types of input data in open-loop cross-validation. Accuracy rate is taken for the best nonadaptive method. All data are from healthy subjects.

4.3.5.2 Open-loop cross-validation - patients

Table 4.12 shows the Nagelkerke R^2 coefficient for logistic regression using different types of input data (performance, biomechanics, psychophysiology, and various combinations) and the subject's preference (easier/harder task) as the binary output.

Table 4.12: Nagelkerke R^2 coefficient for logistic regression using different types of input data and the subject's preference as the binary output. Calculated for patients in the open-loop phase.

input data	Nagelkerke R^2
performance	0.722
biomechanics	0.672
psychophysiology	0.527
performance + biomechanics	0.921
performance + psychophysiology	0.974
performance + biomechanics + psychophysiology	0.975

Classifiers were first built and cross-validated with data from only hemiparetic subjects. Results for established classification methods are shown in Table 4.13. Results for adaptive discriminant analysis are shown in Table 4.14. The best value is bolded and underlined for each input data type. The experimenter and subject agreed on whether difficulty should be increased or decreased in 97.0% of all cases.

Table 4.13: Classification results for hemiparetic patients in the open-loop phase for methods already established in psychophysiology. Results are shown for different input data types (performance, biomechanics, psychophysiology, all) and for different methods of dimension reduction (none, PCA, SFFS).

	performance			biomechanics			psychophysiology			all		
dimension reduction	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS
LDA	81.8	81.8	81.8	75.8	65.2	75.8	54.5	56.1	60.6	75.8	83.3	**89.4**
QDA	80.3	80.3	**83.3**	66.7	62.1	75.8	56.1	57.6	57.6	62.1	63.6	86.4
diagonal LDA	81.8	**83.3**	81.8	71.2	63.6	71.2	60.6	56.1	57.6	75.8	83.3	83.3
diagonal QDA	**83.3**	80.3	**83.3**	66.7	63.6	69.7	56.1	56.1	56.1	81.8	77.3	83.3
kNN (Euclidean)	77.3	77.3	**83.3**	60.6	60.6	66.7	57.6	57.6	57.6	57.6	62.1	75.8
kNN (Mahalanobis)	80.3	80.3	81.8	66.7	65.2	69.7	53.0	50.0	**63.6**	57.6	63.6	80.3
tree	78.8	**83.3**	81.8	75.8	60.6	**78.8**	60.6	54.5	53.0	74.2	66.7	81.8
SVM	78.8	77.3	**83.3**	68.2	56.1	77.3	56.1	60.6	57.6	56.1	56.1	81.8

Table 4.14: Classification results for hemiparetic patients in the open-loop phase for adaptive discriminant analysis. Results are shown separately for supervised and unsupervised methods, for different input data types (performance, biomechanics, psychophysiology, all) and for different methods of dimension reduction (none, PCA, SFFS).

		performance			biomechanics			psychophysiology			all		
	dimension reduction	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS
supervised	KALDA	81.8	81.8	81.8	**75.8**	72.7	**75.8**	68.2	65.2	68.2	75.8	60.6	**89.4**
	ADIM	80.3	80.3	**83.3**	72.7	71.2	**75.8**	66.7	63.6	66.7	62.1	65.2	86.4
	diagonal KALDA	81.8	78.8	81.8	71.2	74.2	**75.8**	68.2	**77.3**	69.7	75.8	80.3	81.8
	diagonal ADIM	**83.3**	75.8	**83.3**	74.2	74.2	**75.8**	68.2	74.2	69.7	81.8	77.3	86.4
unsupervised	KALDA	81.8	81.8	81.8	**75.8**	68.2	**75.8**	63.6	63.6	68.2	75.8	68.2	**89.4**
	ADIM	80.3	80.3	**83.3**	71.2	65.2	**75.8**	62.1	62.1	63.6	62.1	63.6	86.4
	diagonal KALDA	81.8	77.3	81.8	71.2	71.2	72.7	65.2	**71.2**	65.2	75.8	77.3	81.8
	diagonal ADIM	**83.3**	75.8	**83.3**	72.7	69.7	69.7	63.6	69.7	65.2	81.8	77.3	83.3

When ranking features from hemiparetic patients with a F-to-enter threshold of 1.0 and a F-to-remove threshold of 0.8, SFFS took the following steps:

1. Entered: *percentage of balls placed into the basket* (*F* to enter = 100.55),
2. Entered: *respiratory rate variability* (*F* to enter = 3.71),
3. Entered: *total power in the high-frequency heart rate band* (*F* to enter = 7.73),
4. Entered: *RMSSD* (*F* to enter = 5.02),
5. Entered: *final skin temperature* (*F* to enter = 2.82),
6. Entered: *mean absolute force* (*F* to enter = 1.41),
7. Entered: *difficulty level* (*F* to enter = 2.08),
8. Entered: *mean respiratory rate* (*F* to enter = 1.56),
9. Entered: *mean heart rate* (*F* to enter = 1.22),
10. Entered: *SDNN* (*F* to enter = 2.58),
11. Removed: *RMSSD* (*F* to remove = 0.23),
12. Entered: *mean absolute acceleration* (*F* to enter = 1.95),
13. Removed: *mean absolute force* (*F* to remove = 0.20).

Classifiers were also trained with data from healthy subjects, then tested on patient data. In this case, results for established classification methods are shown in Table 4.15 while results for adaptive discriminant analysis are shown in Table 4.16.

Table 4.15: Classification results for methods already established in psychophysiology when classifiers are trained with data from healthy subjects, then tested on patient data. Results are shown for different input data types (performance, biomechanics, psychophysiology, all) and for different methods of dimension reduction (none, PCA, SFFS).

	performance			biomechanics			psychophysiology			all		
dimension reduction	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS
LDA	78.8	78.8	83.3	63.6	65.2	62.1	54.6	54.6	53.0	71.2	48.5	81.8
QDA	80.3	80.3	78.8	65.2	54.6	60.6	45.5	50.0	59.1	60.6	50.0	78.8
diagonal LDA	81.8	83.3	78.8	65.2	68.2	69.7	50.0	54.6	59.1	69.7	50.0	78.8
diagonal QDA	80.3	80.3	78.8	57.6	57.6	60.6	47.0	51.5	60.6	63.6	50.0	78.8
kNN (Euclidean)	80.3	83.3	84.9	63.6	68.2	54.6	43.9	47.0	56.1	50.0	48.5	81.8
kNN (Mahalanobis)	84.9	84.9	81.8	62.1	62.1	62.1	48.5	50.0	59.1	57.6	51.5	81.8
tree	80.3	75.8	<u>**86.4**</u>	<u>**71.2**</u>	<u>**71.2**</u>	54.6	51.5	54.6	56.1	74.2	56.1	<u>**86.4**</u>
SVM	81.8	75.8	84.9	66.7	68.2	66.7	56.1	53.0	<u>**62.1**</u>	56.1	45.5	84.9

Table 4.16: Classification results for adaptive discriminant analysis when classifiers are trained with data from healthy subjects, then tested on patient data. Results are shown separately for supervised and unsupervised methods, for different input data types (performance, biomechanics, psychophysiology, all) and for different methods of dimension reduction (none, PCA, SFFS).

		performance			biomechanics			psychophysiology			all		
	dimension reduction	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS	none	PCA	SFFS
supervised	KALDA	78.8	75.8	**83.3**	68.2	63.6	71.2	**66.7**	59.1	56.1	68.2	54.6	**81.8**
	ADIM	80.3	80.3	78.8	63.6	57.6	69.7	63.6	54.6	59.1	62.1	54.6	78.8
	diagonal KALDA	78.8	81.8	81.8	63.6	68.2	71.2	60.6	60.6	57.6	63.6	57.6	78.8
	diagonal ADIM	80.3	80.3	80.3	65.2	66.7	65.2	62.1	60.6	60.6	63.6	57.6	78.8
unsupervised	KALDA	78.8	75.8	**83.3**	62.1	63.6	60.6	50.0	51.5	**54.6**	65.2	45.5	**81.8**
	ADIM	78.8	80.3	78.8	60.6	57.6	69.7	48.5	47.0	**54.6**	57.6	47.0	75.8
	diagonal KALDA	78.8	80.3	81.8	59.1	54.6	62.1	45.5	47.0	53.0	63.6	50.0	78.8
	diagonal ADIM	80.3	80.3	80.3	65.2	54.6	63.6	48.5	50.0	**54.6**	60.6	48.5	78.8

4.3.5.3 Closed-loop validation

As seen in the open-loop phase, the most accurate type of classifier was (adaptive or nonadaptive) LDA with all data types and SFFS. Thus, LDA with SFFS was chosen for closed-loop testing in both healthy subjects and patients. For healthy subjects, three features exceeded the SFFS F-to-enter threshold and were thus included: the *percentage of caught balls*, *mean SCR amplitude* and *pNN50*. For patients, four features exceeded the SFFS F-to-enter threshold and were thus included: *percentage of balls placed into the basket*, *respiratory rate variability, total power in the high-frequency heart rate band* and *RMSSD*. In closed-loop testing, this approach yielded an accuracy rate of 88.3% for healthy subjects and 88.9% for patients. The experimenter and subject agreed on whether difficulty should be increased or decreased in 91.7% of all cases for healthy subjects and in 97.2% of all cases for patients.

In a follow-up offline analysis, the closed-loop data was also passed through the most accurate classifier based only on performance data (also trained using data from the open-loop phase). Performance data yielded an accuracy rate of 86.7% for healthy subjects and 83.3% for patients. For an example of psychophysiology increasing accuracy, see Figure 4.7. Additionally, in a second follow-up offline analysis, the closed-loop data was passed through supervised KALDA with SFFS. However, the adaptive version yielded the same accuracy rates as the nonadaptive version for both healthy subjects and patients.

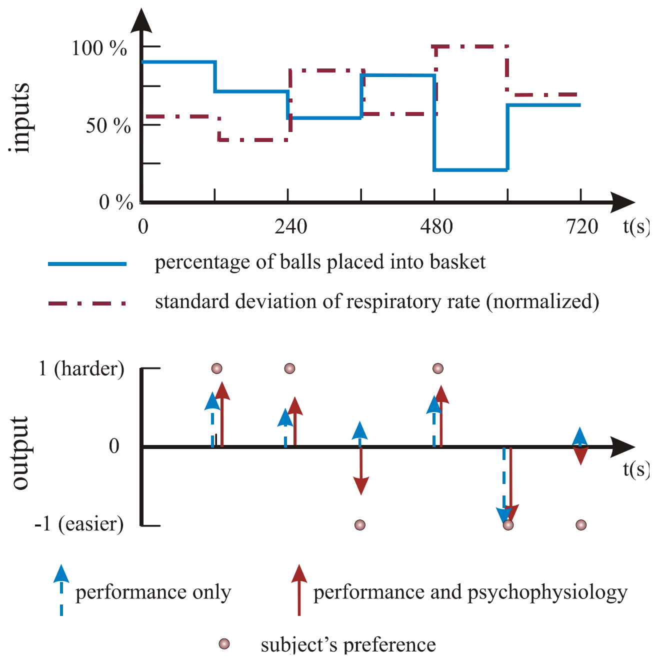


Figure 4.7: One hemiparetic patient in the closed-loop phase: two input features (one performance, one psychophysiological), the output of LDA, and the subject's preferences. High performance and a low *respiratory rate variability* (even, regular breathing) indicate an easy task. For the first, second, fourth and fifth task periods, task performance would have been sufficient to change the difficulty. During the third period, task performance is moderately high, but breathing becomes very uneven, indicating stress. If only task performance had been taken into account in this case, the incorrect decision would have been made (the patient was successful at the task, but was stressed and wanted difficulty to decrease). During the last period, both performance and psychophysiology are unreliable, and the patient stated that he would most prefer difficulty to stay the same.

