



LUND UNIVERSITY

Source Data Selection for Brain-Computer Interfaces

Heskebeck, Frida

2026

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Heskebeck, F. (2026). *Source Data Selection for Brain-Computer Interfaces*. [Doctoral Thesis (compilation), Department of Automatic Control]. Department of Automatic Control, Lund University.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Source Data Selection for Brain-Computer Interfaces

FRIDA HESKEBECK

DEPARTMENT OF AUTOMATIC CONTROL | LUND UNIVERSITY



Source Data Selection for Brain-Computer Interfaces

Frida Heskebeck



LUNDS
UNIVERSITET

AKADEMISK AVHANDLING

som för avläggande av Teknologie Doktorsexamen
vid Tekniska fakulteten, Lunds universitet, kommer att
offentligen försvaras fredagen den 23 oktober 2026, kl. 9.15
i sal M:A, M-huset, Lunds Tekniska Högskola.

Handledare: Prof. Bo Bernhardsson, Dr. Carolina Bergeling, and Prof. Tore Häggglund
Opponent: Asst. Prof. Andreea Sburlea, University of Groningen, Groningen, The Netherlands

Department of Automatic Control Lund University Box 118 SE-221 00 Lund Sweden		<i>Document name</i> DOCTORAL DISSERTATION
		<i>Date of issue</i> October 2026
		<i>Document Number</i> TFRT-1154
<i>Author(s)</i> Frida Heskebeck		<i>Supervisor</i> Bo Bernhardsson, Carolina Bergeling, Tore Hägglund
		<i>Sponsoring organisation</i> WASP
<i>Title and subtitle</i> Source Data Selection for Brain-Computer Interfaces		
<i>Abstract</i> A brain-computer interface (BCI) controls a device using brain signals rather than conventional input modalities such as a computer mouse. Before use, the BCI must be calibrated. During the calibration process, the user repeatedly performs mental tasks while electroencephalography (EEG) data is collected. The calibration process is often time-consuming and tiresome for the user. One approach to reduce calibration time is to use transfer learning. With transfer learning, preexisting so-called source data can be used instead of collecting all data during calibration. BCI performance depends on the source data used, and careful source data selection is important for achieving good performance. As more source data becomes available, effective source data selection methods become increasingly important. Another approach to reducing calibration time for motor imagery-based BCIs is early task selection, in which promising motor imagery tasks are identified during calibration so that data is collected only from those tasks. The first aim of the thesis is to <i>present methods for source data selection</i>. Two methods are presented: the transfer performance predictor method and the pole ratio method. Multi-armed bandits are also investigated for source data selection. To understand why BCI performance depends on the source data used, the second aim is to <i>provide insight and intuition into the limitations of transfer learning for BCIs</i>. Two Riemannian geometry-based transfer learning methods, the Riemannian Procrustes analysis and the tangent space alignment, are examined to investigate this. The third and final aim of the thesis is to <i>study how to improve individualized task selection for BCIs</i>. Two methods based on class-mean distance, the direct prediction band and the mean-difference approach, are presented, and multi-armed bandits are investigated for this purpose. The research presented in this thesis can help reduce BCI calibration time, thereby making BCIs more practical to use. The source data selection methods and insights into Riemannian geometry-based transfer learning could be applied to other fields of research as well.		
<i>Key words</i> Brain-computer interface, Machine learning, Riemannian geometry, Source data selection, Transfer learning		
<i>Classification system and /or index terms (if any)</i>		
<i>Supplementary bibliographical information</i>		
<i>ISSN and key title</i> 0280-5316	<i>ISBN</i> 978-91-6858-005-0 (print), 978-91-6858-006-7 (web)	
<i>Language</i> English	<i>Number of pages</i> 177	<i>Recipient's notes</i>
<i>Security classification</i>		

Distribution by Lund University, Department of Automatic Control, Box 118, SE-221 00 Lund, Sweden.

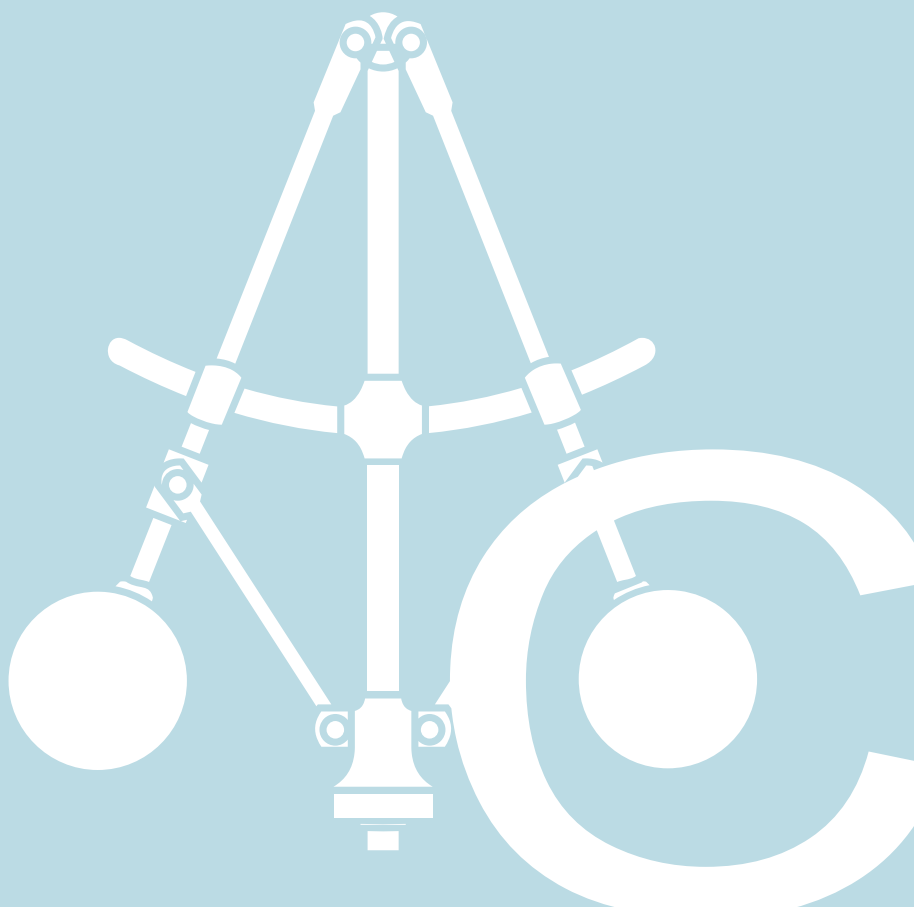
I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature

Date.....

Frida Heskebeck

11/9-2026



Source Data Selection for Brain-Computer Interfaces

Frida Heskebeck



LUND
UNIVERSITY

Department of Automatic Control

Cover illustration by Moa Åkerblom.

Doctoral Thesis TFRT-1154
ISSN 0280-5316

Department of Automatic Control
Lund University
Box 118
SE-221 00 LUND
Sweden

© 2026 by Frida Heskebeck. All rights reserved.
Printed in Sweden by Media-Tryck.
Lund 2026

To my family

Abstract

A brain-computer interface (BCI) controls a device using brain signals rather than conventional input modalities such as a computer mouse. Before use, the BCI must be calibrated. During the calibration process, the user repeatedly performs mental tasks while electroencephalography (EEG) data is collected. The calibration process is often time-consuming and tiresome for the user. One approach to reduce calibration time is to use transfer learning. With transfer learning, preexisting so-called source data can be used instead of collecting all data during calibration. BCI performance depends on the source data used, and careful source data selection is important for achieving good performance. As more source data becomes available, effective source data selection methods become increasingly important. Another approach to reducing calibration time for motor imagery-based BCIs is early task selection, in which promising motor imagery tasks are identified during calibration so that data is collected only from those tasks.

The first aim of the thesis is to *present methods for source data selection*. Two methods are presented: the transfer performance predictor method and the pole ratio method. Multi-armed bandits are also investigated for source data selection. To understand why BCI performance depends on the source data used, the second aim is to *provide insight and intuition into the limitations of transfer learning for BCIs*. Two Riemannian geometry-based transfer learning methods, the Riemannian Procrustes analysis and the tangent space alignment, are examined to investigate this. The third and final aim of the thesis is to *study how to improve individualized task selection for BCIs*. Two methods based on class-mean distance, the direct prediction band and the mean-difference approach, are presented, and multi-armed bandits are investigated for this purpose.

The research presented in this thesis can help reduce BCI calibration time, thereby making BCIs more practical to use. The source data selection methods and insights into Riemannian geometry-based transfer learning could be applied to other fields of research as well.

Acknowledgements

There are some people who have been with me throughout my PhD and the journey leading up to it whom I want to thank especially.

Supervisors

Bo Bernhardsson, thank you for believing in me. I realize that hiring me, with a background significantly different from that of most PhD students in the Department of Automatic Control, must have been a leap of faith. Thank you for sharing your wisdom and for guiding me through these years. Looking back, I have grown a great deal as a person and as a researcher, and much of that is thanks to you.

Carolina Bergeling, you are one of my greatest role models. Having you by my side throughout this PhD has been invaluable, not only for research guidance but also for everything else that comes with being a PhD student, particularly in balancing work and family life with small children at home and in navigating (and surviving in) a male-dominated field.

Tore Hägglund, I did have a rough start, feeling a bit lost as the only person with a biotechnology background in a workplace where most came from mathematics. You were a pillar of support during those early days.

Family

Erik Bergman, my husband, you are the kindest person I know, and I am lucky to spend my life with you. Your endless love and support mean more to me than I can say. I love you!

My children, you bring so much joy and love into my life. I love you!

Jessika Heskebeck, my mum, and **Andreas Heskebeck**, my dad, thank you for all your love and support over the years. I would not be where I am today without you! You are my heroes and role models. I love you!

Johan Heskebeck, my brother, growing up with you gave me countless happy memories, and I am so grateful to have you in my life.

The Bergman family, Lotten, Olle, Moa, Arvid, Ida, Julian, Oskar, Sigge, and all your children. Your family is wonderfully different from the one I grew up in, and I am so grateful that you have welcomed me into it and made me feel at home.

Colleagues

I have had the luxury of working at the Department of Automatic Control, where the work environment is outstanding. Everyone is welcoming, and the room fills with conversation and laughter during fika. Thank you to all of you! Still, there are a few people I would like to give special thanks to.

Pex Tufvesson, what would I do without you? It is rare to find such a good friend, and I am very grateful to have found one in you. We have shared many laughs and thoughts about life. You have also taught me so much.

Ylva Wahlquist, how is it that every third year we realize how well we work together, only to forget about it again? We have so much fun together and so many things in common. I hope we remember that from now on.

Martin Gemborn Nilsson, we have worked together at both the beginning and the end of my PhD, and working with you has been delightful. Our conversations about research have honestly been some of the most stimulating and rewarding I have had as a PhD student.

Emil Vladu, we started our PhDs on the same day, in the same office, and have followed each other ever since. It has always been a pleasure coming to work when you are in the office, and I miss having you as an officemate.

Ask Hällström, you are truly an extraordinary person and an inspiration. If everyone were a little bit more like you, the world would be a better place. I am lucky to have met you.

Julia Adlercreutz, I am so glad you moved to our office. Having you there has made everyday work much more enjoyable. Thank you for all the conversations, laughs, and good company.

Max Nyberg Carlsson, sharing an office with you has been a pleasure. Thank you for the good company and for making the office a nice place to be.

Pontus Giselsson, thank you for trusting me with so much responsibility for the teaching and lectures in the Process Engineering courses. I love teaching, and you let me live that dream.

Eva Westin, the Department of Automatic Control is lucky to have someone who cares so much about the people who work here. Thank you for always being there.

Karl Johan Åström, when I was pregnant with my first child, you told me that there are always more papers to write and deadlines to meet, but time with your children will never come back. Hearing this from you, one of the greatest names in the field of automatic control, was such a comfort. It gave me confidence to prioritize my family when it matters, while still caring deeply about my work.

Finally, thank you to all of you whom I have not yet mentioned but with whom I have shared countless laughs and enjoyable moments during fika: **Alba Gurpegui Ramón, Albin Heimerson, Anders Nilsson, Christian Rosdahl, Emil Sundström, Felix Agner, Frida Nordlund, Johan Lindberg, Johan Ruuskanen, Martin Heyden, Martin Morin, Mattias Fält, Max Nilsson, Nils Vreman, Oskar Stenberg, Talitha Nauta, and Yde Sinnema.**

Friends

Ella Espinoza and **Daniel Espinoza**, when we lived two doors apart, we spent almost every evening together. Now, the logistics of life get in the way. I look forward to watching our children grow up together and to all the fun we will have along the way. Daniel, I am excited about whatever research we find a way to do together.

Moa Åkerblom and **Arvid Åkerblom**, thank you for all the interesting conversations we have shared about life and the universe. I look forward to all the fun our families will have together at Gula huset. I hope we will live closer to each other in the future so that we can share more of those conversations and family moments together. Moa, a special thank you for drawing the cover image.

Emma Kihlberg and **Manfred Klug**, I never laugh as much as when I am with you. Manfred, you were my first friend at LTH. I am glad we found each other.

Caroline Drabe, we have spent many wonderful weekends hiking together, and those trips felt like paradise. Let's do it again when life allows.

Elin Marken Larsson, there are few people I feel so completely on the same wavelength with. We have spent so many hours studying together. I hope we will spend at least as many doing other things together.

Ebba Fogelberg, my oldest friend. Though we now live far apart, I value every chance we get to meet in person and talk.

Teachers

Finally, I want to thank the teachers who, throughout my years in school, have helped me learn, grow, and become the person I am today.

Malin Sjöberg, you encouraged my interest in mathematics in elementary school and gave me a head start in my later studies.

Magnus Ehinger, you sparked my interest in biotechnology, which led me to study for a Master of Science in Engineering in Biotechnology. You are a great example of what it means to be an excellent teacher.

Ulf Ellervik, during my first year of university studies you recruited me for a summer project in your lab. That opportunity meant a lot to me.

Bernt Nilsson, doing my master's thesis with you had a profound impact on me. You are the one who inspired me to pursue a PhD. Although I ultimately did my PhD elsewhere, your belief in me has meant more than you probably realize.

Anton Cervin, during an exam review, you told me something along the lines of, "The Department of Automatic Control is always looking for brilliant people like you. You should consider doing a PhD here." If you had not said that, I would never have considered the Department of Automatic Control.

Martin Morin, you spent so many hours patiently answering my questions. I learned an incredible amount from you, both within and beyond the course material.

Financial support

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. The author is a member of the ELLIIT Strategic Research Area.



LUND
UNIVERSITY

Towards Faster Startup of Brain-Computer Interfaces

Frida Heskebeck

Department of Automatic Control

Popular science summary of the PhD thesis *Source Data Selection for Brain-Computer Interfaces*, October 2026. The thesis can be downloaded from: <http://www.control.lth.se/publications>.

