Non-invasive Neuromodulation in Motor Rehabilitation after Stroke

Rick van der Vliet

Acknowledgements

Financial support for the publication of this thesis by Daniël van der Vliet senior is gratefully acknowledged.

Colophon

ISBN	978-94-028-1993-9
Cover	Jan van Kamphout
Lay-out	Rick van der Vliet
Printing	Ipskamp printing

Non-invasive Neuromodulation in Motor Rehabilitation after Stroke

Niet-invasieve neuromodulatie bij bewegingsrevalidatie na een beroerte

Proefschrift

ter verkrijging van de graad van doctor aan de

Erasmus Universiteit Rotterdam

op gezag van de rector magnificus

Prof.dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 15 april 2020 om 13:30 uur

door

Rick van der Vliet

geboren te Schiedam, Nederland

Erasmus University Rotterdam

Promotiecommissie

Promotoren	Prof.dr. M.A. Frens
	Prof.dr. G.M. Ribbers
Overige leden	Prof.dr. D.W.J. Dippel
	Prof.dr. A.C.H. Geurts
	Prof.dr. G. Kwakkel
Copromotor	Dr. R.W. Selles

Paranimfen	Anne Lenting
	Claire Verhage

Table of contents

Chapter 1. General introduction

Stroke is a common global health-care problem[1] that is serious and disabling.[2] Currently, stroke is defined as an acute loss of neurological function caused by permanent (as opposed to a transient ischemic attack) ischemic damage from infarction or hemorrhage in the cerebrum or the spinal cord.[3] Permanency can be objectified clinically as neurological deficits outlasting 24 hours or radiologically as ischemic damage on computed tomography (CT) or magnetic resonance imaging (MRI). Because most patients with stroke survive the initial injury,[4] the largest effect on patients and families is usually through long-term impairment, limitation of activities (disability), and reduced participation (handicap).[5,6] Motor impairment after stroke, which can be regarded as a loss or limitation of function in muscle control or movement or a limitation in mobility,[7] typically affects the control of movement of the face, arm, and leg of one side of the body in about 80% of patients.[8,9] Therefore, much of the focus of stroke rehabilitation is on the recovery of movement and associated functions with high-intensity, repetitive task-specific practice.[8,9] Disappointingly though, evidence for effectiveness of therapeutic interventions aimed at motor recovery poststroke is limited.[9–11] Therefore, a better understanding of motor rehabilitation after stroke is needed.

Both restoration and compensation of motor function contribute to recovery after stroke.[7] Restoration refers to the recruitment of the same muscle groups as prestroke for a specific movement, for example by non-damaged ipsilesional premotor areas, and is measured on the ICF level of impairment.[7] The biological processes underlying restoration are most active in the first weeks poststroke, also termed the "sensitive" period,[12–15] and act as the main driver of motor recovery poststroke.[16] This specific activity is also referred to as "spontaneous biological recovery", even though studies in non-human primates[17,18] and rodents[12] do suggest that intensive motor learning with the affected limb diminishes impairment. Compensation refers to the recruitment of alternative muscle groups for a specific movement and is measured on the ICF level of activity.[7] An example could be learning how to write with the non-lesioned non-dominant hand. Compensation relies on motor learning mechanisms which are not time-sensitive, as opposed to restoration, and is therefore not restricted to a specific period poststroke.[7] Therefore, studying motor learning and spontaneous biological recovery could help develop more effective treatments for stroke recovery.

In this thesis, we aim to develop more accurate models of motor learning and spontaneous biological recovery, which both contribute to motor recovery after stroke, by applying the principles of optimal control and proportional recovery. Using these models, we aim to sensitively test the role of electrophysiology (electro-encephalography), genetics (common polymorphism in brain-derived neurotrophic factor) and neuromodulation (transcranial direct current stimulation) in healthy subjects and stroke patients.