4.3.6 Discussion

4.3.6.1 The usefulness of different data types

Results of logistic regression, conducted prior to performing open-loop cross-validation of the different classifiers, already suggested that the most useful type of data would be task performance (Nagelkerke R^2 of 0.574 for healthy subjects and 0.722 for patients), with biomechanics providing less information (Nagelkerke R^2 of 0.500 for healthy subjects and 0.672 for patients) and psychophysiology providing the least information (Nagelkerke R^2 of 0.340 for healthy subjects and 0.527 for patients). It also suggested that combining different data sources, particularly performance and psychophysiology (Nagelkerke R^2 of 0.744 for healthy subjects and 0.974 for patients), would yield more accurate information about the subject's preference than using a single data source. Furthermore, it suggested that predicting patients' preferences would be easier than predicting healthy subjects' preferences.

As suggested by logistic regression, task performance was clearly the most accurate type of data in open-loop cross-validation, with an accuracy rate of over 80% for both healthy subjects and patients. Biomechanical data similarly had an accuracy rate of over 75% for both healthy subjects and patients. Psychophysiological measurements, on the other hand, yielded noticeably worse results. Nonadaptive methods yielded an accuracy rate of 62.5% for healthy subjects and 63.6% for patients. Supervised adaptive methods were able to improve the accuracy rate of psychophysiological measurements to 77.1% for healthy subjects and 77.3% for patients, but these results are still worse than results for task performance. This suggests that psychophysiological measurements by themselves are not reliable in a biocooperative feedback loop for upper extremity rehabilitation.

Combining multiple types of data often actually lowers the overall accuracy rate. As in section 4.2, this is most likely due to the small sample size problem: with a large number of features (26 in total) and a limited training set, it is difficult to find an accurate classifier. This is especially noticeable in Figure 4.6, where the accuracy rate when using all data types together rises steadily as the size of the training set increases. Using SFFS can improve the accuracy rate when using all types of data noticeably, and it appears to outperform PCA. This is not surprising, as PCA is unsupervised and the extracted principal components may not be related to the classification problem (as described in section 4.1.1.2). Nonetheless, PCA has its uses: in the case of adaptive methods, the highest accuracy rates are obtained using PCA rather than SFFS.

In open-loop cross-validation, combining multiple data types using SFFS increases the accuracy rate from 82.6% (performance data only) to 84.7% for healthy subjects and from 83.3% to 89.4% for patients. In closed-loop validation, combining multiple types of data increases the accuracy rate from 86.7% (performance data only) to 88.3% for healthy subjects and from 83.3% to 88.9% for patients. While SFFS identifies task performance as the most important source of data, several psychophysiological features are also selected, suggesting that they can provide some supplementary information as predicted by logistic regression.

Of course, an accuracy rate of 100% is most likely unrealistic. In a number of cases, subjects were uncertain how they wanted the difficulty to change (if at all), and responded with comments such as ‘I don’t know, either is fine’. In such a case, the best choice may have been not to change the task difficulty at all. During the closed-loop phase, it was observed (though only on a subjective, qualitative level) that the output of LDA ($D(\boldsymbol{x})$ in Equation 4.5) tended to be closer to zero in such cases as well, suggesting that the output of the discriminant function was also ‘uncertain’ in a way.

The reliability of the subject’s opinion was also taken into account by comparing the subject’s opinion to the experimenter’s opinion. These matched in over 90% of cases, with most disagreements being due to either the subject wanting to try a difficulty level that he/she had never encountered before or the subject being tired despite doing well. Thus, the relatively poor accuracy of psychophysiological measurements cannot be (only) due to subjects’ inaccurate opinions.

To summarize the above paragraphs: if measures of task performance are available and relevant, psychophysiological measurements are probably unnecessary. Designers could take this into account by creating exercises where performance is easy to quantify. However, psychophysiology may prove useful when task performance and biomechanical measures are not readily obtainable or not necessarily connected to the subject’s psychological state. It could also be used to change elements other than the difficulty of the task - for instance, the appearance of a virtual scenario.

Finally, a word on biomechanical measurements: the first five features selected by SFFS include only task performance and psychophysiology. This does not mean that biomechanical measurements are useless. Before any features are included, the F-value (criterion for inclusion) of biomechanical features is higher than that of psychophysiological features. However, once the first feature (a task performance feature for both healthy subjects and patients) has been taken into account, biomechanical features offer less additional information than psychophysiological ones. Similarly, the Nagelkerke R^2-coefficient is higher for biomechanical data features than for psychophysiological features, but the Nagelkerke R^2-coefficient for a combination of performance and psychophysiology is higher than the Nagelkerke R^2-coefficient for a combination of performance of biomechanics. This again suggests that psychophysiological measurements offer information that cannot be obtained from forces and movements.

4.3.6.2 Comparison of classifiers

From Tables 4.10 and 4.13, it is evident that there is no clear 'best' method among the classifiers already established in psychophysiology. All give similar results when used on the same type of data. While some methods give worse results than others with a specific set of data (e.g. *k*-nearest neighbours with Euclidean distance has a worse accuracy rate than other methods when used with biomechanical data from patients), this does not generalize across data types or even to the other type of subjects (e.g. *k*-nearest neighbours with Euclidean distance has a similar accuracy rate as other methods when used with biomechanical data from healthy subjects).

As in section 4.2, the methods that perform best on psychophysiological data all contain some form of dimension reduction: either PCA, SFFS, or the dimension reduction inherently present in the creation of a classification tree. However, due to the low accuracy rates when using psychophysiological data alone, the difference is fairly small and may not be important in practice. Dimension reduction, however, is crucial when combining different types of data. As mentioned, this is most likely due to the small sample size problem: with a large number of features and a limited training set, it is difficult to find an accurate classifier.

In many cases, several different methods produce exactly the same classification accuracy rate. This is especially true for classification of performance data and may seem surprising at first. However, it can be explained by several factors. First of all, given that the accuracy rate is calculated as the percentage of correct classifications out of a limited number of cases (144 for healthy subjects and 60 for patients in the open-loop phase), it cannot take any possible value. For instance, 82.6% in Table 4.10 represents 124 correctly classified data points out of a possible 144. Furthermore, when performance data is included, a single performance feature often contains more information than the other remaining features. For instance, in sections 4.3.5.1 and 4.3.5.2, the first selected feature is a performance feature and has a F-value of over 100. For both healthy subjects and patients, however, the second selected value has a F-value of less than 5. We can see that the majority of relevant information can be obtained from a single feature. It is thus not surprising that all methods that include SFFS give the same accuracy rate since only a few features are included in classification, reducing the problem to a relatively simple one.

If using only performance data, classification is even simpler. SFFS on healthy subjects' performance data is as follows. Before any features are included, the highest-ranked feature is *percentage of caught balls* (F to enter = 129.9). Once this feature has been included, the highest-ranked remaining feature is *difficulty level* (F to enter = 0.5), with *percentage of balls placed into the basket* (F to enter = 0.04) and *time period* (F to enter = 0.00) providing even less additional information. Thus, since practically all relevant information can be obtained from a single feature and since there are only four performance features in total (so that there are no problems due to high data dimensionality), little difference in accuracy is to be expected between different dimension reduction and classification methods in the case of performance data.

A special case among classifiers is adaptive discriminant analysis, covered separately in the next subsection.

4.3.6.3 Adaptive discriminant analysis

In open-loop cross-validation, supervised adaptive discriminant analysis offers no improvement over established classification methods in the case of performance features and only slight improvement in the case of biomechanical features (accuracy rate increases from 76.4% to 82.6% for healthy subjects when adaptive discriminant is used, but classification trees outperform adaptive discriminant analysis for patients). In the case of psychophysiological features, however, supervised adaptive discriminant analysis increases the accuracy rate from 62.5% to 77.1% for healthy subjects and from 63.6% to 77.3% for patients. Unsupervised adaptive discriminant analysis also increases the accuracy rate, though to a lesser degree.

It is currently uncertain why the improvement is greater for psychophysiological features than for other features, but results nonetheless show that the system can gradually adapt itself to a given subject to some degree. Since rehabilitation is usually a long-term process, it would be interesting to see what kind of improvement adaptive methods could provide over multiple sessions.

4.3.6.4 Accuracy as a function of time period

From Figures 4.4 and 4.5, we see that classification accuracy is not the same throughout the experiment. In Figure 4.4b, for instance, there is a noticeable decrease in accuracy rate for all three classifiers in the fifth period of the experiment. Similarly, in Figure 4.5, there is a noticeable decrease in accuracy rate for both classifiers in the fourth period of the experiment. Furthermore, we see that not accuracy rate does not change the same way for all classifiers: in Figure 4.4a, the accuracy rates of the three classifiers change quite differently as a function of time period.

The different accuracy rates of the three classifiers used on psychophysiological data are likely a case of ‘statistical noise’. Since there is a low number of subjects and the classifiers use different rules to classify a large number of features, one classifier may purely randomly exhibit a higher accuracy rate than another in a certain time period. It is less certain why all three classifiers in Figure 4.4b and both classifiers in Figure 4.5 have a lower accuracy rate in one time period than the others. Given the low number of subjects, it may again be caused by statistical noise, but it may be caused by some other, unknown reason (e.g. tiredness).

Finally, Figure 4.5 shows the difference in accuracy rate between nonadaptive and adaptive diagonal LDA. What is surprising is that the difference in accuracy rate does not gradually increase with time as the Kalman filter learns and adapts, but that the difference in accuracy rate is similar across all time periods. Though not shown graphically, similar results were observed for example for nonadaptive and adaptive QDA. One possible explanation is that adaptation is very fast and thus does not increase greatly after the first period. Another possibility is that the state of the patient changes constantly (e.g. increased tiredness, increased experience with the task…), requiring the Kalman filter to constantly adapt in order to maintain the same advantage over nonadaptive discriminant analysis. However, it is impossible to determine the true reason from the data recorded in this study.