A brain-computer interface (BCI) is a device controlled by brain signals. Brain signals are measured using an electroencephalography (EEG) cap placed on the user's scalp as shown in the figure. BCIs can be used for medical applications such as controlling a prosthetic arm or for entertainment applications such as controlling video games.

For a BCI to work, it needs to be fine-tuned to the user's brain. This fine-tuning is performed in a process called calibration. During calibration, the user performs specific mental tasks while EEG data is recorded. The BCI then learns to recognize the patterns in the EEG data for each mental task. The mental tasks are mapped to commands for the application, for example, moving the cursor on a screen to the left when the user imagines moving the left hand. For the BCI to fully recognize the patterns in the EEG data, the user must repeat the mental tasks multiple times. This results in calibration times of around 20 minutes before the BCI can be used. Doing this calibration is tiresome and boring for most users.

For medical applications, long calibration times are generally acceptable, as the additional time is worth it to achieve good BCI performance. In entertainment applications, on the other hand, a short calibration time is often more important than perfect final performance, since few want to use an entertainment device with a 20-minute startup time. For both applications, reducing calibration time is important to improve the user experience. Thus, much research focuses on this.

One approach to reduce the calibration time is to reduce the number of times the mental tasks need to be repeated. This can be done by using already recorded data



A person wearing an EEG cap that measures the brain signals used in BCIs.

from someone else. Using data from one person to train the BCI for another is a form of so-called transfer learning. Formally, data from a new BCI user is called target data, and preexisting data is called source data. A problem with transfer learning is that the source and target data must be sufficiently similar for it to work. The question is then: How can we identify which source data is similar to the target data? In this thesis, two methods for source data selection are presented. One is the transfer performance predictor method, which is particularly useful in settings with a lot of available source data. The other is the pole ratio method, which evaluates how well source and target data are aligned. The thesis also studies so-called multi-armed bandits for source data selection. To understand why data from some source users work better than data from others, the thesis also studies the limitations of transfer learning for two well-known transfer learning methods for BCIs: the Riemannian Procrustes analysis and the tangent space alignment.

Another approach to reduce the number of times the mental tasks need to be repeated is to identify which mental tasks yield good performance and collect data only from those. For example, some users may perform better when imagining right-versus left-hand movements, while others may perform better when imagining foot versus hand movements. The thesis presents two methods for task selection: the direct prediction band and the mean-difference approach. Multi-armed bandits are also investigated for task selection.

With the methods and results of the thesis, source data and mental tasks can be selected more efficiently, thereby reducing calibration times while maintaining good BCI performance. This will allow for faster BCI startup and, in the long run, we will likely see improvements in both medical and entertainment applications.



LUNDS
UNIVERSITET

Mot Snabbare Uppstart av Brain-Computer Interfaces

Frida Heskebeck

Institutionen för Reglerteknik

Populärvetenskaplig sammanfattning av doktorsavhandlingen *Source Data Selection for Brain-Computer Interfaces*, oktober 2026. Avhandlingen kan laddas ner från: <http://www.control.lth.se/publications>

Ett så kallat brain-computer interface (BCI) är en apparat som styrs av hjärnsignaler. Hjärnsignalerna mäts genom elektroencefalografi (EEG) med hjälp av en mössa som är placerad på användarens huvud som figuren visar. BCI:er kan användas för medicinska applikationer som att styra en protes eller för underhållningsapplikationer som att styra tv-spel.

För att en BCI ska fungera behöver den finjusteras till användarens hjärna. Denna finjusteringsprocess kallas kalibrering. Under kalibreringen utför användaren specifika mentala uppgifter samtidigt som EEG-data samlas in. BCI:n lär sig sedan att känna igen mönstren i EEG-datan för respektive mental uppgift. De mentala uppgifterna kopplas till kommandon för applikationen, till exempel att muspekaren på en skärm ska flyttas till vänster när användaren föreställer sig att den rör på vänster hand. För att BCI:n fullt ut ska kunna lära sig mönstren i EEG-datan måste användaren upprepa de mentala uppgifterna flera gånger. Detta leder till kalibreringstider kring 20 minuter innan BCI:n kan användas. Att göra denna kalibrering är tröttsamt och tråkigt för de flesta användare.

För medicinska applikationer är långa kalibreringstider generellt acceptabla eftersom den förbättrade prestandan är värd den extra kalibreringstiden. För underhållningsapplikationer är å andra sidan en kort kalibreringstid ofta viktigare än perfekt prestanda eftersom få vill använda en underhållningsapplikation med en 20 minuter lång uppstartstid. För både medicinska och underhållningsapplikationer är det viktigt att reducera kalibreringstiden för att förbättra användarupplevelsen. Därför fokuserar mycket forskning på detta.

Ett tillvägagångssätt för att minska kalibreringstiden är att minska antalet gånger en mental uppgift måste upprepas. Detta kan göras genom att använda tidigare insam-



En användare bär en EEG-mössa som mäter de hjärnsignaler som styr BCI:n

lad data från en annan användare. Att träna BCI:n för en ny användare på data från en tidigare användare är en form av så kallad transfer learning. Formellt kallas datan från en ny BCI-användare target-data och datan från en tidigare användare source-data. Ett problem med transfer learning är att source- och target-datan måste vara tillräckligt lika för att det ska fungera. Frågan är då: Hur kan man identifiera vilken source-data som liknar target-datan? I denna avhandling presenteras två metoder för så kallad source data selection. Den ena är den så kallade transfer performance predictor-metoden, vilken är speciellt användbar när det finns mycket source-data att välja från, och den andra är den så kallade pole ratio-metoden, vilken utvärderar den geometriska relationen mellan source- och target-datan. I avhandlingen studeras även så kallade multi-armed bandits för source data selection. För att förstå varför viss source-data är mer lämpad för en given target-användare studerar denna avhandling också begränsningarna för transfer learning för två transfer learning-metoder kallade Riemannian Procrustes analysis och tangent space alignment.

Ett annat tillvägagångssätt för att minska antalet gånger som en mental uppgift behöver upprepas är att identifiera vilka mentala uppgifter som ger bra prestanda och enbart samla data från dessa mentala uppgifter. Till exempel, vissa användare får högre prestanda om de föreställer sig att de rör höger och vänster hand, medan andra användare får högre prestanda om de föreställer sig att de rör fötterna och händerna. I avhandlingen presenteras två metoder för uppgiftsselektion kallade direct prediction band och mean-difference approach. Multi-armed bandits undersöks också för uppgiftsselektion.

Med metoderna och resultaten i denna avhandling kan source data selection och uppgiftsselektion utföras mer effektivt, och därmed kan kalibreringstiden för BCI:er minskas medan prestandan bevaras. Detta kommer att medföra snabbare uppstart av BCI:er och i det långa loppet kommer vi troligtvis se förbättringar för både medicinska och underhållsapplikationer.

List of abbreviations

Abbreviation	Meaning
AIR	Affine-invariant Riemannian
BCI	Brain-computer interface
CSP	Common spatial pattern
EEG	Electroencephalography
ERP	Event-related potential
ErrP	Error-related potential
HPD	Hermitian positive definite
ISA	Intra-subject accuracy
LDA	Linear discriminant analysis
LogReg	Logistic regression
MAB	Multi-armed bandit
MDM	Minimum distance to mean
MI	Motor imagery
MOABB	Mother of all BCI benchmarks
P300	Positive potential occurring approximately 300 ms after a stimulus
RCT	Recentering
RMSE	Root mean square error
RPA	Riemannian Procrustes analysis
SCM	Sample covariance matrix
SVD	Singular value decomposition
SVM	Support vector machine
SPD	Symmetric positive definite
SSVEP	Steady-state visually evoked potential
TPP	Transfer performance predictor
TSA	Tangent space alignment
UCB	Upper confidence bound

Contents

Abstract	i
Acknowledgements	iii
Popular summary in English	vii
Populärvetenskaplig sammanfattning på svenska	ix
1. Introduction	1
1.1 Research aims	2
1.2 Who benefits from this research?	3
1.3 Layout of thesis	3
2. Understanding BCIs and their challenges	4
2.1 Measuring brain activity with EEG	4
2.2 Controlling BCIs with imagined movement is harder than it sounds	6
2.3 BCI calibration from a machine learning perspective	7
2.4 BCI development raises ethical questions	8
3. Overview of the methods used	10
3.1 Covariance matrices represent EEG epochs	10
3.2 Riemannian geometry provides a framework for BCI analysis . .	11
3.3 MDM and LogReg are simple but powerful BCI classifiers . . .	14
3.4 Visualization of the RPA and TSA transfer learning methods . .	15
4. Contributions	20
4.1 Included papers	20
4.2 Additional papers	23
5. Discussion of key findings	25
5.1 Selecting source data is not a trivial task	25
5.2 Understanding transfer learning limitations is a first step toward further development	29
5.3 Individualized task selection can reduce calibration time in MI- based BCIs	34

5.4	Cloud computing creates new opportunities for BCI development	37
6.	Conclusions	38
6.1	Research aims revisited	38
6.2	Future work	39
7.	Final words	41
	Bibliography	42
	Paper I. Source Data Selection for Brain-Computer Interfaces based on Simple Features	47
1	Introduction	48
2	Related work	50
3	Background	51
4	Material and Method	54
5	Results and Discussion	61
6	Conclusions and Future Research	64
7	Summary	69
	References	69
	Paper II. Rotation-Based Metric on the Riemannian Manifold of SPD Matrices with Applications to Source Data Selection for Brain-Computer Interface Transfer Learning	73
1	Introduction	74
2	Related works	75
3	Materials and Methods	77
4	Results	84
5	Discussion	88
6	Conclusion	93
	References	95
	Paper III. Class Mean Distance for Early Classification Accuracy Estimation	99
1	Introduction	100
2	Materials and methods	100
3	Results	103
4	Discussion	105
5	Conclusion	106
	References	107
	Paper IV. Comparison and extension of the Riemannian Procrustes analysis and tangent space alignment transfer learning methods	109
1	Introduction	111
2	Methods	112

3	Results	122
4	Discussion	128
5	Conclusion	135
A	Appendix: Mathematical details	138
	References	140
	Paper V. Multi-Armed Bandits in Brain-Computer Interfaces	145
1	Introduction	146
2	Multi-Armed Bandits theory - a crash course	146
3	Current use of Multi-Armed Bandits in Brain-Computer Interfaces	149
4	Discussion of Future use of Multi-Armed Bandits in Brain- Computer Interfaces	151
5	Getting started with multi-armed bandits in BCI research	153
6	Conclusion	153
	References	154

1

Introduction

Imagine that you are sitting at your computer. Instead of using the keyboard and mouse to control the computer, you use your *thoughts*. Such a thought-controlled system is known as a brain-computer interface (BCI). For you, the BCI is mostly a cool gadget you use for fun, but you once saw a paralyzed person control their wheelchair with one. They seemed so happy and free. That made you realize that BCIs can, for some people, be the difference between a life in isolation and a life among family and friends. You are very impressed by this futuristic technology, but two things annoy you.

The first annoying thing is that you have to *calibrate* the BCI system each time you use it. That means the BCI system needs to collect data from you as you perform specific mental tasks, which it learns to recognize. You could accept a short calibration time, but not the 20 minutes it takes now. The second thing that annoys you is that on days when you are tired or emotional, the BCI performs worse than on days when you are happy. The instructions state that you should remain calm and still while using the BCI. You do your best to follow these instructions, but the variability of performance is upsetting you. You strongly suspect that the mental tasks used to control the BCI are more suitable for you on some days, whereas a different set of mental tasks would be better on other days.

The BCI challenges presented above are two of the main reasons why there are few BCI applications available. A BCI needs to be carefully tuned to the user, and this so-called calibration process takes time. For medical applications, carefully selecting mental tasks and collecting large amounts of data are acceptable, since the positive impact on the user's life with a well-functioning BCI is high. For entertainment-oriented applications, on the other hand, it is often more important to have a quick BCI startup than to have perfect performance, as these devices are used more for fun than for practical purposes. Imagine this: if you were paralyzed and a BCI were the only way you could communicate with your family, you would happily be doing a 20-minute calibration just to say hello. If you, on the other hand, are playing a

video game, you don't want to do a 20-minute calibration before you start playing. Regardless of the application and tolerance for long calibration times, reducing calibration time is important for user satisfaction and the future development of BCIs.

One approach to reducing the calibration time is to use preexisting data from another user. Then, less data from the new user needs to be collected. Using data from another person in this way is a form of so-called *transfer learning*. In terms of BCIs and transfer learning, a user is often referred to as a *subject*. The data from a new subject is called *target data*, and preexisting data is called *source data*. The source data can either be data from another subject or from a previous session of the same subject. When doing transfer learning, it is important that the source and target data are compatible. Otherwise, BCI performance will be poor. Thus, efficient methods for *source data selection* are important for the future development of BCIs. To design efficient source data selection methods, it is important to understand the limitations of transfer learning and why some source data yield better performance than others.

Another approach to reduce calibration time is to detect early which mental tasks yield good performance and collect data only from those tasks, rather than collecting data from all tasks and then, after calibration, deciding which to use. In this thesis, this is referred to as *task selection*, which should not be confused with source data selection described previously.

1.1 Research aims

This thesis aims to take a step toward the next generation of BCIs by addressing the challenges for BCIs presented above. The central theme is how calibration time for BCIs can be reduced, but specifically, the thesis sets out to:

- Present methods for source data selection.
- Provide insight and intuition into the limitations of transfer learning for BCIs.
- Study how to improve individualized mental task selection for BCIs.

These research aims are connected to each other and overlap as visualized in [Figure 1.1](#).

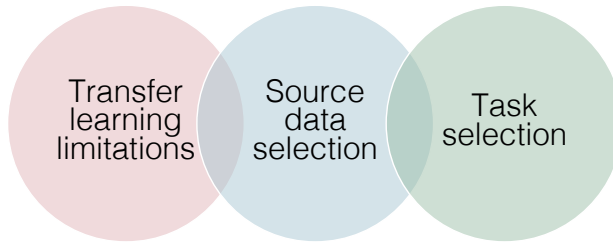


Figure 1.1 Aims of the thesis.

1.2 Who benefits from this research?

Research on BCIs will ultimately improve BCI performance, thereby affecting BCI end users. As of today, many BCI applications are medical, with examples such as prosthetic limbs and communication devices. Some emerging consumer-grade applications exist. These are often aimed at entertainment, for example, meditation aids and control for video games [Nam et al., 2018; Yadav and Maini, 2025]. Improving BCIs for medical applications can be life-changing for end users, while entertainment-oriented applications will be more enjoyable. Regardless of the application, pushing the limits of what BCIs can achieve will expand the field’s knowledge, enabling further research and more advanced applications.

The thesis focuses on BCI data, but the source data selection methods are relevant to any field that requires source data selection, and a deeper understanding of the transfer learning methods studied is relevant to other fields where similar methods are used. Thus, researchers across different fields may also benefit from the work presented in this thesis.