Optimal control models of movement

Fundamental to the optimal control model of movement is the motivation to minimize motor costs and maximize rewards.[19] The optimal control framework is built on four criteria to which the brain must cater; it needs to (1) infer sensory consequences from motor commands (system identification), (2) integrate the predicted sensory consequences with the actual sensory feedback to construct a best estimate of the state of our body and world (state estimation), (3)

estimate costs and rewards of movement (cost estimation), and (4) adjust the feedback gains in order to maximize performance (optimal control). First, *system identification* means discovering the internal dynamics of the musculoskeletal system to predict the results of movement. The nervous system is believed to achieve this goal by translating an efference copy of the motor command into its sensory consequence (located in the cerebellum)[20–22] Second, *state estimation* integrates predictions from the forward model with sensory information to form a belief about the states (position, velocities, etc.) of the world and our body. In essence, the forward model forms a prior estimate of the state of the body and the world,[19,23] which is integrated with proprioceptive and visual feedback to create a posterior belief, using an optimal observer or "Kalman filter"[24] (located in the parietal cortex).[25–28] Third, to select the optimal movement out of the many possible actions, it is needed to estimate the cost function of the movement as well as the rewarding nature of the future sensory state for every time step of the movement[23] (located in the basal ganglia).[23,29,30] Finally, all these components come together in the *feedback control policy*. Here, internal state estimates are transformed into actual motor commands on the basis of feedback gains, that have been deducted from the expected rewards and costs[23] (located in the premotor cortex and primary motor cortex).[31,32] Optimal control models of movement have been successful in explaining a wide variety of movements, amongst other motor rehabilitation after stroke.

Skill acquisition (e.g mastering wheelchair skills or walking stairs with a hemiparesis) is a major part of stroke rehabilitation programs. It involves acquiring new patterns of muscle activation over an extended period ranging from days to months.[7] According to the optimal control model of movement, skill learning involves several steps relying on different areas of the brain: (1) acquiring an internal model that predicts sensory feedback for a given motor command (cerebellum), (2) combining these predictions with actual sensory information to form a belief about the states of the body (parietal cortex) and (3) setting feedback gains to optimally guide movement during execution (motor cortex).[19] Brain injury changes the relation between a motor command and sensory feedback and therefore necessitates reacquiring (1) proper internal models, which is similar to movement adaptation (cerebellum) and (2) optimal feedback gains though extensive practice (motor cortex).

However, even though optimal control models of movement have been successful in explaining motor behavior on a group level, they have hardly been used to investigate differences between individuals which result from talent, disease or neuromodulation. The difficulty lies in estimating the parameters of relatively complex models with enough certainty to discern individuals. To this end, we need more flexible and accurate statistical methods.

Proportional recovery models of stroke

The proportional recovery rule has been instrumental in modeling spontaneous upper extremity recovery by linking baseline FM-UE,[33] to the observed motor recovery (ΔFM-UE), defined as the difference between the measurements early and 3 to 6 months after stroke.[34] More specifically, the proportional recovery rule states that in 3 to 6 months (1) the majority of patients (recoverers) gain a fixed proportion, estimated between 0.55 and 0.85,[29] of their potential recovery, calculated as the difference between baseline FM-UE and the scale's maximum score of 66, while (2) the minority of patients (non-recoverers) show only very moderate improvement which cannot be linked to potential recovery.[34–36] Mechanistically, the key underlying difference

between recoverers and non-recoverers is currently understood as the intactness of the corticospinal tract early after stroke.[37–40]

However, the proportional recovery rule has been criticized for a number of reasons. Recent analyses indicated that a strong correlation between baseline FM-UE and recovery can emerge even when baseline FM-UE is completely uncorrelated to endpoint FM-UE.[41,42] Therefore, even though the proportional recovery rule is not wrong,[35] it probably overstates the predictability of endpoint FM-UE.[41,42] In addition, the proportional recovery rule does not model the time course of recovery early poststroke which means it cannot model the rate of recovery nor update predictions with repeated measurements in time. Finally, predictions of endpoint FM-UE based on the proportional recovery rule and identification of (non)-recoverers have never been cross-validated. To increase our understanding of upper extremity recovery after stroke, we therefore need a model that (1) relates the FM-UE to potential recovery as a function of time after stroke, with (2) separate sets of parameters for different subgroups, including those that show no improvement early poststroke.

Electrophysiology, genetics and neuromodulation

We investigate the electrophysiology of motor learning using electro-encephalography. Since the discovery of error-related negativity[43,44] and feedback-related negativity[45] as markers of cortical processing of binary decisions in electro-encephalography recordings over the anterior cingulate cortex,[46] understanding of these signals has advanced in at least two important ways. First, the role of error-related negativity and feedback-related negativity has been generalized to the processing of continuous error in for example visuomotor adaptation[47,48] and forcefield adaptation.[49] Second, the error-related negativity and feedback-related negativity have been identified as reflections of frontal midline theta activity (FMθ, 4-8Hz) in the frequency domain.[50] Therefore, FMθ activity might be an interesting electrophysiological marker of individual differences in motor learning ability.