4.3.6.5 Differences between healthy subjects and patients

Based on previous studies that have shown weakened psychophysiological responses as a result of stroke and other pathological conditions and experiences from section 3.3, it was expected that fusion of psychophysiological measurements would be less accurate in patients than in healthy subjects. However, this does not appear to be the case; accuracy rates are similar in healthy subjects and patients. Interestingly, accuracy rates are similar for both groups even though the patient group is much smaller.

As Tables 4.15 and 4.16 show, classifiers based on biomechanical or psychophysiological measurements cannot simply be transferred from healthy subjects to patients, as many classifiers suffer a noticeable decrease in accuracy rate. SFFS also selects different features in healthy subjects and patients.

It is easy to understand why results of biomechanical measurements are different between groups: hemiparetic patients, by definition, cannot move their affected limb as well as healthy subjects can. This was evident, for instance, in their response to high difficulty levels. While all healthy subjects reacted to very fast balls by rapidly moving around the virtual table trying to catch the ball, many patients preferred to simply stay in one area of the table and catch only the balls passing through that area.

Psychophysiological measurements are, to some degree, obviously different due to the aforementioned effects of stroke and other pathological conditions. Additionally, for patients, higher task difficulty levels may also be physically demanding since they involve fast movement which the patients may not be physically capable of. This would result in both a physical and cognitive challenge that would evoke strong physiological responses. For healthy subjects, on the other hand, fast movement would not present a physical challenge (since the subjects are physically capable of fast movement) but only a cognitive one (since the movement must be planned more quickly) and thus evoke weaker physiological responses. However, this is primarily speculation. The study in section 3.3, which also includes patients, neither confirms nor denies it since it does not vary physical and cognitive workload in the same task – physical workload is expected to be similar in the physical control task, the virtual rehabilitation task and the harder virtual rehabilitation task. An interesting test would be to perform an experiment similar to the inverted pendulum in section 3.2 (where both cognitive and physical workload are changed as independently as possible in the same task) on hemiparetic patients.

4.3.6.6 Study limitations

In the course of the two data fusion studies presented in section 4, a few limitations became apparent. A few are fairly general and are thus listed in section 5.2 of the overall discussion, but a few are task- or method-specific.

First, the choice of the ball-catching task may not have been optimal for data fusion. Since little psychophysiological work had been done in rehabilitation prior to the beginning of the dissertation, it was not known what type of rehabilitation task would be optimal in a psychophysiological study of rehabilitation. Thus, in section 3.3 a task that had already been developed in our laboratory was used. Since good results were obtained with the task in section 3.3, it was also selected for use in data fusion. However, one component of the task (placing the ball in the basket) does not depend on the difficulty level since the difficulty level only affects the size and speed of the ball. Psychophysiological differences between difficulty levels thus may not have been as large as they would have been if all task components had been affected by the difficulty level, and this may have contributed to the limited usefulness of psychophysiological measurements. Future studies may prefer to focus on a task with only a single component (e.g. only horizontal reaching).

Second, it is impossible to say with absolute confidence that adaptive classifiers better adapt themselves to the individual user's psychophysiological state. Though the results show that adaptation clearly takes place, it is possible that the adaptive classifiers recognize certain other trends not related to the current user. For instance, once users have spent some time performing the task, they are likely to desire higher difficulty levels. Adaptive classifiers could thus learn to increase the chance of choosing 'task is too easy' as time passes. This would not be a negative effect since it would still result in more accurate classification, but we should keep it in mind nonetheless.

Finally, it is difficult to quantitatively validate whether the differences between difficulty levels were sufficiently large to evoke different psychological states and thus cause different psychophysiological responses. In the Corsi task, this had been confirmed by SAM results which showed a significant correlation between *valence* and *difficulty level*, among others. The SAM was not used in this task, primarily since it had not provided reliable results with patients (section 3.3). The subjects' and therapists' verbal comments during the task as well as the experimenter's opinion indicated that the difficulty levels were certainly different enough to induce boredom, frustration or satisfaction. However, it is difficult to assess quantitatively just how different the various difficulty levels felt for the subjects. Thus, it is also impossible to say with absolute certainty that the subjects were optimally challenged; while it is assumed that they were led toward an optimally challenged state by the classifier, future work should make use of questionnaires to validate this assumption.

4.3.7 Others' contributions

Matjaž Mihelj helped design the experiment protocol and also had the idea to combine Kalman filtering with classification, resulting in the use of adaptive discriminant analysis. The ball-catching scenario and the HapticMaster control algorithms were programmed by Jaka Ziherl and Andrej Olenšek. Marko Munih oversaw the study at the Faculty of Electrical Engineering. Nika Goljar, MD, of the University Rehabilitation Institute oversaw the study at the Institute and selected suitable patients. Metka Javh was the occupational therapists who guided the patients during the experiment sessions and ensured their safety.

5 Overall discussion

5.1 The usefulness of psychophysiology in motor rehabilitation

In section 3, it was demonstrated that both physical activity and pathological conditions significantly affect psychophysiological responses. Physical activity partially masks the contribution of psychological states while pathological conditions weaken psychophysiological responses in general. In section 4.3, it was furthermore shown that data fusion in rehabilitation using psychophysiological measurements alone is relatively inaccurate compared to other sources of information that are already available in motor rehabilitation. However, fusion of psychophysiological features yielded higher accuracy in a non-rehabilitation setting. The accuracy rate in a non-rehabilitation setting using only psychophysiological features was 75.0% for the best nonadaptive method and 86.7% for the best supervised adaptive method (section 4.2). On the other hand, the accuracy rate in the ball-catching task using only psychophysiological features was 62.5% for the best nonadaptive method and 77.1% for the best adaptive method (section 4.3, healthy subjects). The difference of 12.5% for nonadaptive methods and 9.6% for adaptive discriminant analysis suggests that the low accuracy rate in the ball-catching task is not the fault of the data fusion methods themselves, although the number of subjects is relatively small (N = ~20 in both sections) and thus cannot guarantee that this difference is significant. Similarly, the experiment protocol is likely not to blame since the same protocol was used in sections 4.2 and 4.3. There are thus three remaining options for the relative inaccuracy of data fusion in rehabilitation: physical activity, pathological conditions or something else.

Pathological conditions apparently did not noticeably affect the effectiveness of data fusion, as classification accuracy was very similar for both healthy subjects and hemiparetic patients. Different features were, however, selected for healthy subjects and patients. This confirms the findings of section 3: that, while psychophysiological responses are generally weakened, they are not all weakened to the same degree. Enough information may remain so that classification accuracy can be similar for healthy subjects and patients. A second possibility is that hemiparetic patients find the task to be physically demanding while healthy subjects do not. Physical activity would be expected to increase with task difficulty, so the physiological effects of physical activity may have made data fusion easier in patients.

Physical activity may have been responsible for the low psychophysiological classification accuracy in rehabilitation, as the task proved to be quite demanding for some subjects. Here, not only the physiological effects of physical activity should be taken into account; there is also the possibility that the measurement process itself is affected by physical activity. As the subject moves around, motion-related artefacts can be recorded by the sensors. For instance, in the case of the ECG, increased noise is to be expected as the cables between the electrodes and the amplifier move and pull on the electrodes. This increased noise could potentially obscure the R-peaks of the ECG itself and can only be avoided by minimizing the movement of the cables. For the skin conductance sensor, it was observed in one test that any damage to the cables can cause motion-related artifacts in the signal that resemble SCRs. Thus, it is important to also keep the sensors in good condition.

Additionally, there is a possible alternate reason for poor data fusion accuracy in the rehabilitation task: that the task itself may not have been optimal for accurate psychophysiological data fusion. As mentioned in section 4.3.6.6, one component of the rehabilitation task (placing the ball in the basket) does not depend on the difficulty level since the difficulty level only affects the size and speed of the ball. Psychophysiological differences between difficulty levels thus may not have been as large as they could have been. However, this weakness would be inherent in most rehabilitation tasks since they are necessarily complex.

Combining psychophysiological measurements with task performance and biomechanics increases classification accuracy in data fusion as long as proper dimension reduction methods are used. This suggests that psychophysiological measurements can provide supplementary information that cannot be gleaned from task performance or biomechanics. The question here is whether the increase in accuracy rate due to psychophysiology is sufficient to justify the increased complexity of the system. If measures of task performance are readily available and relevant, psychophysiological measurements are most likely unnecessary. Designers could take this into account by creating virtual environments in which performance is easy to quantify, although this may be difficult to achieve in non-game scenarios such as activities of daily living. In such cases, psychophysiology could prove useful since task performance measures are often not obtainable or not connected to the subject's psychological state. It could also be used to change elements other than the difficulty of the task - for instance, to change the visual appearance of a scenario or to select the music played.

Nonetheless, despite the discouraging performance of psychophysiological measurements in data fusion and biocoperative feedback in this dissertation, this does not mean that psychophysiological measurements are useless for rehabilitation, and it certainly does not mean that the idea of biocooperative robotics is wrong. Several possibilities for future work into biocooperative robotics are presented in section 5.2, and biocooperative robotics can be considered as an extension of patient-cooperative robotics [5] that attempts to bring the robot closer to the role of the physical or occupational therapist. While the therapist has a complete overview of the patient's biomechanical, psychological and physiological state, patient-cooperative robots only have an insight into the patient's biomechanical state. Biocooperative robotics extend this by attempting to obtain an insight into the patient's physiological and psychological states as well, and this is an idea worth exploring further.

Furthermore, psychophysiological measurements are not necessarily useful only as part of a feedback loop. In general applied psychophysiology, psychophysiological measurements are often used simply as general measures of psychological factors in the same way as questionnaires. Though high accuracy is not expected in such cases (much like questionnaires are not expected to be perfectly accurate in all subjects), a sufficiently large sample of subjects could nonetheless reveal interesting trends. This would also be possible in rehabilitation, as results showed that certain psychophysiological features show significant differences between conditions or correlate significantly with self-report measures. Skin conductance appears to be the most useful of the four evaluated signals in this case, and it could thus be used as an alternative or complementary option to questionnaires (particularly those that measure arousal or related constructs) in motor rehabilitation.

Finally, these physiological measurements should be considered as a supplement to task performance and biomechanics in rehabilitation without necessarily focusing on their psychological component. For instance, treating heart rate simply as a measure of physical workload may prove more useful than trying to determine stress or boredom from it, especially since the physical workload can obscure information about psychological states. This could be an alternative direction for biocooperative rehabilitation, which focuses on both physiological and psychological aspects on the patient. It has been recently explored by Koenig et al. [171], who controlled heart rate in lower extremity rehabilitation by means of visual feedback.

5.2 Possible improvements and further work

5.2.1 A different upper extremity rehabilitation task

As mentioned in both sections 4.3.6.6 and 5.1, the choice of the ball-catching task in section 4.3 may not have been optimal for data fusion. One component of the task (placing the ball in the basket) does not depend on the difficulty level since the difficulty level only affects the size and speed of the ball. Psychophysiological differences between difficulty levels thus may not have been as large as they would have been if all task components had been affected by the difficulty level, and this may have contributed to the limited usefulness of psychophysiological measurements.