1.3 Layout of thesis

The thesis is a compilation thesis, meaning the core of the thesis lies in the included papers. They can be found at the end of the thesis. [Chapters 2](#) and [3](#) provide background to BCIs, the problem formulation, and the main methods that are used in the papers. [Chapter 4](#) presents an overview of the included papers with a short summary of the key findings for each. [Chapter 5](#) discusses the contributions in the papers and highlights interesting aspects and results. Finally, [Chapters 6](#) and [7](#) conclude the thesis. I hope you will enjoy reading the thesis as much as I have enjoyed writing it.

2

Understanding BCIs and their challenges

A BCI uses brain signals as input to the system. In a typical setup, a subject performs one of many predefined mental tasks, and the BCI decodes which task the subject is performing. Each mental task is translated to a command for the device. For example, if the subject is imagining moving the right hand, the cursor on a screen should move to the right.

This chapter provides background on BCIs and the BCI challenges addressed in the thesis. The next chapter includes details on the methods and algorithms used in the thesis.

2.1 Measuring brain activity with EEG

The first step for a BCI to work is to measure the brain signals while the subject performs mental tasks. There are many ways to do that, but electroencephalography (EEG) is most often used. In EEG, the brain's electrical activity is detected by electrodes placed on the scalp and is typically measured in microvolts. Since the electrodes are placed on the skin's surface, EEG is considered a non-invasive measurement method. The brain's electrical activity is measured multiple times per second, often at 250–500 Hz, which means that EEG recordings have a high temporal resolution. The electrodes, also called EEG channels, are placed in a predefined pattern on the scalp called the 10–20 system, as shown in [Figure 2.1 \[Nicolas-Alonso and Gomez-Gil, 2012; Nam et al., 2018\]](#).

There is a lot of noise in EEG data, and it needs to be preprocessed to extract the interesting information. [Figure 2.2](#) shows an example of seven seconds of EEG data measured with ten electrodes. The first step of preprocessing is to remove the noise from the data. To do so, a bandpass filter is applied. The filter's frequency

2.1 Measuring brain activity with EEG

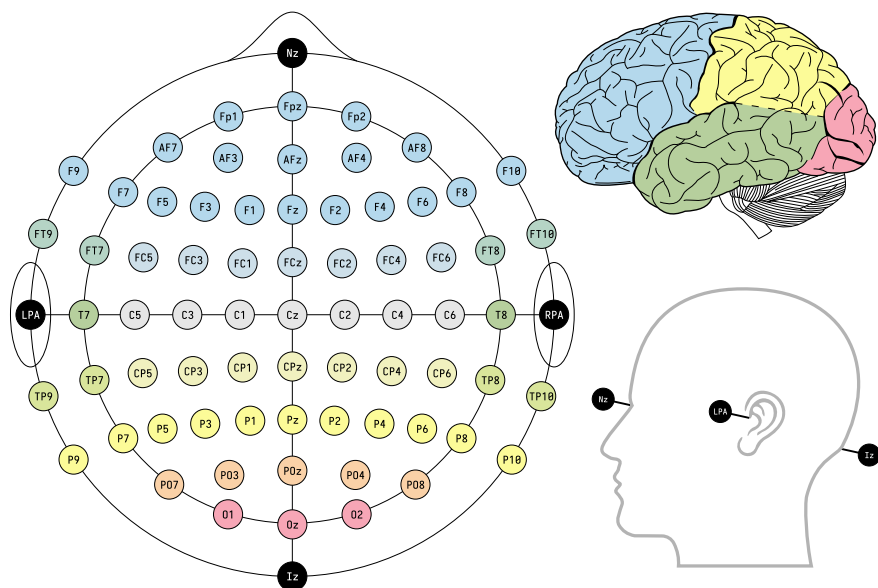


Figure 2.1 EEG electrode placements according to the 10–20 system. [Figure by Laurens R. Krol - Own work, CC0, [Wikimedia Commons](#).]

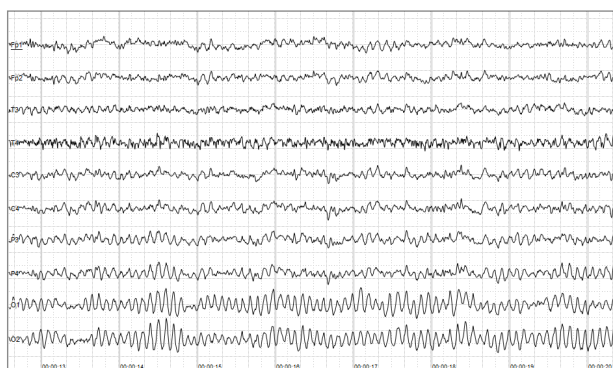


Figure 2.2 Seven seconds of EEG recordings from ten electrodes measured in microvolts. [Figure by Andrii Cherninskyi - Own work, CC BY-SA 4.0, [Wikimedia Commons](#).]

range depends on the mental task performed, as different tasks elicit brain activity at different frequencies. The second step of preprocessing is to select only the EEG channels where the relevant brain activity is expected to occur, since different brain regions are involved in different types of mental tasks. For example, visual stimuli are processed at the back of the head, in the red area of Figure 2.1, while movement-related activity is processed near the top of the head, between the blue and yellow areas of Figure 2.1. The final step of preprocessing is to cut the EEG data into so-called epochs. Each epoch is a short interval of EEG data, typically a few seconds, during which a single mental task is performed. I want to emphasize that the BCI preprocessing presented above is the bare minimum. Additional preprocessing steps can be included. The interested reader is referred to Blankertz (2018) and Nicolas-Alonso and Gomez-Gil (2012) for more details.

Once the preprocessing is completed, the EEG data for an epoch i with n_{ch} channels and n_s time samples is

$$X_i \in \mathbb{R}^{n_{ch} \times n_s}.$$

The mental task performed during an epoch is referred to as the epoch's label and is denoted by y_i . An epoch and its label form a pair (X_i, y_i) , and a BCI dataset consists of hundreds of such pairs with multiple repetitions of each label.

Since EEG-based BCIs are the most common type of BCI, I use only EEG data in my thesis. I perform standard preprocessing in accordance with previous research. I use data from multiple subjects and apply the same preprocessing to all of them. I do not investigate whether alternative preprocessing would be better for any individual subject. This is because a central theme of the thesis is how data from different subjects can be used together to improve performance, and a shared preprocessing pipeline is beneficial as it makes the data more comparable across subjects. The exact preprocessing pipeline is detailed in each of the included papers. I use only publicly available data, accessed via the open-source MOABB project [Aristimunha et al., 2025; Jayaram and Barachant, 2018]. This means that no subject data was collected in this thesis.

2.2 Controlling BCIs with imagined movement is harder than it sounds

As previously introduced, a BCI is controlled by the subject performing a mental task. Some special types of mental tasks, often referred to as BCI paradigms, are commonly used because they generate brain activity that is well-known and detectable in EEG data. One such paradigm is motor imagery (MI), in which the imagined movement of a body part is mapped to a command. For example, if the

2.3 BCI calibration from a machine learning perspective

subject imagines moving their left or right hand, it could be mapped to moving the cursor on a screen left or right. The set of possible MI tasks and their mappings to commands is defined before the BCI is used. Other BCI paradigms include error-related potentials (ErrPs), P300 event-related potentials (ERPs), and steady-state visually evoked potentials (SSVEPs) [Nicolas-Alonso and Gomez-Gil, 2012; Nam et al., 2018]. In this thesis, I focus on the MI paradigm because it presents interesting challenges for transfer learning.

When a subject imagines moving a body part, oscillations in the EEG signal change in specific brain regions. For example, right-hand movements can normally be detected in the C3 electrode, left-hand movements in the C4 electrode, and foot movements in the Cz electrode [Pfurtscheller and Neuper, 2010]. Performing an MI task requires practice for the subject, as imagining movement can be challenging for some people. Because of this, people have slightly different approaches to imagining the movement, so their EEG data differ slightly. This is one of the reasons the MI paradigm is interesting to study: brain activity varies more across subjects than it does in other BCI paradigms, making transfer learning between subjects less straightforward. Additionally, some subjects yield good data for certain classes, for example, left- versus right-hand movements, whereas these classes are indistinguishable for other subjects, motivating the need for task selection during BCI calibration.

2.3 BCI calibration from a machine learning perspective

At the core of a BCI is a machine learning algorithm that classifies the EEG data. To do so, the machine learning algorithm must know the distinguishing patterns in the data, which it learns from so-called *training data*. The training data consist of EEG epochs with known labels. The patterns in EEG data generated from MI tasks are subtle and relatively hard to classify. Generally, the more training data available, the better the machine learning algorithm learns the patterns, and thus the better the BCI performance becomes. To collect sufficient training data, the subject must perform the same mental task multiple times during calibration, leading to long calibration times.

In a real-time BCI, after calibration, the machine learning algorithm is used to classify EEG epochs with unknown labels. The predicted labels are then used to control the BCI. In this thesis, I evaluate BCI performance using so-called *test data*, which is target data with known labels not used to train the machine learning algorithm. The general term *BCI performance* refers to the classification accuracy on target test data, which reflects how well the BCI decodes the mental task performed by the subject. BCI performance is an umbrella term for the more specific terms

transfer learning performance and *intra-subject accuracy*. Transfer learning performance refers to the classification accuracy on the target test data after the machine learning algorithm is trained on source data that has been transformed by a transfer learning method. Intra-subject accuracy refers to the classification accuracy on the target test data when the machine learning algorithm is trained only on the target training data, without any transfer learning involved.

When transfer learning is used to reduce the calibration time, some target data is replaced with source data to train the underlying machine learning algorithm. One approach to transfer learning is to align the source and target data. This is known as *domain adaptation*. Another approach is to adapt a model trained on the source data to the target data. This is known as *rule adaptation* [Jayaram et al., 2018]. This thesis focuses on domain adaptation where data from one source subject is aligned to data from one target subject. However, the presented source data selection methods could also be applied in a rule adaptation setting. In such a setting, the premise is that specialized models trained on large amounts of source data already exist for different subject types, and the task is to identify the subject type of a new target subject in order to select a suitable model.

The EEG epochs need to be transformed to extract the relevant features for the machine learning algorithm. A traditional classification pipeline uses common spatial pattern (CSP) analysis followed by a linear discriminant analysis (LDA) classifier [Blankertz, 2018]. Barachant et al. (2012) introduced Riemannian geometry for classification, which has since been the leading approach [Chevallier et al., 2024]. Much of my work is thus based on Riemannian geometry, with the relevant background provided in Chapter 3. There are many domain adaptation transfer learning methods [Wan et al., 2021; Ko et al., 2021]. In the thesis, I use the Riemannian geometry-based methods Riemannian Procrustes analysis [Rodrigues et al., 2019] and tangent space alignment [Bleuzé et al., 2022]. Further details on those are also given in Chapter 3.

2.4 BCI development raises ethical questions

The first question I often get after saying that I work with BCIs is whether I can read the person's mind. The question reflects the person's fear of losing control of their most private possession: their thoughts. As of today, we cannot read a person's mind with BCIs, only specific patterns, but who knows what can be done in the future? In this section, I discuss some ethical considerations in BCI research. For a thorough review of the risks and ethical aspects of BCIs, the reader is referred to Burwell et al. (2017) and Cao et al. (2026).

A major concern with the development of BCIs is data privacy. As of today,

2.4 BCI development raises ethical questions

it is not possible to decode *all* thoughts and information about a person from the EEG data. However, with improved methods and understanding of the brain, more information might be gained in the future. A question with source data is how it should be used, or if it should be available at all. Should it be possible to download raw EEG data from the cloud, even if it is anonymized? For BCI data, the discussion is similar to that for other medical and personal data: who should have access, and for what purposes should the information gained from it be used?

Another ethical consideration for BCI research is data bias. Using diverse data in research is important to avoid ending up with BCIs that only work for healthy white men. I do want to emphasize that, compared to many other fields of research, BCI data is fairly diverse in terms of health since much of the BCI research is aimed at medical applications. However, as more datasets become available, it is important to keep diversity in mind.

As with many new technologies, BCIs may reach the rich users first. Particularly for medical applications, it is important not to disadvantage any region or ethnic group. With more research, BCI devices will hopefully become cheaper and more widely available.

When BCIs become more integrated into our lives, the question of responsibility arises. Say that the BCI is used to control a wheelchair. Who is responsible if the wheelchair crashes? Is it the person using the BCI or those who designed it? This discussion is similar to that for self-driving cars, except that with BCIs, there is user input. Still, the engineers creating the BCI have a big responsibility for how well it understands the user's intentions.

A final ethical aspect of BCIs that I find interesting is that using BCIs might change human behavior. Say that you are connected to a BCI and it is controlled with the MI paradigm as described previously. Then you might discover that instead of imagining lifting your right arm, it is better to imagine frantically waving it. Then the BCI has affected how you think; it has changed part of your behavior. This example might seem harmless, but it is not hard to imagine more invasive examples. For example, if you are often connected to a BCI, you might start to get paranoid, wondering what thoughts can actually be decoded by the BCI. And who can access this data? This risk of a Big Brother society might affect what you think while wearing the BCI, ultimately influencing your behavior.

When raising ethical concerns about BCIs, it is a slippery slope down to all the negative outcomes. But one has to remember the good things that can come of BCIs. As of today, BCIs for medical applications have helped paralyzed people communicate with their loved ones and helped people with prosthetic limbs recover some functionality. This is what I am doing research for: helping others. Still, we need to carefully consider the risks and drawbacks of the technology and its applications so as not to create something we will later regret.

3

Overview of the methods used

Now that we understand the basics of BCIs and the challenges of their calibration, it is time to move on to more details about the research I have performed. In this chapter, I provide an overview of the Riemannian framework for the analysis of BCI data. I give an introduction to the Riemannian Procrustes analysis (RPA) [Rodrigues et al., 2019] and the tangent space alignment (TSA) [Bleuzé et al., 2022] transfer learning methods. I do my best to visualize these concepts and methods to complement the mathematical presentations.

3.1 Covariance matrices represent EEG epochs

After the data is preprocessed and split into epochs, X_i , as described in Chapter 2, the epochs must be transformed into suitable features before they can be classified by a machine learning algorithm. For MI data, the epoch-wise covariance matrix is often used as such a feature, since it captures correlations between electrodes and reflects differences in brain activity between regions, which is important information for MI BCIs. There are other possible feature descriptions of the epochs, such as time-frequency patterns, but these are less commonly used for MI data. Thus, I consider only the covariance matrix representation of epochs in this thesis.

The sample covariance matrix, $C_i \in \mathbb{R}^{n_{ch} \times n_{ch}}$, for an epoch, $X_i \in \mathbb{R}^{n_{ch} \times n_s}$, with n_{ch} channels and n_s time samples, is calculated as

$$C_i = \frac{1}{n_s} X_i X_i^\top.$$

3.2 Riemannian geometry provides a framework for BCI analysis

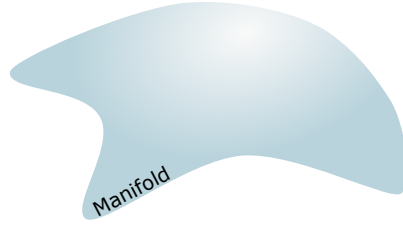


Figure 3.1 Visualization of a Riemannian manifold.

Hereafter, I will write n instead of the channel-count notation n_{ch} and say the covariance matrix is of size $n \times n$. After the EEG data is transformed into covariance matrices, the dataset consists of pairs (C_i, y_i) which are used for further analysis.

The covariance matrices are symmetric positive definite (SPD), or are regularized to be SPD, which means they have certain properties, such as all eigenvalues being positive. It has been shown that analyzing covariance matrices on a specific Riemannian manifold yields superior classification performance compared with analyzing them in Euclidean space [Barachant et al., 2012]. Using the covariance representation of MI data and analyzing them with Riemannian geometry is considered state-of-the-art, which is why I focus on this in my thesis [Chevallier et al., 2024].

3.2 Riemannian geometry provides a framework for BCI analysis

When comparing covariance matrices in Euclidean geometry, some important features of the covariance matrices are lost. This is why Riemannian geometry is used [Barachant et al., 2012; Yger et al., 2017]. In this thesis, the covariance matrices are analyzed on a Riemannian manifold equipped with the affine-invariant Riemannian (AIR) metric. The AIR metric defines how to compare covariance matrices, giving a way to measure distances between them. More formally, it defines the local inner product on the manifold of SPD matrices, from which geodesic distances (shortest paths) can be derived. The Riemannian manifold equipped with the AIR metric is denoted $\mathcal{P}_n$ and is hereafter referred to as the *AIR SPD manifold*. The n in $\mathcal{P}_n$ refers to the size of the $n \times n$ covariance matrices. One often visualizes a manifold as a curved surface, as shown in Figure 3.1. I will not go into all the details of a Riemannian manifold in this thesis, nor derive all equations, but will focus on providing an overview that gives sufficient insight to understand my research. The interested reader is referred to Barachant et al. (2012) and Yger et al. (2017) and references therein for more details.

Chapter 3. Overview of the methods used

The distance (defined by the AIR metric) between two covariance matrices on the AIR SPD manifold, $C_i, C_j \in \mathcal{P}_n$, is

$$\delta_R(C_i, C_j) = \left\| \log(C_i^{-1/2} C_j C_i^{-1/2}) \right\|_F = \sqrt{\sum_{k=1}^n \log^2(\lambda_k)}, \quad (3.1)$$

where $\log$ denotes the matrix logarithm in the first equality and the scalar (standard) logarithm in the second. λ_k , where $k = 1, \dots, n$, are the eigenvalues of the matrix $C_i^{-\frac{1}{2}} C_j C_i^{-\frac{1}{2}}$. I will hereafter refer to (3.1) as the *Riemannian distance*.

The Riemannian distance has some interesting properties [Rodrigues et al., 2019]. One that is of particular interest for BCIs is that it is invariant to affine transformations with any invertible matrix $A \in \mathbb{R}^{n \times n}$, meaning that

$$\delta_R(C_i, C_j) = \delta_R(AC_i A^\top, AC_j A^\top).$$

This means that if two covariance matrices are linearly transformed by the same matrix A , the distance between them remains unchanged. This can be interpreted as follows. Suppose two covariance matrices are computed from EEG data recorded during a session and their distance is measured. If, in another session, the same underlying brain activity is recorded using a slightly different channel representation, for example due to a different electrode ordering or an invertible linear mixing of the channels, then this change can be represented as an affine transformation of the covariance matrices. This means that the distance between two covariance matrices in the second session is the same as in the first session. Intuitively, it makes sense that the distance is the same in both sessions since the same underlying brain activity is compared in both.

A dataset consists of many labeled covariance matrices, each forming the pair (C_i, y_i) . All covariance matrices with the same label $y_i = k$ can be grouped into one class set denoted $C_k = \{C_i \in \mathcal{P}_n \mid y_i = k\}$. The mean covariance matrix for each class set, called a *class mean* and denoted $M_k \in \mathcal{P}_n$, is used to represent the covariance matrices within each class set in classification and transfer learning algorithms. The class mean of a class set is the matrix on the manifold that minimizes the sum of squared Riemannian distances to the class set's covariance matrices. It is calculated as

$$M_k = \arg \min_{Q \in \mathcal{P}_n} \sum_{C_i \in C_k} \delta_R^2(Q, C_i). \quad (3.2)$$

The class mean, M , of an example dataset is visualized in Figure 3.2.

The dispersion, σ , of a dataset $C = \{C_i \in \mathcal{P}_n\}$ around its mean M is defined as

$$\sigma^2 = \frac{1}{|C|} \sum_{C_i \in C} \delta_R^2(M, C_i), \quad (3.3)$$

3.2 Riemannian geometry provides a framework for BCI analysis

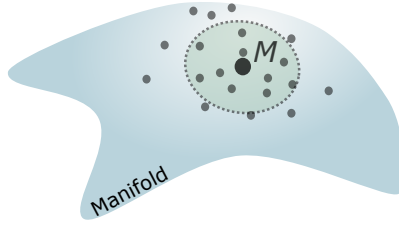


Figure 3.2 Visualization of class mean (M) and dispersion (green area) for an example dataset on a Riemannian manifold.

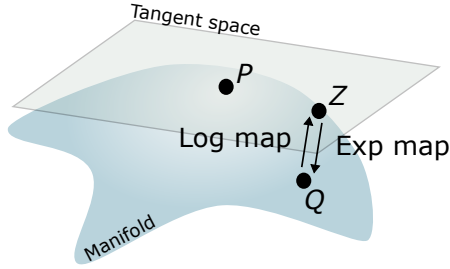


Figure 3.3 Visualization of projection between a Riemannian manifold and its tangent space.

where $|C|$ is the number of covariance matrices in the set C . The dispersion of an example dataset is visualized in Figure 3.2 as the green area.

A final important feature of the Riemannian geometry framework for this thesis is that data can be projected onto a tangent space of the manifold and then back again, as visualized in Figure 3.3. On the AIR SPD manifold, the points are SPD matrices, while the corresponding points in the tangent space are symmetric matrices. Each point on the manifold has a tangent space defined by the AIR metric, and the tangent space at a point $P \in \mathcal{P}_n$ is denoted $T_P \mathcal{M}$. The point P is, in almost all cases, the identity matrix in my research. The projection from the tangent space to the manifold is done with the so-called exponential map, and from the manifold to the tangent space with the so-called logarithmic map. These maps are defined for the tangent space at point P , mapping the points $Q \in \mathcal{P}_n$ and $Z \in T_P \mathcal{M}$, as

$$\text{Exp}_P(Z) = P^{\frac{1}{2}} \exp \left(P^{-\frac{1}{2}} Z P^{-\frac{1}{2}} \right) P^{\frac{1}{2}}$$

and

$$\text{Log}_P(Q) = P^{\frac{1}{2}} \log \left(P^{-\frac{1}{2}} Q P^{-\frac{1}{2}} \right) P^{\frac{1}{2}},$$

where $\exp$ and $\log$ denote the matrix exponential and matrix logarithm, respectively.

After projection onto the tangent space, the matrices are often vectorized into a vector $V \in \mathbb{R}^{n(n+1)/2}$ for computational reasons. This means they are stacked as vectors rather than matrices. The matrices are symmetric, so only the upper triangular part, including the diagonal, is stacked as detailed in Paper IV [Gemborn Nilsson and Heskebeck, 2026]. When expressed as vectors, all standard machine learning methods developed for Euclidean space can be applied, such as logistic regression, support vector machines (SVMs), and neural networks.

3.3 MDM and LogReg are simple but powerful BCI classifiers

The overall task of the BCI is to predict the class of a new observation. During classifier training, the pairs (C_i, y_i) are available. As a reminder, $C_i \in \mathcal{P}_n$ is the covariance matrix of the epoch and y_i is the class label. The available target data is split into training and test sets. The training set is used for transfer learning and classifier training, and the test set is used to evaluate performance. There are many methods to classify EEG data [Lotte et al., 2018]. In this thesis, I mostly use minimum distance to mean (MDM) based on the Riemannian distance, but also logistic regression (LogReg).

With an MDM classifier, a new observation, $C_{\text{new}} \in \mathcal{P}_n$, is assigned to the class with the closest class mean $M_k \in \mathcal{P}_n$ as

$$y_{\text{new}} = \arg \min_{k \in \{1, \dots, K\}} \delta_R(M_k, C_{\text{new}}).$$

The MDM classifier is visualized in Figure 3.4, where every point on the manifold in the pink area is assigned to class 1, and every point in the green area is assigned to class 2. The MDM classifier is arguably simple, as it only cares about the distance to the class means. But since the Riemannian distance is used, it incorporates the underlying structure of the covariance matrices, which gives good performance.

When using a LogReg classifier, a new observation $C_{\text{new}} \in \mathcal{P}_n$ is first projected onto a tangent space of the manifold, then vectorized into a vector, $V_{\text{new}} \in \mathbb{R}^{n(n+1)/2}$, before the logistic regressor is applied, which is visualized in Figure 3.5. The logistic regressor returns the probability that the observed sample belongs to class 1. For the vector, V_{new} , the probability is estimated as

$$\hat{p}(V_{\text{new}}) = \frac{1}{1 + \exp(-V_{\text{new}}w - w_0)},$$

where w and w_0 are weights tuned to the training data during classifier training.

3.4 Visualization of the RPA and TSA transfer learning methods

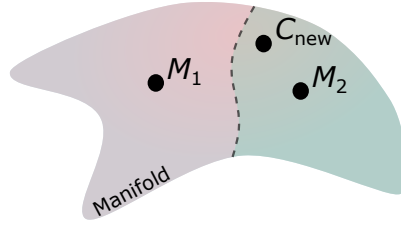


Figure 3.4 Visualization of the MDM classifier and its decision boundary on a Riemannian manifold with a new observation $C_{\text{new}} \in \mathcal{P}_n$ which is classified as class 2 in this case.

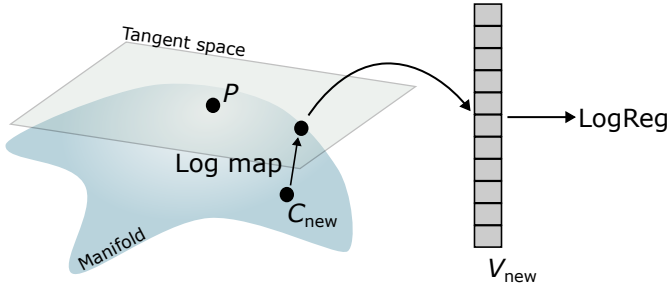


Figure 3.5 Visualization of the steps to apply a logistic regression classifier for a new observation $C_{\text{new}} \in \mathcal{P}_n$. First, projection to tangent space, then vectorization to $V_{\text{new}} \in \mathbb{R}^{n(n+1)/2}$, and finally application of the logistic regression classifier.

Since I am interested in the source data selection and transfer learning processes for BCIs, I have not put much effort into optimizing classifier choice. In particular, using the MDM classifier has helped separate the effects of transfer learning from those introduced by the classifier itself.

3.4 Visualization of the RPA and TSA transfer learning methods

Transfer learning is used to align source and target data. Recall that target data refers to data from a new subject for whom the machine learning algorithm should be trained, while source data refers to existing data from another subject or a previous session of the same subject. Here, I provide a brief introduction to the RPA and TSA transfer learning methods used in the thesis. For mathematical details on RPA and TSA, see the original papers by Rodrigues et al. (2019) and Bleuzé et al. (2022), respectively.

Both methods are domain adaptation methods that transform the data so that the source and target data align, taking the AIR SPD manifold into account. The methods consist of three main steps: recentering, scaling, and rotation. The recentering step is the same in both methods. The RPA method scales and rotates the data on the AIR SPD manifold, while the TSA method performs the scaling and rotation in the manifold's tangent space. Below is an overview of the steps for each method. In the following descriptions, the term domain refers to the source and target domains.

RPA

The transfer learning steps for the RPA method are visualized in [Figure 3.6](#). For more details on the method, see Rodrigues et al. (2019). $C_i \in \mathcal{P}_n$ is a covariance matrix in the dataset, either source or target data, depending on context.

1. **Recentering:** Recenter the data so that the identity becomes the mean of all data. This step is done on the source and target data separately. First, the mean of all data in each domain, $M \in \mathcal{P}_n$, is computed with (3.2). Then the data in each domain is recentered to the identity with

$$C_i^{\text{rec}} = M^{-1/2} C_i M^{-1/2}.$$

2. **Scaling:** The dispersion, σ , of the source and target data is calculated with (3.3). The source and target data are scaled separately to have dispersion 1 with

$$C_i^{\text{sca}} = (C_i^{\text{rec}})^{1/\sigma}.$$

3. **Rotation:** The source data are rotated to align with the target data using the congruence transformation

$$C_i^{\text{rot}} = U^\top C_i^{\text{sca}} U, \quad (3.4)$$

where $U \in \mathbb{R}^{n \times n}$ satisfies $U^\top U = I$. The rotation matrix U is found through an optimization problem that minimizes the sum of squared distances between the class means of the source data, M_k^S , and target data, M_k^T , for all K classes, as

$$U = \arg \min_{U^\top U = I} \sum_{k=1}^K \delta_R^2(U^\top M_k^S U, M_k^T). \quad (3.5)$$

The target data is not rotated.

3.4 Visualization of the RPA and TSA transfer learning methods

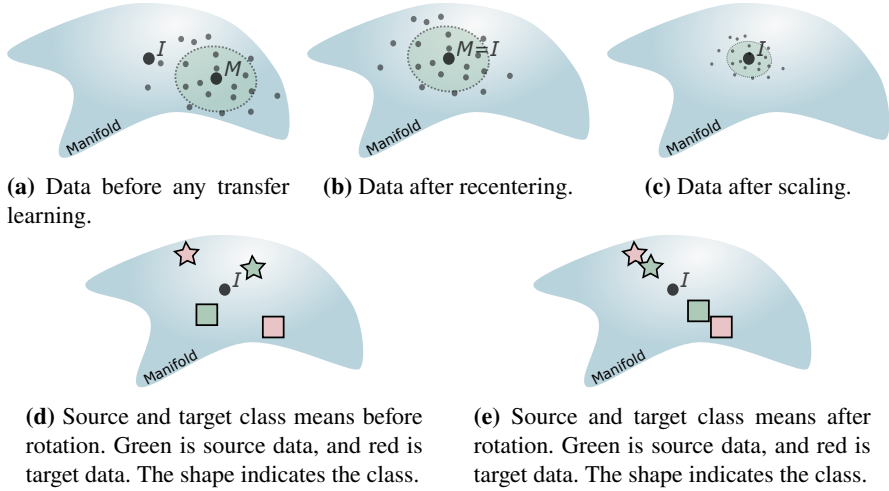


Figure 3.6 Visualization of the steps in the RPA method for some example data.