Most tasks performed in upper extremity rehabilitation are activities of daily living and are thus necessarily complex, involving many different components (e.g. reaching, grasping, lifting). Since such tasks lead to more effective rehabilitation, they should certainly not be changed for the sake of psychophysiology. However, since psychophysiology in motor rehabilitation is not yet a mature field, future psychophysiological studies in upper limb rehabilitation may prefer to start small: by having, for instance, a simple task with few components so that the effect of each component on psychophysiological responses would be easier to discern. If psychophysiological data fusion in such a simple task proved accurate, it would then be possible to gradually add more components, studying their effects on psychophysiology and data fusion. In such a way, it would eventually be possible to perform accurate psychophysiological data fusion in complex rehabilitation tasks or at least identify how complex a task can be before psychophysiological data fusion becomes impractical.

For instance, the ball-catching task from sections 3.3 and 4.3 could be simplified to only require horizontal movement as follows: The ball would appear at the top of the slope and begin rolling downward. The subject would then need to reach it as in the existing task, but once the ball was reached, it would for instance bounce back to the top of the slope. In this simplified task, it would still be possible to adjust task difficulty by changing the size and speed of the ball. If data fusion in this task were accurate, grasping and lifting the ball could be added, gradually extending the knowledge of psychophysiological responses to different components of the task. The ultimate goal would, of course, be to implement data fusion in complex tasks that yield the best rehabilitation outcome. A preliminary study with a simple task that only requires horizontal reaching has recently been conducted by Guerrero et al. [172], though not enough has been done yet to ascertain the effectiveness of the approach.

5.2.2 Lower extremity rehabilitation

All of the studies in this dissertation were performed with hardware and virtual scenarios for upper extremity rehabilitation. However, lower extremity rehabilitation also represents a major research field, and psychophysiological measurements could be potentially useful there. A team at ETH Zurich and the Neurological Clinic of Bad Aibling investigated the use of psychophysiological measurements for classification and control of cognitive workload in the Lokomat, a driven gait orthosis. Their initial approach used neural networks and has been recently published [173].

Following the use of adaptive discriminant analysis in the ball-catching scenario (section 4.3), ETH Zurich also chose to adopt adaptive discriminant analysis, and the Ljubljana team assisted them with transferring the classification algorithms to the Lokomat platform. The Lokomat implementation has been expanded to include multiple possible classes (task is too easy / task is appropriate / task is too hard), and attempts have also been made to separate the effects of physical and cognitive workload. The multiclass classification has proven effective. Though task performance again provides the most information, psychophysiological signals increase accuracy to a greater degree than in the upper extremity rehablitation task covered in section 4.3. A joint paper (though ETH Zurich did the overwhelming majority of the work) describing this work has also recently been published [174].

5.2.3 Additional sensors

A biocooperative rehabilitation system is one in which the parameters of the task are automatically adjusted so that the patient is challenged in a moderate but engaging and motivating way without causing undue stress or harm. The first ideas on the topic emphasized psychophysiological measurements of the autonomic nervous system as a convenient way of measuring psychological factors such as boredom, stress or motivation [9], so the dissertation also focused on these measurements. However, they are not the only way to measure psychological factors.

The general field of systems that can recognize human emotions is called affective computing and covers many possible affect recognition methods. Two methods that could be especially useful in rehabilitation are facial expression recognition and eye movement analysis. It was previously mentioned that many psychophysiological studies have monitored facial expressions through electromyography [16, 46] and found them to be a very useful complement to autonomic nervous system responses since they are very good at recognizing emotional valence while autonomic nervous system responses are more sensitive to arousal. However, facial electromyography is unlikely to be clinically practical since the needed electrodes require precise positioning, are time-consuming to apply, and are considered fairly obtrusive by the subject (since they are placed around the eyes and along the jaw). A good alternative would be to recognize facial expressions with machine vision methods, which have also proven to be a useful complement to autonomic nervous system responses in affective computing [106, 121].

Similar machine vision equipment could be used to measure and analyze eye movements, which have also proven to be a useful complement to autonomic nervous system responses in affective computing [69, 121]. Eye movements could also be measured using electrooculography [122, 139, 140], but this approach suffers from the same weakness as facial electromyography: the needed electrodes require precise positioning, are time-consuming to apply, and are considered fairly obtrusive by the subject. A potential recently developed solution are wearable goggles with built-in dry electrodes [175], which would be much faster to apply and far less obtrusive.

Of course, though the above two may be the most promising, many other affect recognition methods exist. Electroencephalography has seen extensive research [116, 122, 129], but also suffers from the problem of obtrusive, time-consuming electrodes. Speech recognition and general body movement recognition in general could also be useful and have recently been reviewed in an extensive paper [176]. Though biocooperative rehabilitation robotics has so far been mostly studied with psychophysiology, there is no need to limit ourselves to physiological measurements in the future; the ultimate goal is to keep the patient from becoming bored or frustrated, no matter what sensors are used.

It is thus evident that many additional sensors could be used. However, in addition to effectiveness, user-friendliness should be considered. The optimal selection of sensors would be able to determine the suitability of the task for the patient while not obstructing or annoying the patient. If only using autonomic nervous system responses, we might want to actually reduce the number of sensors by omitting the respiration sensor (which, of the four physiological sensors used in the dissertation, was the most unpleasant for the patient) and using a finger photoplethysmography sensor instead of ECG electrodes (since heart rate was found to be mostly an indicator of physical workload). Such a physiological sensor setup would only require the attachment of three sensors to a single hand and would thus be very user-friendly. For better accuracy, it could be expanded with a camera for eye tracking and/or facial expression recognition, which would not require attaching anything to the subject. This may be the most practical option for biocooperative robotics. It would also be relatively accurate, as heart rate, skin conductance and skin temperature would allow accurate estimation of the subject's arousal while eye tracking and facial expression recognition systems would allow contactless estimation of valence, which cannot be estimated by the autonomic nervous system measurements used in this dissertation.

5.2.4 Expanded classification options and validation measures

A limitation of the data fusion studies in this dissertation is that subjects were only given two choices: to 'prefer easier' or to 'prefer harder' task difficulty. There were thus only two possible states for classification. Obviously, it is possible that a subject finds the difficulty to be 'just right' and does not wish to change it. A possible follow-up study would thus be to utilize more than two states. The simplest option would be to define a third class called 'task is appropriate' where task difficulty would not be changed. Such a study has recently been made with our assistance by a team at ETH Zurich as mentioned in 5.2 and achieved promising results [174].

Another possibility would be a four-class setup based on the four quadrants of the arousal-valence space. Such a psychological model has already been used in many psychophysiological studies (as described in section 4.1.2) and would be quite relevant for rehabilitation. The goal would be to keep the patient in the high arousal/positive valence quadrant, and different actions could be taken depending on the current quadrant. This was considered early on, but was not used since autonomic nervous system measurements are relatively poor at distinguishing different levels of valence. However, with additional sensors such as facial expression recognition and eye tracking, it should be possible to estimate both arousal and valence to some degree. This has been shown to be effective outside rehabilitation [16, 75], but it is uncertain how useful this would be in rehabilitation practice. A third possibility would be to use estimation rather than classification, using methods such as fuzzy logic to map psychophysiological features to variables such as 'stress' or 'workload'. This was also briefly considered early on, but not used since classification is much better-established in psychophysiological literature. However, taking these ideas into account, there are three broad possibilities for psychophysiological feedback in rehabilitation:

- feedback without mapping psychophysiological features to psychological states;
- feedback by first mapping psychophysiological features to discrete psychological classes;
- feedback by first mapping psychophysiological features to continuous psychological variables.

Larger amounts of classes or the use of continuous psychological variables would require proper validation that the subject is actually in a certain state. This would most likely need to be done with self-report measures (questionnaires) or the opinions of independent observers. The data fusion studies in this dissertation could, in fact, have benefitted from the use of more detailed questionnaires, even though their reliability is uncertain. Even for a simple two- or three-state model (task is too easy / appropriate / too hard), more detailed self-report methods can be envisioned. For instance, the subject could provide his or her opinion on a 5- or 7-point Likert scale (from 'task should be much easier' through 'task should stay the same' to 'task should be much harder'), and these responses could be compared to biomechanical and psychophysiological measurements. Even if data fusion is only performed with a limited number of classes, it may still make sense to track a larger number of classes or psychological dimensions using questionnaires, though the questionnaires would need to be short enough to avoid excessively prolonging the experiment. These questionnaires would also allow us to determine whether the patient is actually optimally challenged as desired, something that could not be determined with absolute certainty in this dissertation.

5.2.5 Larger and better-controlled sample groups

The four studies conducted in this dissertation had relatively small sample sizes compared to most psychophysiological literature. The majority of psychophysiological studies include 20-50 subjects (e.g. 24 in Pastor-Sanz et al. [135], 28 in Frantzidis et al. [122], 34 in Christie and Friedman [66], 35 in Fairclough and Venables [144], 37 in Kreibig et al. [53], 41 in Bailenson et al. [106], 42 in Blechert et al. [128], 43 in Rainville et al. [107]). This dissertation, on the other hand, involves smaller sample groups (e.g. 11 patients in section 4.3, 23 patients in section 3.3). Furthermore, the sample groups are relatively nonhomogenous: patients have different diagnoses, different levels of functional ability, and are treated with different drugs. Even in section 3.2, where the sample group consists of 30 healthy subjects, several results are just barely significant and would ideally require a larger sample to properly verify. For instance, the effect of cognitive workload on skin temperature is significant with $p = 0.048$ and partial $\eta^2 = 0.35$. Additionally, important differences or correlations may have been missed because the sample size was too small to reach statistical significance.

However, small sample sizes are also common in other applied psychophysiological studies where it is difficult to recruit a larger group (e.g. six autistic children in Liu et al. [18], seven air traffic controllers in Wilson and Russell [116]). For this dissertation, recruiting a larger number of patients also proved impractical since a limited number of patients were available at the University Rehabilitation Institute. Liu et al. [18] and Wilson and Russell [116] partially offset the small sample size by performing multiple recordings with each subject, which was also done in sections 4.2 and 4.3 of this dissertation. However, particularly in sections 3.2 and 3.3, larger and more homogenous sample groups (e.g. similar levels of functional ability, no drugs that could affect psychophysiological responses) may have revealed additional useful information.

5.3 Adaptive discriminant analysis

Previously never used outside electroencephalography, adaptive discriminant analysis [26] represents a promising method for psychophysiological data fusion and biofeedback due to its ability to gradually adapt to the current subject. Though the overall goal of this dissertation was primarily to use it in rehabilitation, it can also be used in non-rehabilitation settings, achieving relatively high classification accuracy. It could thus be very useful for psychophysiologists in general human-computer interaction and may be potentially applicable to other data sources with high variability.