TSA

The transfer learning steps for the TSA method are visualized in Figure 3.7. For more details on the method, see Bleuzé et al. (2022). $C_i \in \mathcal{P}_n$ is a covariance matrix in the dataset, either source or target data, depending on context. $V_i \in \mathbb{R}^{n(n+1)/2}$ is the vectorized version of C_i after projection to the tangent space at the identity matrix, $T_I \mathcal{M}$.

1. **Recentering:** Recenter the data so that the identity becomes the mean of all data. This step is done on the source and target data separately. First, the mean of all data in each domain, $M \in \mathcal{P}_n$, is computed with (3.2). Then the data in each domain is recentered to the identity with

$$C_i^{\text{rec}} = M^{-1/2} C_i M^{-1/2}.$$

2. **Projection to tangent space and vectorization:** After recentering, the data C_i^{rec} is projected to the tangent space at the identity matrix, $T_I \mathcal{M}$, and vectorized to V_i^{rec} .
3. **Scaling:** The vectors V_i^{rec} form a set $\mathcal{V}$ and are scaled so the average norm is one by

$$V_i^{\text{sca}} = \frac{V_i^{\text{rec}}}{\frac{1}{|\mathcal{V}|} \sum_{V_i^{\text{rec}} \in \mathcal{V}} \|V_i^{\text{rec}}\|},$$

Chapter 3. Overview of the methods used

where $|\mathcal{V}|$ is the number of data points in the set $\mathcal{V}$. This scaling is performed separately for the source and target data.

4. **Rotation:** The target data is rotated with an orthogonal matrix $R \in O(n(n+1)/2)$ to align with the source data as

$$V_i^{\text{rot}} = R V_i^{\text{sca}}.$$

R is the orthonormal matrix that rotates T to align as well as possible with S according to the Frobenius norm,

$$\arg \min_R (\|RT - S\|_F),$$

where T and S are matrices constructed from stacked vectors T_j and S_j , representing the class means of the target and source data, respectively.

3.4 Visualization of the RPA and TSA transfer learning methods

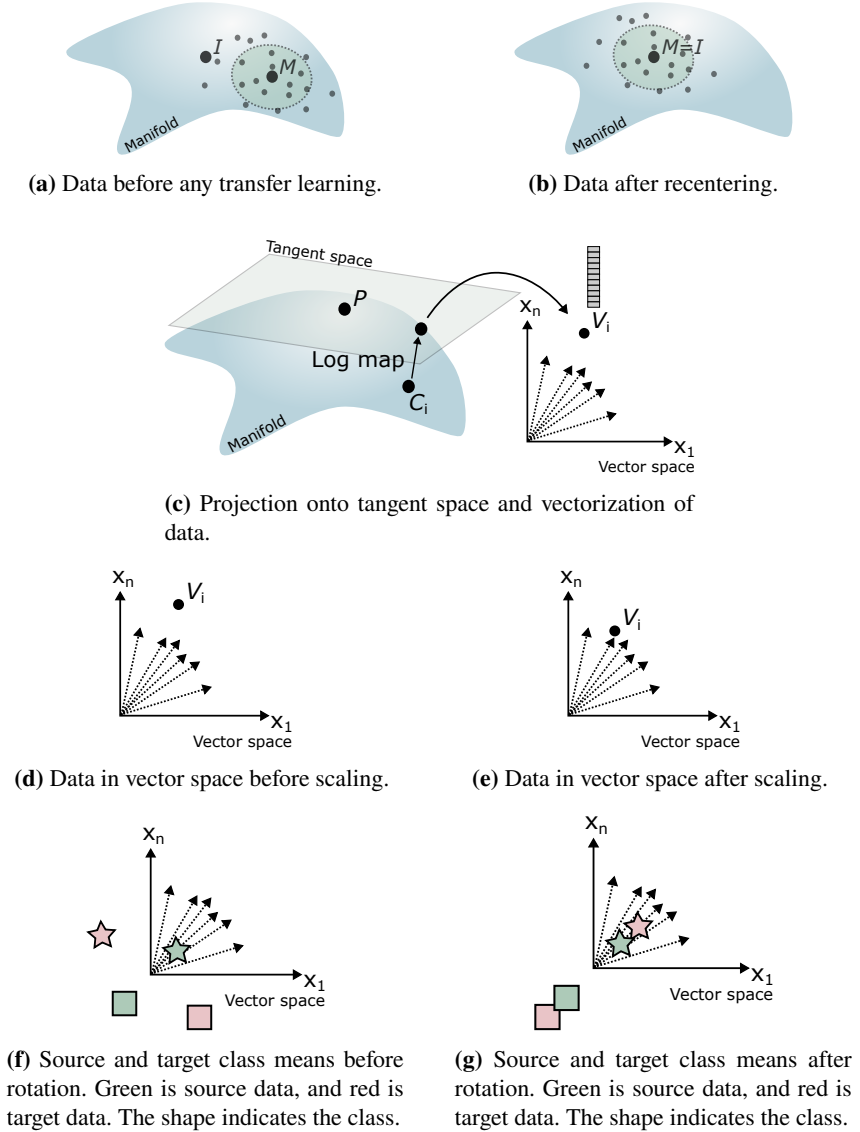


Figure 3.7 Visualization of the steps in the TSA method for some example data.

4

Contributions

The aims of the thesis are to *present methods for source data selection, provide insight and intuition into the limitations of transfer learning for BCIs, and study how to improve individualized task selection for BCIs*. [Papers I, II and V](#) are related to source data selection, [Papers II and IV](#) to understanding the limitations of transfer learning, and [Papers III and V](#) to task selection.

4.1 Included papers

The following five papers constitute the core of the thesis. They can be found at the end of the thesis.

Paper I

F. Heskebeck, C. Bergeling, and B. Bernhardsson (2026a). “Source Data Selection for Brain–Computer Interfaces Based on Simple Features”. *IEEE Access* **14**, pp. 36191–36201. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2026.3670905](#)

Summary. This paper presents the transfer performance predictor (TPP) method. The principle behind the TPP method is that it identifies source subjects whose transfer learning performance is similar to that of a new target subject and uses their optimal source data for transfer learning. The TPP method outperforms other source data selection methods when a large amount of source data is available.

Contribution. F. Heskebeck developed the TPP method, implemented it, and carried out all analyses. C. Bergeling and B. Bernhardsson supervised the work through weekly meetings and provided feedback on the direction of the research and interpretation of results. F. Heskebeck wrote the manuscript and implemented a majority

of the revisions. C. Bergeling and B. Bernhardsson read the manuscript, provided feedback, and made minor edits.

Paper II

F. Heskebeck, B. Bernhardsson, and C. Bergeling (2026c). “Rotation-based metric on the Riemannian manifold of SPD matrices with applications to source data selection for brain-computer interface transfer learning”. *Frontiers in Human Neuroscience* **20**. ISSN: 1662-5161. DOI: [10.3389/fnhum.2026.1824613](https://doi.org/10.3389/fnhum.2026.1824613)

Summary. This paper presents the pole ratio metric and a sphere-based view of rotations of SPD matrices on the Riemannian manifold of SPD matrices equipped with the AIR metric, referred to as the AIR SPD manifold. An important contribution of the paper is the insight into how the orthogonal congruence transformation limits the rotation in the RPA method. The insight follows from the fact that the ray of scaled identity matrices $\{\lambda I : \lambda > 0\}$ is invariant under orthogonal congruence transformations. The pole ratio metric describes how data on the AIR SPD manifold are positioned relative to this ray. The pole ratio is a useful metric for source data selection, as data with similar pole ratios tend to achieve higher transfer learning accuracy than data with different pole ratios.

Contribution. C. Bergeling suggested investigating the rotation step of the RPA transfer learning method to better understand rotations on the AIR SPD manifold. F. Heskebeck performed the mathematical derivations, implemented the methods and experiments, and carried out all analyses, including identifying the invariant ray and proposing the pole ratio metric. C. Bergeling and B. Bernhardsson supervised the work through weekly meetings and provided feedback on the direction of the research and interpretation of results. F. Heskebeck wrote the manuscript and implemented a majority of the revisions. C. Bergeling and B. Bernhardsson read the manuscript, provided feedback, and made minor edits.

Paper III

F. Heskebeck, B. Bernhardsson, and C. Bergeling (2026b). “Class mean distance for early classification accuracy estimation”. Manuscript presented at NeuroAdaptive conference 2026 (NAT’26). URL: <https://neuroadaptive.org/conference/2026/> (visited on 2026-03-05)

Summary. This paper introduces the class-mean distance as a metric for predicting the final classification accuracy in MI BCIs. Two methods are presented: the direct prediction band and the mean-difference approach. The methods can be used early

Chapter 4. Contributions

in the calibration phase of a BCI to predict whether the selected MI classes will yield good performance. If not, changing the MI classes might improve performance.

Contribution. F. Heskebeck carried out all analyses. F. Heskebeck and C. Bergeling translated the results into methods for predicting final classification accuracy. C. Bergeling and B. Bernhardsson supervised the work through weekly meetings and provided feedback on the direction of the research and interpretation of results. F. Heskebeck wrote the manuscript and implemented the revisions. C. Bergeling and B. Bernhardsson read the manuscript and provided feedback.

Paper IV

M. Gemborn Nilsson and F. Heskebeck (2026). “Comparison and extension of the Riemannian Procrustes analysis and Tangent space alignment transfer learning methods”. Unpublished manuscript. Under review.

Summary. This paper extends the RPA and TSA transfer learning methods, developed for real-valued SPD matrices, to complex-valued Hermitian positive definite (HPD) matrices. A comparison of the RPA and TSA methods, both mathematically and empirically across several BCI datasets, is also performed. The comparison finds that, despite their similarities, the TSA method is more flexible in its rotation, which generally results in higher transfer learning performance than the RPA method when a lot of target domain knowledge is available. In a more calibration-like setting with limited target-domain knowledge, simpler transfer learning methods outperform both RPA and TSA when source data selection is successful, highlighting the importance of effective source data selection methods.

Contribution. The two authors share first authorship of this paper. The idea of the paper was developed by both authors. Simulations, visualizations, and experiments were designed and conducted by both authors. Mathematical derivations were mainly conducted by M. Gemborn Nilsson, while statistical analysis was mainly conducted by F. Heskebeck. The original draft was written by both authors, who also contributed to the paper’s review and editing.

Paper V

F. Heskebeck, C. Bergeling, and B. Bernhardsson (2022). “Multi-Armed Bandits in Brain-Computer Interfaces”. *Frontiers in Human Neuroscience* **16**. ISSN: 1662-5161. DOI: [10.3389/fnhum.2022.931085](https://doi.org/10.3389/fnhum.2022.931085)

Summary. This paper is a mini-review that presents the theory of multi-armed bandits (MABs) and its applications to BCIs. Relevant applications in the context of this thesis include source data and task selection for MI-based BCIs.

Contribution. The conceptualization of the paper was done collectively by all authors. F. Heskebeck conducted the literature review and wrote the manuscript, with input and revisions from all co-authors.

4.2 Additional papers

The following papers, to which F. Heskebeck contributed as an author or co-author, are not included in the thesis for various reasons.

BCI-related papers

The following papers are related to BCIs but out of scope for this thesis.

- F. Heskebeck and P. Tufvesson (2024). “Automatic Control of a Wheelchair Using a Brain Computer Interface and Real-Time Decision-Making”. In: *2024 European Control Conference (ECC)*, pp. 3892–3897. doi: [10.23919/ECC64448.2024.10591133](https://doi.org/10.23919/ECC64448.2024.10591133)
- P. Tufvesson and F. Heskebeck (2024a). “Automatic control of reactive brain computer interfaces”. *IFAC Journal of Systems and Control* **27**, p. 100251. ISSN: 2468-6018. DOI: [10.1016/j.ifacsc.2024.100251](https://doi.org/10.1016/j.ifacsc.2024.100251)
- J. Wilroth, B. Bernhardsson, F. Heskebeck, M. A. Skoglund, C. Bergeling, and E. Alickovic (2023). “Improving EEG-based decoding of the locus of auditory attention through domain adaptation”. *J. Neural Eng.* **20**:6, p. 066022. ISSN: 1741-2552. doi: [10.1088/1741-2552/ad0e7b](https://doi.org/10.1088/1741-2552/ad0e7b)
- M. Gemborn Nilsson, P. Tufvesson, F. Heskebeck, and M. Johansson (2023). “An open-source human-in-the-loop BCI research framework: method and design”. *Frontiers in Human Neuroscience* **17**. ISSN: 1662-5161. doi: [10.3389/fnhum.2023.1129362](https://doi.org/10.3389/fnhum.2023.1129362)
- F. Heskebeck and C. Bergeling (2021). “An Adaptive Approach for Task-Driven BCI Calibration”. In: *BCI Meeting 2021*

Pedagogical papers

The following papers are related to pedagogy with a focus on teaching automatic control, and thus out of scope for the thesis.

- F. Heskebeck, P. Tufvesson, and A. Hällström (2025). “Student usage of Lund University Pole-Zero Explorer interactive tool in automatic control teaching”. *European Journal of Control*, p. 101313. ISSN: 0947-3580. DOI: [10.1016/j.ejcon.2025.101313](https://doi.org/10.1016/j.ejcon.2025.101313)
- F. Heskebeck (2024). “Discussion Seminars – a Complement to Online Teaching”. *IFAC-PapersOnLine*. 4th IFAC Workshop on Internet Based Control Education - IBCE 2024 **58**:26, pp. 70–75. ISSN: 2405-8963. DOI: [10.1016/j.ifacol.2024.10.273](https://doi.org/10.1016/j.ifacol.2024.10.273)
- P. Tufvesson and F. Heskebeck (2024b). “LU-PZE: Lund University Pole-Zero Explorer”. *IFAC-PapersOnLine*. 4th IFAC Workshop on Internet Based Control Education - IBCE 2024 **58**:26, pp. 7–12. ISSN: 2405-8963. DOI: [10.1016/j.ifacol.2024.10.262](https://doi.org/10.1016/j.ifacol.2024.10.262)
- F. Heskebeck (2023). “Discussion Seminars – an Award-winning Pedagogical Method”. In: *Pedagogiska Inspirationskonferensen-Genombrottet*. Vol. 12. Lund
- M. Bauer and F. Heskebeck (2022). “A Workplace Equality Workshop for the Control Engineering Classroom”. *IFAC-PapersOnLine*. 13th IFAC Symposium on Advances in Control Education ACE 2022 **55**:17, pp. 97–102. ISSN: 2405-8963. DOI: [10.1016/j.ifacol.2022.09.231](https://doi.org/10.1016/j.ifacol.2022.09.231)

5

Discussion of key findings

At this point in the thesis, you are familiar with the BCI challenges and the methods I have used to address them. From the previous chapter, you are also somewhat familiar with my research. If you need more details, you can find the papers at the end of the thesis. In this chapter, I will discuss my research, explore connections between the papers, and highlight interesting aspects of the results. The purpose is to examine each paper's key contributions and place them in a broader perspective. First, I discuss the source data selection problem, followed by insights into the limitations of transfer learning. Next, the task selection problem is addressed, and finally, the chapter briefly discusses cloud computing for BCIs.

5.1 Selecting source data is not a trivial task

Despite the availability of good transfer learning methods, some source subjects yield better transfer learning performance for a target subject than others. This is particularly true when little target training data is available, such as in a calibration setting, as seen in [Paper IV \[Gemborn Nilsson and Heskebeck, 2026\]](#). Thus, selecting good source data for the target subject is important for effective transfer learning. Effective transfer learning is important for BCI performance, which ultimately affects the user experience and which BCI applications will be available. In this section, I discuss the TPP method presented in [Paper I \[Heskebeck et al., 2026a\]](#), the pole ratio used for source data selection presented in [Paper II \[Heskebeck et al., 2026c\]](#), and MABs used for source data selection summarized in [Paper V \[Heskebeck et al., 2022\]](#).

The requirement for successful transfer learning is that the source and target data can be aligned with the transfer learning method in a way that yields good transfer learning performance. That is, the transfer learning method can align the data so that the subsequent classifier yields good classification accuracy. For that, the data must

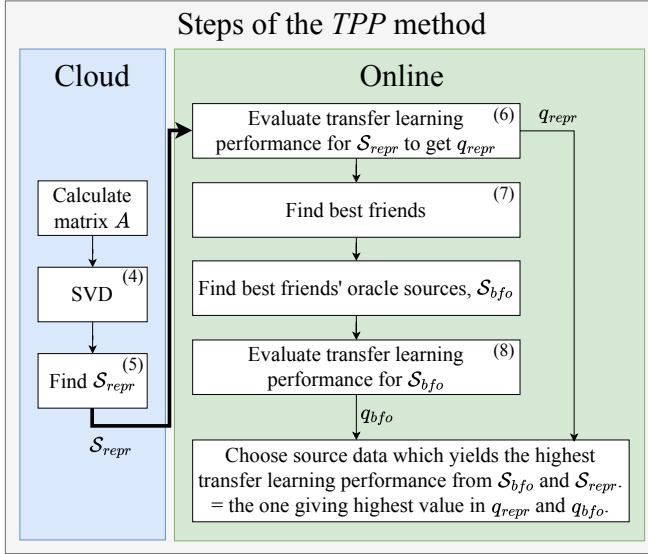


Figure 5.1 Flowchart for the TPP method. The feature matrix $\mathbf{A}$, which contains the transfer learning performance for all combinations of source subjects and source subjects acting as target subjects, is used to identify source subjects similar to the new target subject. The optimal source data for these similar source subjects is evaluated for source data selection. [Figure from Heskebeck et al. (2026a), where more details can be found.]

be similar in the aspects that the transfer learning method and the classifier cannot adjust for. In my work, I have considered both similar transfer learning performance and similar data distributions as metrics for similarity between source and target data. Which approach is best is up for debate. One might assume that data-level similarities are more informative. But which features of the source and target data are relevant for determining similarity depends on the transfer learning method used, since different methods can compensate for different types of differences, as illustrated by the pole ratio for the RPA and TSA methods. Examining similarity in transfer learning performance, as in the TPP method, may be an approach to create a more robust source data selection method.

The TPP method identifies subjects with similar transfer learning performance

The TPP method is a clever method for source data selection that, at first glance, can seem somewhat complicated. The method's steps are visualized in Figure 5.1. The underlying assumption for the TPP method is that a target subject, called target A,

has a similar transfer learning performance on source data to that of another target subject, called target B. This means that if we know the optimal source subject for target B, then that source subject will most likely perform well for target A as well. The TPP method identifies similar target subjects by comparing their transfer learning performance on a subset of source subjects. The optimal source data for these similar target subjects is evaluated for source data selection.

A strength of the TPP method is that it is not bound to any specific transfer learning method and that it can handle a large amount of available source data. The only requirement is that the transfer learning performance (for a specific choice of transfer learning method and classifier) for all combinations of source subjects and source subjects acting as target subjects can be calculated. These transfer learning performance results are collected in a feature matrix denoted $\mathbf{A}$ in the paper. If the transfer learning method or classifier is changed, the matrix $\mathbf{A}$ must be recomputed. The main drawback of the TPP method is that computing the matrix $\mathbf{A}$ is computationally expensive, especially for large datasets. Here, the method can benefit from centralized cloud computing, which can offload these computations.

The core idea of the TPP method is that some target subjects have similar transfer learning performance on source data. As hinted in [Section 2.3](#), there could be clusters of target and source subjects that all have similar transfer learning performance. Thus, for a new target subject, rather than finding a similar target subject, one might instead identify a cluster of similar target subjects in order to select a cluster of source subjects. The approach used in the TPP method for identifying representative source subjects could be used to identify these clusters of subjects.

The pole ratio method leverages knowledge of how data is rotated on the manifold

The pole ratio describes how data is positioned on what I call a geodesic sphere on the AIR SPD manifold, similar to how latitude describes points on Earth. The pole ratio is preserved during the rotational step of the RPA transfer learning method and provides information about the similarity between the source and target data. Selecting a source subject with a pole ratio close to that of the target subject is beneficial when using the RPA transfer learning method.

A strength of the pole ratio approach for source data selection is that it, in effect, compares the similarity of the EEG data between source and target subjects. The main drawback of the method is that it is only relevant for transfer learning methods that preserve the pole ratio. For example, the RPA method preserves the pole ratio while the TSA method does not. When discussing domain adaptation methods, one should bear in mind that the classifier applied after transfer learning may be able to handle some of the remaining differences between the source and target data. I have

deliberately used the MDM classifier to facilitate evaluation of how well the source and target data align after transfer learning. Still, transfer learning performance is measured by the classification accuracy after transfer learning and is therefore inherently coupled to the combination of transfer learning method and classifier.

In a source data selection setting with two classes, the pole ratio for one class is sufficient to identify which source and target subjects are similar since the two class means are placed on opposite sides of the geodesic sphere. If more than two classes are used, the pole ratio for each class as well as the relative position between the classes on the geodesic sphere would likely need to be used for source data selection.

Multi-armed bandits gave inspiration for the source data selection research

Using MABs for source data selection is a research path I explored early in my PhD studies, but have not studied further. Still, they deserve a mention in this thesis, as this early research opened my eyes to the problem of selecting source data for BCIs. MABs provide a structured way to balance exploration of new options against exploitation of known successes. For example, should you visit a restaurant you know is good or try a new one that might be better? In a BCI source data selection setting, the restaurants correspond to the source subjects. To make informed decisions when selecting source data with the MAB strategy, metadata about the source and target data could be used, such as age, gender, and other physiological parameters.

A strength of the MAB approach is that it is not bound to any specific transfer learning method. A drawback is that many trials and errors are needed to identify good source data. One can compare MABs for source data selection to the use of MABs for selecting advertisements on websites. When a website user opens a webpage, the MAB strategy is used to select which advertisement to show. Based on the outcome of the selected advertisement, the next time a similar user opens the webpage, the same or a different advertisement is shown. The consequence of showing a poorly suited advertisement to a website user is small. However, for a BCI user, poorly selected source data can reduce BCI performance, which, depending on the application, could have a significant impact on the user's life. So the cost of the trial-and-error required to fine-tune the MAB strategy might be too high for source data selection, whereas it is worthwhile in a website advertisement setting.

Comparing and combining methods for source data selection

As always, one method is not the solution to all problems. The TPP and MAB methods are similar in that they are not bound to any specific transfer learning method, so they can be used in the same way even if the transfer learning method changes. This is a strength of these methods. The TPP and MAB methods rely on

5.2 *Understanding transfer learning limitations is a first step toward further development*

transfer learning performance, whereas the pole ratio metric uses the EEG data more directly. This is a strong argument for the pole ratio method, as using metrics based on EEG data, or derived representations such as covariance matrices, is of great interest for the future development of source data selection methods. However, as discussed above, the pole ratio method is inherently relevant only to transfer learning methods where the pole ratio contains important information about the transfer learning. Which source data selection method to use depends on the chosen transfer learning method and the requirements for BCI performance.

The pole ratio and TPP methods can potentially be combined to enhance performance. The pole ratio could serve as a first selection of similar subjects, and then the TPP method could identify the optimal source data from these. The pole ratio could also serve as metadata for the MAB strategy, potentially enhancing the selection process. Whether combining these methods improves performance remains to be determined in future research.

Central to source data selection is the similarity between source and target subjects. How similarities in the data and transfer learning performance relate to underlying neurological similarities between subjects is an intriguing question. Is the brain's shape important for identifying similar subjects, since it affects how EEG measures brain activity? Is the way a person thinks important for the similarity of the data? For example, in the MI setting, is how the subjects imagine their hand movements important for the similarity of the EEG data they generate? Further research is needed to address these questions, but their answers may provide valuable insights into the source data selection problem.

When discussing source data selection, it is easy to focus on the task of finding the optimal source data, where optimal source data refers to the source data that yields the highest transfer learning performance for a target subject. While this is a good aim, finding one of the best-performing sources rather than the optimal one might be enough to achieve good BCI performance. The more source data to choose from, the harder it is to find the best-performing data, simply because of the sheer volume of it. Then, effective methods for selecting source data are even more important.

5.2 Understanding transfer learning limitations is a first step toward further development

To create informed source data selection methods, a useful approach is to understand what happens during the transfer learning. In addition, understanding the strengths and weaknesses of current transfer learning methods for BCIs allows for further development of these methods. If transfer learning performance can be improved

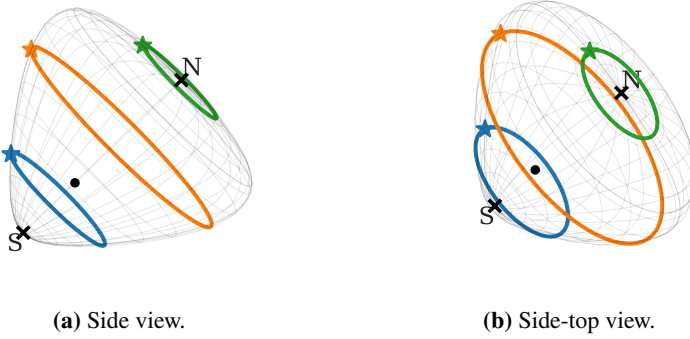


Figure 5.2 The geodesic sphere centered at I for 2×2 matrices with $r = 1$ visualized in $\mathbb{R}^3$. Three diagonal matrices, marked with a star, and their orthogonal congruence rotations are highlighted in blue, orange, and green, respectively. The identity matrix is marked with a black dot, and the so-called north (N) and south (S) poles are marked with crosses. [Figure is adapted from Gemborn Nilsson and Heskebeck (2026)]

when there is little target training data, it would mean reduced calibration time and improved user experience. My research in this field contributes to the understanding of the RPA [Rodrigues et al., 2019] and TSA [Bleuzé et al., 2022] methods, which are transfer learning methods based on Riemannian geometry and are summarized in Section 3.4. My main contributions are the sphere-based view of RPA rotations along with the pole ratio presented in Paper II [Heskebeck et al., 2026c], and the comparison of the RPA and TSA methods in Paper IV [Gemborn Nilsson and Heskebeck, 2026].

Central to the understanding of the transfer learning limitations is what I call the geodesic sphere. The geodesic sphere consists of all points at Riemannian distance r from the identity matrix, $\mathcal{S}(r) := \{C \in \mathcal{P}_n \mid \delta_R(I, C) = r\}$. The geodesic sphere for 2×2 SPD matrices can be visualized in $\mathbb{R}^3$ as in Figure 5.2, but this is not possible for matrices of larger size. However, the definition of the geodesic sphere is the same regardless of size: all points at a given Riemannian distance from a central point. The so-called north and south poles, denoted $e^\gamma I$ and $e^{-\gamma} I$ respectively, are invariant under orthogonal congruence transformations (3.4) and serve as fixed points on the geodesic sphere. The pole ratio, $\rho(C)$, for a matrix $C \in \mathcal{S}(r)$, describes the position of C relative to these fixed points, similar to latitude on Earth, and is calculated using the Riemannian distance (3.1) as

$$\rho(C) = \frac{\delta_R(C, e^\gamma I)}{\delta_R(C, e^\gamma I) + \delta_R(C, e^{-\gamma} I)}. \quad (5.1)$$

In the visualized 2×2 case, all points on a colored ring share the same pole ratio.

The importance of the geodesic sphere and the pole ratio is that they give intuition and understanding for rotations on the AIR SPD manifold. The geodesic sphere is central for the comparison of the RPA and TSA methods, which will be discussed next. The pole ratio is thoroughly reviewed in the subsequent section.

Rotation on the manifold is more limited than rotation in the tangent space

The RPA method rotates points on the AIR SPD manifold around the geodesic sphere while keeping the pole ratio fixed, as illustrated in [Figure 5.2](#) with the colored rings. If there are two classes in the data, the class means will be on opposite sides of the geodesic sphere. The same orthogonal congruence rotation (3.4) is applied to all source data, meaning that the class means will stay on opposite sides of the geodesic sphere after rotation. Exemplified by the 2×2 case in the figure, each class mean would rotate around its corresponding colored ring while staying on opposite sides of the geodesic sphere. The TSA method, on the other hand, can move points arbitrarily on the geodesic sphere. So, in some sense, the TSA method is freer in the rotation than the RPA method. Still, all source data are rotated by the same transformation, so the two class means are rotated together and will therefore remain on more or less opposite sides of each other during rotation. A main difference between the RPA and TSA methods is thus that the RPA method keeps the pole ratio for the class means whereas the TSA method can change it.

A possible interpretation of the orthogonal congruence transformation used in the RPA method is as a linear mixing of the EEG channels. As discussed in [Section 3.2](#), if the same brain activity is measured with, for example, slightly different placements of the EEG cap, these measurements can be translated through a linear mixing. So when the RPA method rotates the data, it can be interpreted as aligning the source data's EEG channel representation with that of the target data. The TSA method, on the other hand, performs the rotation in the tangent space of the AIR SPD manifold, and the corresponding transformation on the AIR SPD manifold need not be an orthogonal congruence transformation. Thus, the TSA method does not have such a straightforward interpretation for what is happening with the EEG data as the RPA method has.

For both the RPA and TSA methods, there is no requirement that the underlying brain activity is the same for the alignment to be successful. It could be that the source subject performs a right- versus left-hand motor movement task while the target subject is thinking about cats versus dogs. Particularly for the TSA method, the alignment would still work regardless of the underlying mental tasks. What is mathematically doable in terms of alignment and how this can be motivated from a neurological perspective is open for discussion. Is it important that the source and

target subjects have performed the same mental task for the transfer learning to be meaningful? Being able to do transfer learning regardless of mental tasks could be beneficial for BCI calibration.

A very interesting observation from comparing the RPA and TSA methods is that when there is a lot of target domain knowledge, the TSA method outperforms the RPA method. But when the target domain knowledge is limited, as in a calibration setting, the RPA method (and even simpler methods, such as only performing recentering) is superior to the TSA method if effective source data selection can be performed. This is likely because the TSA and RPA methods are limited by the accuracy of the estimated class means. If the class means are poor estimates of the data, the TSA may distort the data while the RPA method is forced to keep some underlying structure of the data, which could be a strength when the target domain knowledge is limited. The observation that successful transfer learning in cases with limited target domain knowledge depends on effective source data selection is a strong argument for the importance of source data selection.