In the supervised adaptive discriminant analysis, the system was provided with the subject's preference so that it could adapt the discriminant function with accurate information. Since this information is generally unavailable, an unsupervised version was also developed where the discriminant function is adapted online using the system's own estimate of the subject’s preference. This is probably not the optimal unsupervised adaptive discriminant analysis, as it was validated empirically rather than theoretically. With improperly selected parameters of the update process, instability could occur, leading to poor classification and decisions that could be detrimental to the patient. An alternative unsupervised adaptive version [177] has been recently developed by the original authors of adaptive discriminant analysis and is most likely superior to ours. Nonetheless, the implementation used in this dissertation serves as a demonstration that even unsupervised adaptation can lead to improved classification results. Two other possibilities are also foreseen for adaptive classification.

In one alternative implementation of adaptive discriminant analysis, the patient’s first session with the system is a supervised session where the patient regularly inputs his or her preference into the system, enabling accurate adaptation. In later sessions, the adaptation is turned off. Thus, the system uses the first session to adapt to the patient to some degree, and this information is incorporated into the system during later sessions.

In a second alternative implementation of adaptive discriminant analysis, the discriminant function would not be adapted on its own, but the subject could manually input his or her own preference at any time. The system would then not only change the difficulty of the task, but also update its discriminant function with the subject’s input. Another possibility would be for the system to explicitly ask the subject for input if certain potentially erroneous trends are detected (e.g. if the system repeatedly estimates that the task is too easy even though the subject has reached a very high difficulty level).

6 Conclusions

This dissertation focuses on the nearly unexplored field of psychophysiology in motor rehabilitation, particularly on the creation of a biocooperative feedback loop: a system that uses psychophysiological and other measurements to determine the suitability of the task for the current patient and then adjust the task in order to keep the patient from becoming bored or frustrated. Four psychophysiological responses were measured, analyzed and used in data fusion and feedback: heart rate, skin conductance, respiration and skin temperature.

An analysis of the effects of physical activity on psychophysiological responses found that, at least in the inverted pendulum task, heart rate and skin conductance are strongly affected by physical activity. At high levels of physical activity, it is difficult to discern any psychological influence on these two physiological responses. Respiration and skin temperature, on the other hand, are less strongly affected and show significant differences between different levels of cognitive workload even in the presence of physical activity. However, multiple physiological sensors are recommended for cognitive workload estimation in haptic interaction, and nonphysiological sensors such as force sensors and accelerometers should be used to gauge the level of physical activity.

An analysis of the effects of stroke on psychophysiological responses found that psychophysiological responses are weakened by stroke, though some are more strongly impaired than others. In patients, skin conductance was found to be the most useful for psychological state assessment, as skin conductance level differentiated between different difficulty levels of a task while skin conductance response frequency was correlated with self-reported arousal. Additionally, skin conductance sensors are very easy to attach and use. Skin temperature, which is also easy to use, unfortunately showed different results in control and patient groups. It also responded to stimuli much more slowly than skin conductance. Heart rate offered uncertain results in patients with regard to psychophysiology, but could at least be used as a measure of physical effort (in which case, a simpler measuring method than electrocardiography would suffice).

Data fusion was performed first in a non-rehabilitation setting with healthy subjects and no physical activity. Several classification methods were tested, but it appears that, at least in this case the choice of classifier is not quite as important as selecting the most informative features. The exception was adaptive discriminant analysis, which was more accurate than all other classification methods in fusion of only psychophysiological features even though it has not been previously used outside of electroencephalography (section 4.2: accuracy rate of 75.0% with the best nonadaptive method and 86.7% with adaptive discriminant analysis; section 4.3 – healthy subjects: 62.5% with the best nonadaptive method and 77.1% with adaptive discriminant analysis; section 4.3 – patients: 63.6% with the best nonadaptive method and 77.3% with adaptive discriminant analysis). Adaptive discriminant analysis was, however, no more useful than other classification methods when task performance features were included. In any case, classification accuracy in a non-rehabilitation setting was over 85% for a two-state problem (is the task too easy or too hard?), showing that the utilized data fusion methods are viable. Such classification accuracy could even be achieved with only psychophysiological features, proving that they can be a useful primary source of information in a non-rehabilitation setting.

In a motor rehabilitation task, data fusion of the four psychophysiological responses alone was not very accurate, although adaptive discriminant analysis improved accuracy. Data fusion was much more accurate with task performance and biomechanics. Psychophysiological responses thus cannot be used as a primary source of information in rehabilitation. If dimension reduction is used, a combination of task performance and psychophysiology can achieve the highest accuracy. Psychophysiological measurements can thus serve as a supplementary source of information, although it is uncertain whether they provide enough additional information to justify the increased cost and complexity of the system. They may also be a useful source of information in tasks and environments where task performance or biomechanical measurements are either not available or are not at all connected to the subject’s mood.

However, despite the somewhat discouraging results, there is much room for improvement. A number of ideas for future work in biocooperative robotics have been suggested, ranging from more thorough validation using questionnaires to the utilization of other, nonphysiological sensors. This dissertation lays out some of the first steps in implementing biocooperative control, including an operational biocooperative feedback loop using psychophysiological measurements. It can be thought of as an extension of patient-cooperative robotics that attempts to bring the robot closer to the role of the physical or occupational therapist. While the therapist has a complete overview of the patient's biomechanical, psychological and physiological state, patient-cooperative robots only have an insight into the patient's biomechanical state. The biocooperative feedback loop presented in this dissertation extends this by attempting to obtain an insight into the patient's physiological and psychological states as well. Whether the idea of biocooperative rehabilitation will gain ground remains an open question, but the author of this dissertation firmly believes that adjusting the difficulty of the task to keep the patient appropriately challenged and motivated would be an important addition to rehabilitation robotics and could potentially lead to an improved rehabilitation outcome.

7 Original scientific contributions

- **Analysis of healthy subjects' psychophysiological responses to a combination of psychological and physical activity in haptic human-robot interaction;**

The analysis was performed in a study where 30 subjects performed an inverted pendulum balancing task with the HapticMaster haptic robot at two levels of physical workload and three levels of cognitive workload. Heart rate and skin conductance level were primarily influenced by physical workload, and there was also a noticeable influence of physical workload on skin conductance response frequency. Neither respiration nor peripheral skin temperature were significantly affected by physical workload. Respiratory variability decreased from baseline during the moderately cognitively challenging condition while skin temperature decreased during the cognitively overchallenging condition. This suggests that respiration and skin temperature are effective for the estimation of cognitive workload in haptic interaction.

- **Analysis of psychophysiological differences between healthy subjects and hemiparetic patients in clinical rehabilitation scenarios;**

The analysis was performed in a study where 23 stroke and 23 control subjects performed a virtual rehabilitation task and a simple cognitive task (the Stroop word-colour interference task). Significant differences between stroke and control groups were found especially for heart rate and peripheral skin temperature, with the stroke group exhibiting weaker responses to both the rehabilitation task and the cognitive task. Skin conductance appears to be the most useful psychophysiological signal in the stroke group, as there is a significant correlation with self-reported arousal as well as a significant difference between different difficulty levels of the virtual rehabilitation task.

- **Psychophysiological sensor fusion for task suitability assessment in rehabilitation robotics using different methods**

A number of different sensor fusion methods were implemented for task suitability assessment. Dimension reduction was performed using principal component analysis and sequential floating forward selection. Discriminant analysis, diagonal discriminant analysis, nearest-neighbor classification, classification trees and support vector machines were used to classify psychophysiological and other variables into two classes: the task is too easy or too hard. They were implemented in both a simple cognitively challenging task and a virtual rehabilitation task. The subject's and experimenter's opinions were used as validation measures. Psychophysiological variables were less accurate in classification than task performance and biomechanics, but provided supplementary information.

- **An adaptive method that can adapt to intersubject differences in psychophysiological responses**

Kalman adaptive linear discriminant analysis and the adaptive information matrix, previously only used in electroencephalography, were transferred to autonomic nervous system responses and used to perform online adaptation of the classification rules. They were able to improve the classification accuracy for psychophysiological variables over established classification methods in both the simple cognitively challenging task and the virtual rehabilitation task. Both supervised and unsupervised adaptation was demonstrated, though the unsupervised implementation is not optimal.

- **A biocooperative controller that can adapt the parameters of a rehabilitation task based on adaptive fusion of psychophysiological, biomechanical and other sensors.**

After training different data fusion methods on a larger group of subjects, a biocooperative feedback loop was implemented and tested on 10 healthy subjects and 6 stroke subjects performing a virtual rehabilitation task. The controller used sequential floating forward selection and discriminant analysis to fuse task performance, biomechanics and psychophysiology into an estimate of whether the task difficulty should be increased or decreased. Task difficulty was then adjusted accordingly by changing the parameters of the virtual scenario. The controller reached approximately 90% agreement with the subject's opinion for both healthy and stroke subjects.

Publications arising from this work

Original scientific articles

NOVAK, Domen, ZIHERL, Jaka, OLENŠEK, Andrej, MILAVEC, Maja, PODOBNIK, Janez, MIHELJ, Matjaž, MUNIH, Marko. Psychophysiological responses to robotic rehabilitation tasks in stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2010, vol. 18, no. 4, pp. 351-361.

NOVAK, Domen, MIHELJ, Matjaž, MUNIH, Marko. Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction. *Robotica*, 2011, vol. 29, no. 3, pp. 367-374.

NOVAK, Domen, MIHELJ, Matjaž, ZIHERL, Jaka, OLENŠEK, Andrej, MUNIH, Marko. Psychophysiological responses in a biocooperative feedback loop for upper extremity rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2011, vol. 19, pp. 400-410.

NOVAK, Domen, MIHELJ, Matjaž, MILAVEC, Maja, MUNIH, Marko. Task difficulty estimation in human-computer interaction using psychophysiology and adaptive discriminant analysis. Submitted for publication.

NOVAK, Domen, MIHELJ, Matjaž, MUNIH, Marko. A survey of methods for data fusion and biofeedback using autonomic nervous system responses in psychophysiology. Submitted for publication.

Published scientific conference contributions

NOVAK, Domen, MIHELJ, Matjaž, MUNIH, Marko. Psychophysiological indicators in virtual reality-assisted motor rehabilitation. In: *International Neurorehabilitation Symposium 2009 - Program and abstracts*, 2009, pp. 59-60.

NOVAK, Domen, MIHELJ, Matjaž, MUNIH, Marko. Using psychophysiological measurements in physically demanding virtual environments. In: *Proceedings of INTERACT 2009* (*Lecture Notes on Computer Science)*, 2009, pp. 490-493.

NOVAK, Domen, ZIHERL, Jaka, OLENŠEK, Andrej, PODOBNIK, Janez, MIHELJ, Matjaž, MUNIH, Marko. Robotska rehabilitacija z navidezno resničnostjo in psihofiziološkimi meritvami. In: *Zbornik 12. mednarodne multikonference Informacijska družba - IS 2009,* 2009, pp. 423-426.

NOVAK, Domen, MIHELJ, Matjaž, ZIHERL, Jaka, OLENŠEK, Andrej, MUNIH, Marko. Measuring motor actions and psychophysiology for task difficulty estimation in human-robot interaction. In: *Proceedings of Measuring Behavior 2010*, 2010, pp. 269-272.

NOVAK, Domen, MIHELJ, Matjaž, ZIHERL, Jaka, OLENŠEK, Andrej, MUNIH, Marko. Adaptivna senzorna integracija na osnovi biomehanskih in fizioloških meritev v rehabilitacijski robotiki. In: *Zbornik 13. mednarodne multikonference Informacijska družba - IS 2010*, 2010, pp. 357-360.

NOVAK, Domen, MIHELJ, Matjaž, ZIHERL, Jaka, OLENŠEK, Andrej, MUNIH, Marko. Task difficulty adjustment in biocooperative rehabilitation using psychophysiological responses. In: *Proceedings of the 12th International Conference on Rehabilitation Robotics*, 2011, pp. 381-386.

References

1. Liebermann D, Buchman A and Franks I. *Enhancement of motor rehabilitation through the use of information technologies.* Clinical Biomechanics 2006; vol. 21, pp. 8-20.

2. Johnson M. *Recent trends in robot-assisted therapy environments to improve real-life functional performance after stroke.* Journal of Neuroengineering and Rehabilitation 2006; 3:29.

3. Lo AC et al. *Robot-assisted therapy for long-term upper-limb impairment after stroke.* New England Journal of Medicine 2010; vol. 362, pp. 1772-1783..

4. Holden MK. *Virtual environments for motor rehabilitation: Review.* CyberPsychology & Behavior 2005; vol. 8, pp. 187-219.

5. Riener R et al. *Patient-cooperative strategies for robot-aided treadmill training: First experimental results.* IEEE Transactions on Neural Systems and Rehabilitation Engineering 2005; vol. 13, pp. 380-394.

6. Emken JL, Benitez R and Reinkensmeyer DJ. *Human-robot cooperative movement training: Learning a novel sensory motor transformation during walking with robotic assistance-as-needed.* Journal of Neuroengineering and Rehabilitation, 2007; vol. 4:8.

7. Wolbrecht ET et al. *Optimizing compliant, model-based robotic assistance to promote neurorehabilitation.* IEEE Transactions on Neural Systems and Rehabilitation Engineering 2008; vol. 16, pp. 286-297.

8. Barkana DE and Sarkar N. *Towards a smooth human-robot interaction for rehabilitation robotic systems.* Advanced Robotics 2009; vol. 23, pp. 1641-1662.

9. Munih M et al. *MIMICS: Multimodal immersive motion rehabilitation of upper and lower extremities by exploiting biocooperation principles.* IEEE Conference on Rehabilitation Robotics 2009, pp. 127-132. Kyoto, Japan: IEEE Press, 2009.

10. Riener R and Munih M. *Guest editorial special section on rehabiliation via bio-cooperative control*. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2010; vol. 18, pp. 337-338.

11. Gomez P and Danuser B. *Affective and physiological responses to environmental noises and music.* International Journal of Psychophysiology 2004; vol. 53, pp. 91-103.
12. Collet C, Averty P and Dittmar A. *Autonomic nervous system and subjective ratings of strain in air-traffic control.* Applied Ergonomics 2009; vol. 40, pp. 23-32.
13. Veltman JA and Gaillard AWK. *Physiological workload reactions to increasing levels of task difficulty.* Ergonomics 1998; vol. 41, pp. 656-669.
14. Backs RW et al. *Cardiac measures of driver workload during simulated driving with and without visual occlusion.* Human Factors 2003; vol. 45, pp. 525-538.
15. Carrillo E et al. *Gender differences in cardiovascular and electrodermal responses to public speaking task: the role of anxiety and mood states.* International Journal of Psychophysiology 2001; vol. 42, pp. 253-264.
16. Mandryk RL and Atkins SM. *A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies.* International Journal of Human-Computer Studies 2007; vol. 65, pp. 329-347.
17. Rani P et al. *Anxiety detecting robotic system - towards implicit human-robot collaboration.* Robotica 2004; vol. 22, pp. 85-95.
18. Liu C et al. *Online affect detection and robot behavior adaptation for intervention of children with autism.* IEEE Transactions on Robotics 2008; vol. 24, pp. 883-896.
19. Kreibig SD. *Autonomic nervous system activity in emotion: A review.* Biological Psychology 2010; vol. 84, pp. 394-421.
20. van der Linde RQ and Lammertse P. *HapticMaster - a generic force controlled robot for human interaction.* Industrial Robot: An International Journal 2003; vol. 30, pp. 515-524.
21. Roth DL, Bachtler SD and Fillingim RB. *Acute emotional and cardiovascular effects of stressful mental work during aerobic exercise.* Psychophysiology 1990; vol. 27, pp. 694-701.
22. Webb HE et al. *Psychological stress during exercise: cardiorespiratory and hormonal responses.* European Journal of Applied Physiology 2008; vol. 104, pp. 973-981.
23. Korpelainen JT, Sotaniemi KA and Myllylä, V. V. *Autonomic nervous system disorders in stroke.* Clinical Autonomic Research 1999; vol. 9, pp. 325-333.
24. Andersson S and Finset A. *Heart rate and skin conductance reactivity to brief psychological stress in brain-injured patients.* Journal of Psychosomatic Research 1998; vol. 44, pp. 645-656.

25. Macleod CM. *Half a century of research on the Stroop effect: an integrative review.* Psychological Bulletin 1991; vol. 109, pp. 163-203.

26. Vidaurre C et al. *Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces.* IEEE Transactions on Biomedical Engineering 2007; vol. 54, pp. 550-556.

27. Berch DB, Krikorian R and Huha EM. *The Corsi Block-Tapping Task: Methodological and Theoretical Considerations.* Brain and Cognition 1998; vol. 38, pp. 317-338.

28. Krebs HI et al. *Robot-aided neurorehabilitation.* IEEE Transactions on Rehabilitation Engineering 1998; vol. 6, pp. 75-87.

29. Loureiro R et al. *Upper limb robot mediated stroke therapy - GENTLE/s approach.* Autonomous Robots 2003; vol. 15, pp. 35-51.

30. Nef T, Mihelj M and Riener R. *ARMin: a robot for patient-cooperative arm therapy.* Medical and Biological Engineering and Computing 2007; vol. 45, pp. 887-900.

31. Cacioppo JT, Tassinary LG and Berntson GG [eds.]. *Handbook of Psychophysiology,* 3rd edition. Cambridge: Cambridge University Press, 2007.

32. Stern RM, Ray WJ and Quigley KS. *Psychophysiological Recording,* 2nd edition. Oxford: Oxford University Press, 2001.

33. Ekman P, Levenson RW and Friesen WV. *Autonomic nervous system activity distinguishes among emotions.* Science 1983; vol. 221, pp. 1208-1210.

34. Cameirão MS et al. *Physiological responses during performance within a virtual scenario for the rehabilitation of motor deficits.* Proceedings of PRESENCE 2007, pp. 85-88. Barcelona, Spain: Starlab Barcelona, 2007.

35. Cacioppo JT and Tassinary LG. *Inferring psychological significance from physiological signals.* American Psychologist 1990; vol. 45, pp. 16-28.

36. Ekman P. *An argument for basic emotions.* Cognition and Emotion 1992; vol. 6, pp. 169-200.

37. Russell JA. *A circumplex model of affect.* Journal of Personality and Social Psychology 1980; vol. 39, pp. 1161-1178.

38. Mehrabian A. *Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament.* Current Psychology 1996; vol. 14, pp. 261-292.

39. Podobnik J. *Reaching and grasping in haptic virtual environments.* Doctoral dissertation, University of Ljubljana, Faculty of Electrical Engineering, 2009.

40. Byrne EA and Parasuraman R. *Psychophysiology and adaptive automation.* Biological Psychology 1996; vol. 42, pp. 249-268.

41. Peter C and Herbon A. *Emotion representation and physiology assignments in digital systems.* Interacting with Computers 2006; vol. 18, pp. 139-170.

42. Fowles DC et al. *Publication recommendations for electrodermal measurements.* Psychophysiology 1981; vol. 18, pp. 232-239.

43. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal* 1996; vol. 17, pp. 354-381.

44. Berntson GG et al. *Heart rate variability: Origins, methods and interpretive caveats.* Psychophysiology 1997; vol. 34, pp. 623-648.

45. Liu C et al. *Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback.* International Journal of Human-Computer Interaction 2009; vol. 25, pp. 506-529.

46. Picard RW, Vyzas E and Healey J. *Toward machine emotional intelligence: Analysis of affective physiological state.* IEEE Transactions on Pattern Analysis and Machine Intelligence 2001; vol. 23, pp. 1175-1191.

47. Christov II, Dotsinsky IA and Daskalov IK. *High pass filtering of ECG signals using QRS elimination.* Medical and Biological Engineering and Computing 1992; vol. 30, pp. 253-256.

48. Friesen GM et al. *A comparison of the noise sensitivity of nine QRS detection algorithms.* IEEE Transactions on Biomedical Engineering 1990; vol. 37, pp. 85-98.

49. Proakis JG and Manolakis DK. *Digital Signal Processing: Principles, Algorithms and Applications,* 4th edition. Prentice Hall, 2006.

50. Yao YJ et al. *Heart rate and respiration responses to real traffic pattern flight.* Applied Psychophysiology and Biofeedback 2008; vol. 33, pp. 203-209.

51. McCraty R et al. *The effects of emotions on short-term power spectrum analysis of heart rate variability.* The American Journal of Cardiology 1995; vol. 76, pp. 1089-1093.

52. Pauls CA and Stemmler G. *Repressive and defensive coping during fear and anger.* Emotion 2003; vol. 3, pp. 284-302.

53. Kreibig SD et al. *Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films.* Psychophysiology 2007; vol. 44, pp. 787-806.

54. Malmivuo J and Plonsey R. The Electrodermal Response. In: *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields.* New York: Oxford University Press, 1995.

55. Bach DR et al. *Modelling event-related skin conductance responses.* International Journal of Psychophysiology 2010; vol. 75, pp. 349-356.

56. Boucsein W. *Electrodermal activity.* New York: Plenum Press, 1992.

57. Haarmann A, Boucsein W and Schaefer F. *Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight.* Applied Ergonomics 2009; vol. 40, pp. 1026-1040.

58. Nikula R. *Psychological correlates of nonspecific skin conductance responses.* Psychophysiology 1991; vol. 28, pp. 86-90.

59. Boiten FA, Frijda NH and Wientjes CJ. *Emotions and respiratory patterns: review and critical analysis.* International Journal of Psychophysiology 1994; vol. 17, pp. 103-128.