A deeper dive into the pole ratio – what it does and does not say about the similarity of subject data

As a reminder, the pole ratio describes a point's relative position on the geodesic sphere between the north and south poles. The north and south poles are invariant to orthogonal congruence transformations (3.4) and thus serve as fixed points on the geodesic sphere. Congruence transformations with orthogonal matrices are of interest because they are used in the RPA transfer learning method. In the remainder of this subsection, I will show that two matrices that are orthogonally congruent have the same pole ratio, but any two matrices with the same pole ratio are not necessarily orthogonally congruent. That is, for SPD matrices $C_1, C_2 \in \mathcal{P}_n$, with $\rho(C_i)$ denoting the pole ratio of C_i ,

$$\exists Q \in O(n) : C_2 = Q^\top C_1 Q \implies \rho(C_1) = \rho(C_2), \quad (5.2)$$

whereas

$$\rho(C_1) = \rho(C_2) \not\implies \exists Q \in O(n) : C_2 = Q^\top C_1 Q. \quad (5.3)$$

First, let the eigenvalues of an $n \times n$ SPD matrix $C \in \mathcal{P}_n$ be $\mu_i = e^{\lambda_i}$ where $i = 1, \dots, n$ (this is valid since all eigenvalues of SPD matrices are positive). Secondly, let $e^\gamma I$ be the north pole and $e^{-\gamma} I$ the south pole on the geodesic sphere with radius $r > 0$ where $\gamma = \frac{r}{\sqrt{n}}$ as described in [Heskebeck et al., 2026c].

To establish (5.2), consider an SPD matrix C . It can be diagonalized with some orthogonal matrix Q such that $C = Q\Lambda Q^\top$ where the diagonal matrix Λ has the eigenvalues e^{λ_i} on the diagonal. Equivalently, $\Lambda = Q^\top C Q$, which is an orthogonal congruence transformation. This shows that orthogonal diagonalization can be

5.2 Understanding transfer learning limitations is a first step toward further development

viewed as a congruence transformation. Thus, two SPD matrices are orthogonally congruent if and only if they have the same eigenvalues, including multiplicities. It follows that two orthogonally congruent matrices have the same distance to the identity matrix since the Riemannian distance (3.1) gives

$$r = \delta_R(I, C) = \sqrt{\sum_{i=1}^n \lambda_i^2}.$$

Consequently, they also have the same value of $\gamma = r/\sqrt{n}$. The pole ratio, given in (5.1), can then be written as

$$\rho(C) = \frac{\sqrt{\sum_{i=1}^n \log^2(e^{\lambda_i} e^{-\gamma})}}{\sqrt{\sum_{i=1}^n \log^2(e^{\lambda_i} e^{-\gamma})} + \sqrt{\sum_{i=1}^n \log^2(e^{\lambda_i} e^{\gamma})}}.$$

For the same set of eigenvalues e^{λ_i} , this expression has the same value. Thus, the pole ratio is preserved under orthogonal congruence transformations.

To establish (5.3), consider all matrices C at a Riemannian distance r from the identity matrix I with a given pole ratio ρ . The squared Riemannian distance from C to the identity matrix is

$$r^2 = \delta_R^2(C, I) = \sum_{i=1}^n \log^2(\mu_i) = \sum_{i=1}^n \lambda_i^2. \quad (5.4)$$

The squared Riemannian distance from C to the north pole is

$$\delta_R^2(C, e^\gamma I) = \sum_{i=1}^n \log^2(e^{\lambda_i} e^{-\gamma}) = \sum_{i=1}^n (\lambda_i - \gamma)^2 = 2r^2 - 2\gamma \sum_{i=1}^n \lambda_i.$$

Using $S = \sum_{i=1}^n \lambda_i$, the pole ratio is

$$\rho = \frac{\sqrt{2r^2 - 2\gamma S}}{\sqrt{2r^2 - 2\gamma S} + \sqrt{2r^2 + 2\gamma S}}.$$

Consequently, for fixed r , n , and ρ , solving for S gives

$$S = \sum_{i=1}^n \lambda_i = r\sqrt{n} \frac{1 - 2\rho}{2\rho^2 - 2\rho + 1}. \quad (5.5)$$

The first requirement, (5.4), corresponds to a sphere in the space of log-eigenvalues, while the second requirement, (5.5), corresponds to a hyperplane. Thus, all log-eigenvalue vectors $\lambda = (\lambda_1, \dots, \lambda_n)$ that satisfy these requirements are found at the

intersection of this sphere and hyperplane. For 2×2 matrices, there are two unknown log-eigenvalues and two equations, giving a determined system. The two solutions differ only by a permutation of the eigenvalues and therefore correspond to the same eigenvalue multiset. However, for matrices of larger size, the system is underdetermined, and the solution generally contains infinitely many distinct eigenvalue sets. Consequently, two matrices with the same pole ratio do not necessarily have the same eigenvalues and are therefore not necessarily orthogonally congruent.

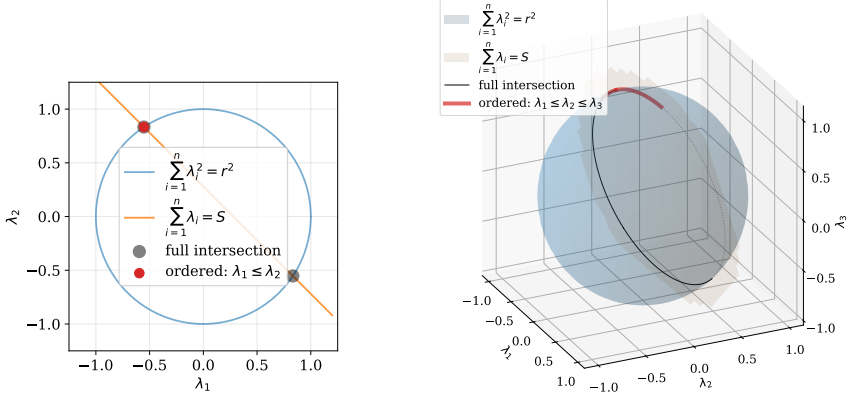
For 2×2 and 3×3 matrices, these requirements are visualized in Figure 5.3, for $r = 1$ and $\rho = 0.45$. The intersections are highlighted in Figures 5.3a and 5.3b, with the corresponding ordered eigenvalues shown in Figures 5.3c and 5.3d. As discussed above, in the 2×2 case, the intersection corresponds to a single ordered eigenvalue set for a fixed r and ρ . Therefore, all 2×2 SPD matrices on the same geodesic sphere and with the same pole ratio have the same eigenvalues and are orthogonally congruent. Geometrically, they can therefore be rotated into one another as illustrated by the rings in Figure 5.2.

It is worth noting that the RPA algorithm finds a rotation matrix that minimizes the sum of squared distances between the source and target data across all classes (see (3.5)). Thus, the class mean matrices in the source and target data need not be exactly related by an orthogonal congruence transformation for the rotation to be useful. If the rotation brings the source and target data sufficiently close together, this may be enough to improve transfer learning performance. However, the discussion above on orthogonal congruence remains relevant, as it provides insight into the limitations of rotations on the AIR SPD manifold.

5.3 Individualized task selection can reduce calibration time in MI-based BCIs

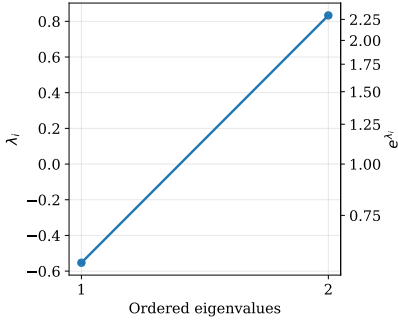
Individualized task selection is particularly important for MI-based BCIs, where the optimal MI classes differ across subjects. For example, for some subjects, the right- versus left-hand classes might be hard to distinguish, in which case the hands versus feet classes might be more distinguishable. The naive approach is to collect training data from all classes during calibration and then select the classes that yield the best performance. If classes that perform well can be identified during calibration, only data from these classes need to be collected, effectively reducing calibration time. This section discusses my work on task selection with the class mean distance methods presented in Paper III [Heskebeck et al., 2026b] and MABs for task selection summarized in Paper V [Heskebeck et al., 2022].

5.3 Individualized task selection can reduce calibration time in MI-based BCIs

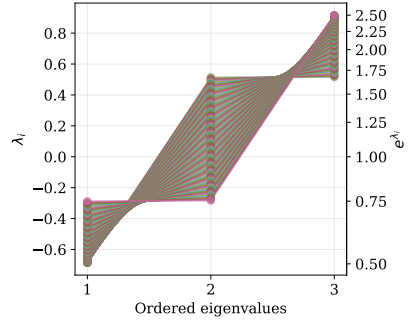


(a) Intersection of requirements for 2×2 matrices.

(b) Intersection of requirements for 3×3 matrices.



(c) Corresponding (logarithmic) eigenvalues for 2×2 matrices.



(d) Corresponding (logarithmic) eigenvalues for 3×3 matrices.

Figure 5.3 Visualization of requirements for logarithmic eigenvalues. In the 2×2 case, the intersection corresponds to a single ordered eigenvalue set for a fixed r and ρ . In the 3×3 case, there are infinitely many eigenvalue sets.

The class mean distance methods can be used for early performance estimation

The class mean distance methods are a way to estimate the classification accuracy that would be obtained if more training data were collected. This research is based on the observation of a trend between class mean distance and intra-subject performance. I present the two methods, direct prediction band and mean-difference approach, for classification accuracy estimation based on the class means, both grounded in empirical evidence. The methods cannot perfectly predict the final accuracy, but can serve as early guidance for changing classes if classification performance seems to be low.

A strength of these methods is that they appear to work across datasets, suggesting they can be used in a new BCI setup. A disadvantage is that they are based on empirical evidence and give a relatively wide range of predicted classification accuracy. The observation that a greater distance between class means implies higher intra-subject performance is unsurprising, since classification accuracy is directly related to data overlap, and greater distance typically indicates less overlap. The novel aspect of this research is how to practically use this observation in a BCI setting to reduce calibration time.

Trial and error with multi-armed bandits

The core idea of using MABs for task selection is the same as for source data selection: MABs provide a structured way to balance exploration of new options against exploitation of known successes. Here, the restaurants in the analogy presented above correspond to the MI tasks used for classification. The setup for the MAB is slightly different between the source data and task selection problems. In the source data selection problem, the approach is to select source data based on the known features of the target subject. If the selected source data give bad transfer learning performance, that source data is not selected for that type of target subject in the future. So the MAB approach is applied at a high level, collecting data across source and target subjects. In the task selection problem, on the other hand, the MAB approach is applied at an individual level for each target, selecting which class the next epoch of data should belong to in order to maximize performance. Here, there is a point in exploring the same class multiple times, even if its performance was low last time, since collecting more data might improve classification accuracy. This is like trying the same restaurant again: another dish might be better than the one you selected last time. In contrast to the source data selection problem, the MAB approach is well suited for the task selection problem.

5.4 Cloud computing creates new opportunities for BCI development

Cloud computing is mentioned in some of my papers and deserves some time in the spotlight. Simply put, cloud computing is the use of remote computers to store data, run software, or perform calculations instead of a local computer. With more mobile BCI applications running on small devices such as cellphones, cloud computing will likely become a natural component of BCIs. For example, a BCI in hearing aids will most likely rely on the cloud to run software and process data. The TPP method is an example that benefits from centralized cloud computing, as the steps in the blue area of [Figure 5.1](#) are the same regardless of the target subject and can be performed in the cloud beforehand. The steps in the green area are target subject-specific and must be performed during calibration. Splitting the TPP method into two parts reduces the computational time during calibration.

In the future, there will likely be a large amount of source data available for transfer learning. This data will probably be accessible in the cloud, and the question will be who can access it and how it may be used. Assuming the cloud can be used for heavy calculations, the source data selection method referred to as Oracle in the TPP paper will be a viable option, as it is then possible to evaluate the transfer learning performance for all available source data. However, accessing source data and performing calculations in the cloud will likely come at a cost, which is an argument for the continued need for source data selection methods.

Despite the data storage and computational offloading the cloud offers, collecting target training data remains a challenge for reducing BCI calibration times. Methods that use target training data for source data selection, transfer learning, or classifier training still need to collect it. Thus, the calibration time is not necessarily reduced by using the cloud alone.

6

Conclusions

My research is one piece of the puzzle to reach the next generation of BCIs. As hinted in the story introducing the thesis, there are many BCI challenges to overcome before BCIs can become part of our everyday lives. But with an improved understanding of source data selection, task selection, and transfer learning methods, the research community is well-positioned to develop better methods and, in turn, better BCIs. This means that BCI devices with quick startup times are one step closer to reality, improving the lives of users in medical settings and enabling more entertainment-oriented applications.

My research has focused on BCIs. But the methods can be applied to other domains as well. The TPP method is by no means bound to BCI data and is directly applicable to any field where transfer learning is used and there is a need for source data selection. A deeper understanding of the RPA and TSA methods is relevant to any field in which data are represented as SPD matrices and transfer learning on the AIR SPD manifold is applicable.

6.1 Research aims revisited

The three research aims: to *present methods for source data selection*, to *provide insight and intuition into the limitations of transfer learning for BCIs*, and to *study how to improve individualized task selection for BCIs* are illustrated in [Figure 6.1](#). My work has covered them all in different aspects, as shown in the figure. The research outcomes I am most proud of are the TPP method for source data selection, presented in [Paper I](#), and the understanding of transfer learning with the RPA and TSA methods, presented in [Papers II](#) and [IV](#). Having effective methods for source data selection is important for BCI performance, and understanding the limitations of transfer learning methods is important when developing new ones.

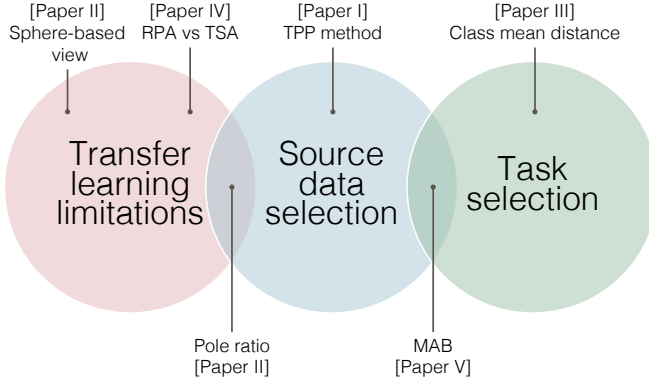


Figure 6.1 Aims of the thesis with my contributions highlighted.

The first aim, to *present methods for source data selection*, is met by the TPP and pole ratio methods, presented in [Papers I](#) and [II](#), respectively. The pole ratio method is relevant to the RPA transfer learning method, whereas the TPP method is more general and particularly useful in settings with abundant source data. Selecting source data with the MAB method, as presented in [Paper V](#), is an approach I explored early in my PhD studies.