60. Boiten F. *Component analysis of task-related respiratory patterns.* International Journal of Psychophysiology 1993; vol. 15, pp. 91-104.

61. Gomez P, Stahel WA and Danuser B. *Respiratory responses during affective picture viewing.* Biological Psychology 2004; vol. 67, pp. 359-373.

62. van Diest I et al. *Anxiety and respiratory variability.* Physiology & Behavior 2006; vol. 89, pp. 189-195.

63. Kistler A, Mariauzouls C and von Berlepsch K. *Fingertip temperature as an indicator for sympathetic responses.* International Journal of Psychophysiology 1998; vol. 29, pp. 35-41.

64. Ohsuga M, Shimono F and Genno H. *Assessment of phasic work stress using autonomic indices.* International Journal of Psychophysiology 2001; vol. 40, pp. 211-220.

65. Min BC et al. *Autonomic responses of young passengers contingent to the speed and driving mode of a vehicle.* International Journal of Industrial Ergonomics 2002; vol. 29, pp. 187-198.

66. Christie IC and Friedman BH. *Autonomic specificity of discrete emotion and dimensions of affective space.* International Journal of Psychophysiology 2004; vol. 51, pp. 143-153.

67. Kim KH, Bang SW and Kim SR. *Emotion recognition system using short-term monitoring of physiological signals.* Medical and Biological Engineering and Computing 2004; vol. 42, pp. 419-427.

68. Stephens CL, Christie IC and Friedman BH. *Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis.* Biological Psychology 2010; vol. 84, pp. 463-473.

69. Zhai J and Barreto A. *Stress detection in computer users through non-invasive monitoring of physiological signals.* Biomedical Sciences Instrumentation 2006; vol. 42, pp. 495-500.

70. Arroyo-Palacios J and Romano DM. *A flexible bio-affective gaming interface.* AISB 2010 Symposium: AI & Games. Leicester, UK, 2010.

71. Nasoz F et al. *Emotion recognition from physiological signals for presence technologies.* International Journal of Cognition, Technology and Work 2003; vol. 6, pp. 4-14.

72. Lisetti CL and Nasoz F. *Using noninvasive wearable computers to recognize human emotions from physiological signals.* EURASIP Journal on Applied Signal Processing 2004; vol. 11, pp. 1672-1687.

73. Mohammad Y and Nishida T. *Using physiological signals to detect natural interactive behavior.* Applied Intelligence 2010; vol. 33, pp. 79-92.

74. Setz C et al. *Discriminating stress from cognitive load using a wearable EDA device.* IEEE Transactions on Information Technology in Biomedicine 2010; vol. 14, pp. 410-417.

75. van den Broek EL et al. *Affective man-machine interface: Unveiling human emotions through biosignals.* Biomedical Engineering Systems and Technologies: BIOSTEC2009, pp. 21-47. Berlin Heidelberg: Springer-Verlag, 2010.

76. Haag A et al. *Emotion recognition using bio-sensors: First steps towards an automatic system.* Affective Dialogue Systems 2004, pp. 36-48. Berlin Heidelberg : Springer-Verlag, 2004.

77. Kulić D and Croft EA. *Affective state estimation for human-robot interaction.* IEEE Transactions on Robotics 2007; vol. 23, pp. 991-1000.

78. Yannakakis GN, Hallam J and Lund HH. *Entertainment capture through heart rate activity in physical interactive playgrounds.* User Modeling and User-Adapted Interaction 2008; vol. 18, pp. 207-243.

79. Hill SG et al. *Comparison of four subjective workload rating scales.* Human Factors 1992; vol. 34, pp. 429-439.

80. Carver CS and White TL. *Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales.* Journal of Personality & Social Psychology 1994; vol. 67, pp. 319-333.

81. Bradley MM and Lang PJ. *Measuring emotion: the self-assessment manikin and the semantic differential.* Journal of Behavior Therapy and Experimental Psychiatry 1994; vol. 25, pp. 49-59.

82. Feldman PJ et al. *Negative emotions and acute physiological responses to stress.* Annals of Behavioral Medicine 1999; vol. 21, pp. 216-222.

83. Schwerdtfeger A. *Predicting autonomic reactivity to public speaking: don't get fixed on self-report data!* International Journal of Psychophysiology 2004; vol. 52, pp. 217-224.

84. Fowles DC. *Psychophysiology and psychopathology: A motivational approach.* Psychophysiology 1988; vol. 25, pp. 373-391.

85. Wilhelm FH et al. *Mechanisms of virtual reality exposure therapy: The role of the behavioral activation and behavioral inhibition systems.* Applied Psychophysiology and Biofeedback 2005; vol. 30, pp. 271-284.

86. Kulić D and Croft EA. *Physiological and subjective responses to articulated robot motion.* Robotica 2007; vol. 25, pp. 13-27.

87. Lakusic N, Mahovic D and Babic T. *Gradual recovery of impaired cardiac autonomic balance within first six months after ischemic cerebral stroke.* Acta Neurologica Belgica 2005; vol. 105, pp. 39-42.

88. Korpelainen JT et al. *Suppressed sympathetic skin response in brain infarction.* Stroke 1993; vol. 24, pp. 1389-1392.

89. Muslumanoglu L et al. *Evaluation of autonomic nervous system functions in post-stroke patients.* American Journal of Physical Medicine and Rehabilitation 2002; vol. 81, pp. 721-725.

90. Curt A, Weinhardt C and Dietz V. *Significance of sympathetic skin response in the assessment of autonomic failure in patients with spinal cord injury.* Journal of the Autonomic Nervous System 1996; vol. 61, pp. 175-180.

91. Andreassi JJ. *Psychophysiology: Human Behaviour and Physiological Response.* London: Lawrence Erlbaum Associates, 2000.

92. Meehan M et al. *Physiological measures of presence in stressful virtual environments.* ACM Transactions on Graphics 2002; vol. 21, pp. 645-653.

93. Cohen J. *Eta-squared and partial eta-squared in fixed factor ANOVA designs.* Educational and Psychological Measurement 1973; vol. 33, pp. 107-112.
94. Abdi H. The Bonferroni and Šidák corrections for multiple comparisons. In: N. Salkind [ed.], *Encyclpedia of Measurement and Statistics.* Thousand Oaks (CA): Sage, 2007.
95. Huynh H. and Feldt LS. *Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs*. Journal of Educational Statistics 1976; vol. 1, pp. 69-92.
96. Lilliefors HW. *On the Kolmogorov-Smirnov test for normality with mean and variance unknown*. Journal of the American Statistical Association 1967; vol. 62, pp. 399-402.
97. Koenig AC et al. *Model-based heart rate prediction during Lokomat walking.* Proceedings of the 31st Annual International Conference of IEEE EMBC. Minneapolis: IEEE Press, 2009.
98. Ziherl J et al. *Pick to place trajectories in human arm training movement.* Technology and Health Care 2009; vol. 17, pp. 323-335.
99. Folstein MF, Folstein SE and McHugh PR. *'Mini-mental state': A practical method for grading the cognitive state of patients for the clinician.* Journal of Psychiatric Research 1975; vol. 12, pp. 189-198.
100. Kidd D et al. *The functional independence measure: A comparative validity and reliability study.* Disability and Rehabilitation 1995; vol. 17, pp. 10-14.
101. Ziherl J et al. *Evaluation of upper extremity robot-assistances in subacute and chronic stroke subjects.* Journal of Neuroengineering and Rehabilitation 2010; vol. 7:52.
102. Siepmann M et al. *The effects of lorazepam on skin conductance responses to aversive stimuli in healthy subjects.* Clinical Autonomic Research 2007; vol. 17, pp. 160-164.
103. Wagner J, Kim J and André E. *From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification.* 2005 IEEE International Conference on Multimedia and Expo, pp. 940-943. Amsterdam, Netherlands, 2005.
104. Pour PA et al. *The impact of system feedback on learners' affective and physiological states.* Intelligent Tutoring Systems 2010; pp. 264-273. Pittsburgh, PA.

105. Rani P, Sarkar N and Adams J. *Anxiety-based affective communication for implicit human-machine interaction.* Advanced Engineering Informatics 2007; vol. 21, pp. 323-334.
106. Bailenson JN et al. *Real-time classification of evoked emotions using facial feature tracking and physiological responses.* International Journal of Human-Computer Studies 2008; vol. 66, pp. 303-317.
107. Rainville P et al. *Basic emotions are associated with distinct patterns of cardiorespiratory activity.* International Journal of Psychophysiology 2006; vol. 61, pp. 5-18.
108. Rigas G et al. *A user independent, biosignal based, emotion recognition method.* User Modeling 2007, pp. 314-318. Berlin Heidelberg: Springer-Verlag, 2007.
109. Healey JA and Picard RW. *Detecting stress during real-world driving tasks using physiological sensors.* IEEE Transactions on Intelligent Transportation Systems 2005; vol. 6, pp. 156-166.
110. Gu Y et al. *A biometric signature based system for improved emotion recognition using physiological responses from multiple subjects.* 2010 8th IEEE International Conference on Industrial Informatics, pp. 61-66. Osaka, Japan, 2010.
111. Alpers GW, Wilhelm FH and Roth WT. *Psychophysiological assessment during exposure in driving phobic patients.* Journal of Abnormal Psychology 2005; vol. 114, pp. 126-139.
112. Yannakakis GN and Hallam J. *Entertainment modeling through physiology in physical play.* International Journal of Human-Computer Studies 2008; vol. 66, pp. 741-755.
113. Tognetti S et al. *Enjoyment recognition from physiological data in a car racing game.* Proceedings of the 3rd International Workshop on Affective Interaction in Natural Environments (AFFINE '10). pp. 3-8, 2010.
114. Kim J. Bimodal emotion recognition using speech and physiological changes. In: R. Plutchik and H. Kellerman [eds.], *Robust speech recognition and understanding.* Vienna: I-Tech Education and Publishing, 2007.
115. Kim J and Andre E. *Emotion recognition based on physiological changes in music listening.* IEEE Transactions on Pattern Analysis and Machine Intelligence 2008; vol. 30, pp. 2067-2083.

116. Wilson GF and Russell CA. *Operator functional state classification using multiple psychophysiological features in an air traffic control task.* Human Factors 2003; vol. 45, pp. 381-389.

117. Bishop CM. *Pattern Recognition and Machine Learning.* Springer, 2006.

118. Lisetti C et al. *Developing multimodal intelligent affective interfaces for tele-home health care.* International Journal of Human-Computer Studies 2003; vol. 59, pp. 245-255.

119. Rani P et al. *An empirical study of machine learning techniques for affect recognition in human-robot interaction.* Pattern Analysis and Applications 2006; vol. 9, pp. 58-69.

120. Shen L, Wang M and Shen R. *Affective e-Learning: Using "Emotional" data to improve learning in pervasive learning environment.* Educational Technology & Society 2009; vol. 12, pp. 176-189.