The second aim, to *provide insight and intuition into the limitations of transfer learning for BCIs*, is met through the comparison of the RPA and TSA transfer learning methods in [Paper IV](#), which builds on the sphere-based view of the RPA method presented in [Paper II](#). There is a sea of transfer learning methods to choose from; in this thesis, I limit my analysis to the RPA and TSA methods. The thesis work provides a deep understanding of both. Additionally, in the thesis, I discuss the implications of the pole ratio in greater detail, thereby further deepening the understanding of the RPA method.

The final aim, to *study how to improve individualized task selection for BCIs*, is met with the MAB studies in [Paper V](#) and the class mean distance methods presented in [Paper III](#). Again, the MAB method is a research path I explored early on in my studies. The relationship between class-mean distance and intra-subject accuracy was discovered as a result of the pole ratio research, highlighting that, as a researcher, you can never know what you will find in your research.

6.2 Future work

There are many interesting paths for future research that stem from this thesis. Here, I highlight those that I am most intrigued by and would continue on if I had more

time.

A highly relevant path for future studies is the question of what makes two subjects suitable for transfer learning. What features should be considered for source data selection? Should it be transfer learning performance as considered by the TPP method? Or data features as considered by the pole ratio? Should physiological aspects, such as the brain's shape, also be considered? Or is it something else entirely that is important?

A question that has surfaced repeatedly throughout my research is how much training data is needed to obtain a good estimate of the class means. Since both the RPA and TSA methods use the class means as features for the transfer learning, it is important that these class means represent the data well. I have not explicitly studied this, but the more data there is, the better the transfer learning performance seems to become. A metric that can be used during calibration to determine when sufficient data has been collected to accurately estimate the class means would be an effective way to reduce calibration time. Still, as seen in the results of [Paper IV](#), performance can be good with few target training samples, provided the source data selection is successful, highlighting the importance of future research on source data selection methods.

The studied source data selection and transfer learning methods do not account for the distributions of the source and target data around their respective class means. Integrating that into the methods might be an effective way to improve transfer learning performance. This research path is also related to the question of whether the source and target subjects must have performed the same mental task. I would guess that different mental tasks yield different distributions around their class means. When only considering the class means, it is reasonable to assume that the mental task might be irrelevant. But if the distribution around the class mean is also considered, it could be important that the source and target subjects performed the same or at least a similar mental task. Whether this is true or not is up for future research to figure out.

In my work, I mostly look at two-class MI task classification. How the multi-class classification differs from the two-class classification is very interesting. For example, defining the pole ratio for multi-class BCIs as a single number is not straightforward. Also, in the multi-class case, the relative positioning of the classes is likely important for classification and transfer learning, as mentioned previously. A way to assess the relative positions of the classes is likely important for source data selection in the multi-class case. Taking this one step further and drawing conclusions about the subject's cognitive processes based on the classes' relative positions would be exciting.

Finally, if you ever happen to answer any of these questions, please get in touch with me so that I can still my curiosity!

7

Final words

Doing a PhD is a long journey with many ups and downs. And for me, this journey has now come to an end. I am so thankful that I have been given this opportunity to learn and grow. It would not have been possible without the people around me: my family, supervisors, colleagues, and friends. Once again, thank you all for your endless support and for being sources of joy.

Thank you for reading my thesis!

FRIDA HESKEBECK

Bibliography

- Aristimunha, B., I. Carrara, P. Guetschel, S. Sedlar, P. Rodrigues, J. Sosulski, D. Narayanan, E. Bjareholt, Q. Barthelemy, R. T. Schirrmeister, R. Kobler, E. Kalunga, L. Darmet, C. Gregoire, A. Abdul Hussain, R. Gatti, V. Goncharenko, A. Andreev, J. Thielen, T. Moreau, Y. Roy, V. Jayaram, A. Barachant, and S. Chevallier (2025). *Mother of all BCI benchmarks*. Zenodo. doi: [10.5281/zenodo.10034223](https://doi.org/10.5281/zenodo.10034223).
- Barachant, A., S. Bonnet, M. Congedo, and C. Jutten (2012). “Multiclass Brain–Computer Interface Classification by Riemannian Geometry”. *IEEE Transactions on Biomedical Engineering* **59**:4, pp. 920–928. issn: 1558-2531. doi: [10.1109/TBME.2011.2172210](https://doi.org/10.1109/TBME.2011.2172210).
- Bauer, M. and F. Heskebeck (2022). “A Workplace Equality Workshop for the Control Engineering Classroom”. *IFAC-PapersOnLine*. 13th IFAC Symposium on Advances in Control Education ACE 2022 **55**:17, pp. 97–102. issn: 2405-8963. doi: [10.1016/j.ifacol.2022.09.231](https://doi.org/10.1016/j.ifacol.2022.09.231).
- Blankertz, B. (2018). “Gentle Introduction to Signal Processing and Classification for Single-Trial EEG Analysis”. In: Nam, C. S. et al. (Eds.). *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*. CRC Press, New York, pp. 343–370. isbn: 978-1-351-23195-4. doi: [10.1201/9781351231954](https://doi.org/10.1201/9781351231954).
- Bleuzé, A., J. Mattout, and M. Congedo (2022). “Tangent space alignment: Transfer learning for Brain–Computer Interface”. *Frontiers in Human Neuroscience* **16**. issn: 1662-5161. doi: [10.3389/fnhum.2022.1049985](https://doi.org/10.3389/fnhum.2022.1049985).
- Burwell, S., M. Sample, and E. Racine (2017). “Ethical aspects of brain computer interfaces: a scoping review”. *BMC Medical Ethics* **18**:1, p. 60. issn: 1472-6939. doi: [10.1186/s12910-017-0220-y](https://doi.org/10.1186/s12910-017-0220-y).

- Cao, Y., H. Yang, Y. Xue, F. Wang, T. Li, L. Zhao, and Y. Fu (2026). “Ethical risks and considerations of the integration of Brain-Computer Interfaces with Artificial Intelligence”. *Cognitive Neurodynamics* **20**:1, p. 46. ISSN: 1871-4099. DOI: [10.1007/s11571-025-10380-5](https://doi.org/10.1007/s11571-025-10380-5).
- Chevallier, S., I. Carrara, B. Aristimunha, P. Guetschel, S. Sedlar, B. Lopes, S. Velut, S. Khazem, and T. Moreau (2024). *The largest EEG-based BCI reproducibility study for open science: the MOABB benchmark*. DOI: [10.48550/arXiv.2404.15319](https://doi.org/10.48550/arXiv.2404.15319). arXiv: [2404.15319](https://arxiv.org/abs/2404.15319) [eess].
- Gemborn Nilsson, M. and F. Heskebeck (2026). “Comparison and extension of the Riemannian Procrustes analysis and Tangent space alignment transfer learning methods”. Unpublished manuscript. Under review.
- Gemborn Nilsson, M., P. Tufvesson, F. Heskebeck, and M. Johansson (2023). “An open-source human-in-the-loop BCI research framework: method and design”. *Frontiers in Human Neuroscience* **17**. ISSN: 1662-5161. DOI: [10.3389/fnhum.2023.1129362](https://doi.org/10.3389/fnhum.2023.1129362).
- Heskebeck, F. (2023). “Discussion Seminars – an Award-winning Pedagogical Method”. In: *Pedagogiska Inspirationskonferensen-Genombrottet*. Vol. 12. Lund.
- Heskebeck, F. (2024). “Discussion Seminars – a Complement to Online Teaching”. *IFAC-PapersOnLine*. 4th IFAC Workshop on Internet Based Control Education - IBCE 2024 **58**:26, pp. 70–75. ISSN: 2405-8963. DOI: [10.1016/j.ifacol.2024.10.273](https://doi.org/10.1016/j.ifacol.2024.10.273).
- Heskebeck, F. and C. Bergeling (2021). “An Adaptive Approach for Task-Driven BCI Calibration”. In: *BCI Meeting 2021*.
- Heskebeck, F., C. Bergeling, and B. Bernhardsson (2022). “Multi-Armed Bandits in Brain-Computer Interfaces”. *Frontiers in Human Neuroscience* **16**. ISSN: 1662-5161. DOI: [10.3389/fnhum.2022.931085](https://doi.org/10.3389/fnhum.2022.931085).
- Heskebeck, F., C. Bergeling, and B. Bernhardsson (2026a). “Source Data Selection for Brain-Computer Interfaces Based on Simple Features”. *IEEE Access* **14**, pp. 36191–36201. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2026.3670905](https://doi.org/10.1109/ACCESS.2026.3670905).
- Heskebeck, F., B. Bernhardsson, and C. Bergeling (2026b). “Class mean distance for early classification accuracy estimation”. Manuscript presented at NeuroAdaptive conference 2026 (NAT’26). URL: <https://neuroadaptive.org/conference/2026/> (visited on 2026-03-05).
- Heskebeck, F., B. Bernhardsson, and C. Bergeling (2026c). “Rotation-based metric on the Riemannian manifold of SPD matrices with applications to source data selection for brain-computer interface transfer learning”. *Frontiers in Human Neuroscience* **20**. ISSN: 1662-5161. DOI: [10.3389/fnhum.2026.1824613](https://doi.org/10.3389/fnhum.2026.1824613).

Bibliography

- Heskebeck, F. and P. Tufvesson (2024). “Automatic Control of a Wheelchair Using a Brain Computer Interface and Real-Time Decision-Making”. In: *2024 European Control Conference (ECC)*, pp. 3892–3897. DOI: [10.23919/ECC64448.2024.10591133](https://doi.org/10.23919/ECC64448.2024.10591133).
- Heskebeck, F., P. Tufvesson, and A. Hällström (2025). “Student usage of Lund University Pole-Zero Explorer interactive tool in automatic control teaching”. *European Journal of Control*, p. 101313. ISSN: 0947-3580. DOI: [10.1016/j.ejcon.2025.101313](https://doi.org/10.1016/j.ejcon.2025.101313).
- Jayaram, V. and A. Barachant (2018). “MOABB: trustworthy algorithm benchmarking for BCIs”. *Journal of Neural Engineering* **15**:6, p. 066011. ISSN: 1741-2552. DOI: [10.1088/1741-2552/aadea0](https://doi.org/10.1088/1741-2552/aadea0).
- Jayaram, V., K.-H. Fiebig, J. Peters, and M. Grosse-Wentrup (2018). “Transfer Learning for BCIs”. In: Nam, C. S. et al. (Eds.). *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*. CRC Press, New York, pp. 425–441. ISBN: 978-1-351-23195-4. DOI: [10.1201/9781351231954](https://doi.org/10.1201/9781351231954).
- Ko, W., E. Jeon, S. Jeong, J. Phy, and H.-I. Suk (2021). “A Survey on Deep Learning-Based Short/Zero-Calibration Approaches for EEG-Based Brain–Computer Interfaces”. *Front. Hum. Neurosci.* **15**. ISSN: 1662-5161. DOI: [10.3389/fnhum.2021.643386](https://doi.org/10.3389/fnhum.2021.643386).
- Lotte, F., L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger (2018). “A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update”. *Journal of Neural Engineering* **15**:3, p. 031005. ISSN: 1741-2552. DOI: [10.1088/1741-2552/aab2f2](https://doi.org/10.1088/1741-2552/aab2f2).
- Nam, C. S., I. Choi, A. Wadson, and M. Whang (2018). “Brain–Computer Interface: An Emerging Interaction Technology”. In: Nam, C. S. et al. (Eds.). *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*. CRC Press, New York, pp. 12–52. ISBN: 978-1-351-23195-4. DOI: [10.1201/9781351231954](https://doi.org/10.1201/9781351231954).
- Nicolas-Alonso, L. F. and J. Gomez-Gil (2012). “Brain Computer Interfaces, a Review”. *Sensors (Basel, Switzerland)* **12**:2, pp. 1211–1279. ISSN: 1424-8220. DOI: [10.3390/s120201211](https://doi.org/10.3390/s120201211).
- Pfurtscheller, G. and C. Neuper (2010). “Dynamics of Sensorimotor Oscillations in a Motor Task”. In: Graimann, B. et al. (Eds.). *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. The Frontiers Collection. Springer, Berlin, Heidelberg, pp. 47–64. ISBN: 978-3-642-02091-9. DOI: [10.1007/978-3-642-02091-9_3](https://doi.org/10.1007/978-3-642-02091-9_3).

- Rodrigues, P. L. C., C. Jutten, and M. Congedo (2019). “Riemannian Procrustes Analysis: Transfer Learning for Brain–Computer Interfaces”. *IEEE Transactions on Biomedical Engineering* **66**:8, pp. 2390–2401. ISSN: 1558-2531. DOI: [10.1109/TBME.2018.2889705](https://doi.org/10.1109/TBME.2018.2889705).
- Tufvesson, P. and F. Heskebeck (2024a). “Automatic control of reactive brain computer interfaces”. *IFAC Journal of Systems and Control* **27**, p. 100251. ISSN: 2468-6018. DOI: [10.1016/j.ifacsc.2024.100251](https://doi.org/10.1016/j.ifacsc.2024.100251).
- Tufvesson, P. and F. Heskebeck (2024b). “LU-PZE: Lund University Pole-Zero Explorer”. *IFAC-PapersOnLine*. 4th IFAC Workshop on Internet Based Control Education - IBCE 2024 **58**:26, pp. 7–12. ISSN: 2405-8963. DOI: [10.1016/j.ifacol.2024.10.262](https://doi.org/10.1016/j.ifacol.2024.10.262).
- Wan, Z., R. Yang, M. Huang, N. Zeng, and X. Liu (2021). “A review on transfer learning in EEG signal analysis”. *Neurocomputing* **421**, pp. 1–14. ISSN: 0925-2312. DOI: [10.1016/j.neucom.2020.09.017](https://doi.org/10.1016/j.neucom.2020.09.017).
- Wilroth, J., B. Bernhardsson, F. Heskebeck, M. A. Skoglund, C. Bergeling, and E. Alickovic (2023). “Improving EEG-based decoding of the locus of auditory attention through domain adaptation”. *J. Neural Eng.* **20**:6, p. 066022. ISSN: 1741-2552. DOI: [10.1088/1741-2552/ad0e7b](https://doi.org/10.1088/1741-2552/ad0e7b).
- Yadav, H. and S. Maini (2025). “Decoding brain signals: A comprehensive review of EEG-Based BCI paradigms, signal processing and applications”. *Computers in Biology and Medicine* **196**, p. 110937. ISSN: 0010-4825. DOI: [10.1016/j.combiomed.2025.110937](https://doi.org/10.1016/j.combiomed.2025.110937).
- Yger, F., M. Berar, and F. Lotte (2017). “Riemannian Approaches in Brain-Computer Interfaces: A Review”. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**:10, pp. 1753–1762. ISSN: 1558-0210. DOI: [10.1109/TNSRE.2016.2627016](https://doi.org/10.1109/TNSRE.2016.2627016).