121. Kapoor A, Burleson W and Picard RW. *Automatic prediction of frustration.* International Journal of Human-Computer Studies 2007; vol. 65, pp. 724-736.

122. Frantzidis CA et al. *On the classification of emotional biosignals evoked while viewing affective pictures: An integrated data-mining-based approach for healthcare applications.* IEEE Transactions on Information Technology in Biomedicine 2010; vol. 14, pp. 309-318.

123. Hand DJ and Yu K. *Idiot's Bayes - Not so stupid after all?* International Statistical Review 2001; vol. 69, pp. 385-398.

124. Calvo RA, Brown I and Scheding S. *Effect of experimental factors on the recognition of affective mental states through physiological measures.* Proceedings of the 22nd Australasian Joint Conference on Artificial Intelligence. pp. 62-70, 2009.

125. Müller ME. *Why some emotional states are easier to be recognized than others: A thorough data analysis and a very accurate rough set classifier.* 2006 IEEE International Conference on Systems, Man and Cybernetics. pp. 1624-1629, 2006.

126. Conati C. *Probabilistic assessment of user's emotions in educational games.* Applied Artificial Intelligence 2002, vol. 16, pp. 555-575.

127. Janssen JH, van den Broek EL and Westerink JHDM. *Personalized affective music player.* 3rd International Conference on Affective Computing and Intelligent Interaction. pp. 1-6. Amsterdam, 2009.

128. Blechert J et al. *Identifying anxiety states using broad sampling and advanced processing of peripheral physiological information.* Biomedical Sciences Instrumentation 2006; vol. 42, pp. 136-141.

129. Chanel G et al. *Short-term emotion assessment in a recall paradigm.* International Journal of Human-Computer Studies 2009; vol. 67, pp. 607-627.

130. Setz C et al. *Combining worthless sensor data.* Measuring Mobile Emotions Workshop at MobileHCI, 2009.

131. Dudoit S, Fridlyand J and Speed TP. *Comparison of discrimination methods for the classification of tumors using gene expression data.* Journal of the American Statistical Association 2002; vol. 97, pp. 77-87.

132. Schölkopf B. and Smola AJ. *Learning with Kernels.* MIT Press, 2001.

133. Katsis CD, Ganiatsas G and Fotiadis DI. *An integrated telemedicine platform for the assessment of affective physiological states.* Diagnostic Pathology 2006; vol. 1:16.

134. Katsis CD et al. *Toward emotion recognition in car-racing drivers: A biosignal processing approach.* IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans 2008; vol. 38, pp. 502-512.

135. Pastor-Sanz L et al. *Clinical validation of a wearable system for emotional recognition based on biosignals.* Journal of Telemedicine and Telecare 2008; vol. 14, pp. 152-154.

136. Levillain F et al. *Characterizing players experience from physiological signals using fuzzy decision trees.* 2010 IEEE Conference on Computational Intelligence and Games. pp. 75-82, 2010.

137. Leon E et al. *Real-time detection of emotional changes for inhabited environments.* Computers & Graphics 2004; vol. 28, pp. 635-642.

138. Leon E et al. *A user-independent real-time emotion recognition system for software agents in domestic environments.* Engineering Applications of Artificial Intelligence 2007; vol. 20, pp. 337-345.

139. Wilson GF and Russell CA. *Real-time assessment of mental workload using psychophysiological measures and artificial neural networks.* Human Factors 2003; vol. 45, pp. 635-643.

140. Wilson GF and Russell CA. *Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding.* Human Factors 2007; vol. 49, pp. 1005-1018.

141. Scheirer J et al. *Frustrating the user on purpose: a step toward building an affective computer.* Interacting with Computers 2002; vol. 14, pp. 93-118.

142. Toups ZO et al. *A design for using physiological signals to affect team game play.* Foundations of Augmented Cognition. pp. 134-139. San Francisco, CA, 2006.

143. Grigore O et al. *Stochastic algorithms for adaptive lighting control using psycho-physiological features.* International Journal of Biology and Biomedical Engineering 2008; vol. 2, pp. 9-18.

144. Fairclough SH and Venables L. *Prediction of subjective states from psychophysiology: a multivariate approach.* Biological Psychology 2006; vol. 71, pp. 100-110.

145. Grundlehner B et al. *The design and analysis of a real-time, continuous arousal monitor.* 2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks, pp. 156-161, 2009.

146. Rani P et al. *Online stress detection using psychophysiological signals for implicit human-robot cooperation.* Robotica 2002; vol. 20, pp. 673-685.

147. Kumar M et al. *Fuzzy evaluation of heart rate signals for mental stress assessment.* IEEE Transactions on Fuzzy Systems 2007; vol. 15, pp. 791-808.

148. Ting C et al. *Real-time adaptive automation system based on identification of operator functional state in simulated process control operations.* IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans 2010; vol. 40, pp. 251-262.

149. King RD, Feng C and Shutherland A. *StatLog: Comparison of classification algorithms on large real-world problems.* Applied Artificial Intelligence 1995; vol. 9, pp. 259-287.

150. Harper PR. *A review and comparison of classification algorithms for medical decision making.* Health Policy 2005; vol. 71, pp. 315-331.

151. Hua J et al. *Optimal number of features as a function of sample size for various classification rules.* Bioinformatics 2005; vol. 21, pp. 1509-1515.

152. Caruana R and Niculescu-Mizil A. *An empirical comparison of supervised learning algorithms.* Proceedings of the 23rd International Conference on Machine Learning, pp. 161-168. Pittsburgh, PA, 2006.

153. Caruana R, Karampatziakis N and Yessenalina A. *An empirical evaluation of supervised learning in high dimensions.* Proceedings of the 25th International Conference on Machine Learning, pp. 96-103. Helsinki, Finland, 2008.

154. Lotte F et al. *A review of classification algorithms for EEG-based brain-computer interfaces.* Journal of Neural Engineering 2007; vol. 4, pp. R1-R13.

155. Benitez JM, Castro JL and Requena I. *Are artificial neural networks black boxes?* IEEE Transactions on Neural Networks 1997; vol. 8, pp. 1156-1164.

156. Dekker A and Champion E. *Please biofeed the zombies: Enhancing the gameplay and display of a horror game using biofeedback.* Proceedings of DiGRA 2007: Situated Play, pp. 550-558. Tokyo, Japan, 2007.

157. Groenegress C, Spanlang B and Slater M. *The physiological mirror: A system for unconscious control of a virtual environment through physiological activity.* The Visual Computer 2010; vol. 26, pp. 649-657.

158. Prinzel III LJ et al. *Three experiments examining the use of electroencephalogram, event-related potentials, and heart-rate variability for real-time human-centered adaptive automation design.* Virginia: NASA Langley Research Center 2003. NASA/TP-2003-212442.

159. Liao W et al. *A decision theoretic model for stress recognition and user assistance.* Twentieth National Conference on Artificial Intelligence (AAAI), pp. 529-534. Pittsburgh, PA, 2005.

160. Bersak D et al. *Biofeedback using an immersive competitive environment.* Online Proceedings for the Designing Ubiquitous Computing Games Workshop, 2001.

161. Nenonen V et al. *Using heart rate to control an interactive game.* CHI 2007, pp. 853-856. San Jose, CA, 2007.

162. Kuikkaniemi K et al. *The influence of implicit and explicit biofeedback in first-person shooter games.* Proceedings of the 28th ACM International Conference on Human Factors in Computing Systems (CHI 2010). New York: ACM, 2010.

163. Wang H, Prendinger H and Igarashi T. *Communicating emotions in online chat using physiological sensors and animated text.* CHI 2004 Extended Abstracts on Human Factors, pp. 1171-1174. New York : ACM, 2004.

164. Oliver N and Kreger-Stickles L. *PAPA: Physiology and purpose-aware automatic playlist generation.* Proceedings of 7th International Conference on Music Information Retrieval, pp. 250-253, 2006.

165. Liu H, Hu J and Rauterberg M. *iHeartrate: A heart rate controlled in-flight music recommendation system.* Proceedings of Measuring Behavior 2010, pp. 265-268. Eindhoven, Netherlands: Noldus Information Technology, 2010.

166. Shye A et al. *Power to the people: Leveraging human physiological traits to control microprocessor frequency.* 41st IEEE/ACM International Symposium on Microarchitecture, pp. 188-199, 2008.

167. Parente A and Parente R. *Mind-operated devices: Mental control of a computer using biofeedback.* Cyberpsychology & Behavior 2006; vol. 9, pp. 1-4.

168. Kamen EW and Su JK. *Introduction to Optimal Estimation.* Springer, 1999.

169. Nagelkerke NJD. *A note on a general definition of the coefficient of determination.* Biometrika 1991; vol. 78, pp. 691-692.

170. Rowe DW, Sibert J and Irwin D. *Heart rate variability: Indicator of user state as an aid to human-computer interaction.* Proceedings of the SIGCHI Conference on Human Factors in Computing systems (CHI '98), pp. 480-487. New York: ACM Press/Addison-Wesley Publishing Co., 1998.

171. Koenig A et al. *Controlling patient participation during robot-assisted gait training.* Journal of NeuroEngineering and Rehabilitation 2011; vol. 8:14.

172. Guerrero CR et al. *Bio cooperative robotic platform for motor function recovery of the upper limb after stroke.* Proceedings of the 32nd Annual International Conference of IEEE EMBC, pp. 4472-4475, 2010.

173. Koenig A et al. *Psychological state estimation from physiological recordings during robot assisted gait rehabilitation.* Journal of Rehabilitation Research and Development 2011; vol. 48, pp. 367-386.

174. Koenig A et al. Real-time closed-loop control of cognitive load in neurological patients during robot-assisted gait training. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2011; vol. 19, pp. 453-464.

175. Bulling A, Roggen D and Tröster, G. *Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments.* Journal of Ambient Intelligence and Smart Environments 2009; vol. 1, pp. 157-171.

176. Zeng Z et al. *A survey of affect recognition methods: Audio, visual and spontaneous expressions.* IEEE Transactions on Pattern Analysis and Machine Intelligence 2009; vol. 31, pp. 39-58.

177. Vidaurre C et al. *Unsupervised adaptation of the LDA classifier for brain-computer interfaces.* Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course, pp. 122-127. Graz, 2008.

Statement / Izjava

I am the author of this dissertation, which came about as a result of my research under the mentorship of prof. dr. Matjaž Mihelj. The assistance of my colleagues is disclosed in the acknowledgments, and already published research by other authors is listed in the References.

Izjavljam, da sem avtor te disertacije, ki je nastala kot plod raziskovalnega dela pod mentorstvom prof. dr. Matjaža Mihlja. Vsa pomoč drugih sodelavcev je izkazana v zahvali. Že objavljeni dosežki drugih avtorjev so navedeni v okviru bibliografije.

Domen Novak

Ljubljana, September 2011.