As a neuromodulation tool, we choose transcranial direct current stimulation (tDCS). tDCS is a safe,[51] non-invasive technique that delivers low-intensity current to the scalp through a pair of electrodes.[52,53] Depending on the polarity of the electrodes and the spatial orientation of the underlying neurons,[54,55] tDCS was found to alter the excitability of the motor cortex, as measured with transcranial magnetic stimulation, for approximately an hour.[56–58] In addition, tDCS has been reported to improve motor skill learning in healthy subjects[59–66] and chronic stroke patients,[67,68] and upper limb rehabilitation in subacute and chronic stroke patients with moderately severe cortical damage,[69–73] presumably by releasing brain-derived neurotrophic factor[62], down-regulating GABA[74–77] and, in stroke patients, restoring the interhemispheric imbalance between the affected motor cortex and the unaffected motor cortex.[78–80]

We limit our search for genetic contributors to variations in motor skill learning to brain-derived neurotrophic growth factor (BDNF). Brain-derived neurotrophic factor (BDNF) plays a role in long-term potentiation of horizontal connections[62] and therefore motor skill learning[81,82] and is believed to be important for realizing the behavioral effects of tDCS. Activity-dependent release of BDNF has been related to motor skill learning in healthy subjects by studying the role of the common (approximately 30% of the Caucasian population[83,84]) secretion-limiting[85] BDNF Val66Met polymorphism. Agreeing with the function of BDNF in motor cortex long-term potentiation, carriers of this polymorphism were found to more slowly acquire a new

motor skill.[62,86] Therefore, since activity-dependent release of BDNF is important for motor skill learning and possibly also for translating tDCS into motor skill learning gains.

Scope of this thesis

We introduce Bayesian hierarchical modeling in Chapter 2.1 to estimate individual parameters of motor adaptation, which is the component of optimal control of movement necessary to calibrate the forward model. This statistical approach is combined with electro-encephalography in Chapter 2.2 to attribute individual differences to variations in cortical brain activity.

In chapter 3.1, we develop a longitudinal model of spontaneous recovery of motor impairment after stroke, which describes the different patterns of recovery over time using exponential functions, and identifies subgroups based on: (1) the degree of recovery as a fraction of potential recovery, (2) the rate of recovery, and (3) the initial FM-UE score. In Chapter 2.2, we compare the power to detect an intervention effect with this longitudinal mixture model of stroke to a cross-sectional, non-parametric (Mann-Whitney U) test.

In Chapter 4.1, we first study the properties of the motor map, which is the area on the skull where transcranial magnetic stimulation elicits a motor evoked potential. The motor map is an interesting measurement because the area has been found to increase in size after motor learning,[87] while the peak is known to increase following tDCS. Potentially, the motor map is therefore able to capture motor learning and the influence of neuromodulation on an electrophysiological level. Next, we study the contribution of cerebellar tDCS to cerebellar-dependent motor learning (Chapter 4.2) and cognition (Chapter 4.3) in healthy individuals. Cerebellar tDCS has been studied less than motor cortex tDCS but is from a theoretical perspective as least as interesting for rehabilitation purposes. Finally, we investigate the role of motor cortex stimulation in motor learning in the chronic phase after stroke (Chapter 4.4) and rehabilitation in the subacute phase after stroke (Chapter 4.5).

References

1. Feigin, V.L. et al. *Lancet (London, England)* 383, 245–54 (2014).
2. Hankey, G.J. *Lancet (London, England)* 389, 641–654 (2017).
3. Sacco, R.L. et al. *Stroke* 44, 2064–2089 (2013).
4. Davenport, R.J., Dennis, M.S., Wellwood, I. & Warlow, C.P. *Stroke* 27, 415–420 (1996).
5. Luengo-Fernandez, R. et al. *Stroke* 44, 2854–2861 (2013).
6. Poon, M.T.C., Fonville, A.F. & Al-Shahi Salman, R. *J. Neurol. Neurosurg. Psychiatry* 85, 660–667 (2014).
7. Krakauer, J.W. *Curr. Opin. Neurol.* 19, 84–90 (2006).
8. Langhorne, P., Coupar, F. & Pollock, A. *Lancet Neurol.* 8, 741–54 (2009).
9. Langhorne, P., Bernhardt, J. & Kwakkel, G. *Lancet* 377, 1693–1702 (2011).
10. Veerbeek, J.M. et al. *PLoS One* 9, e87987 (2014).
11. Pollock, A. et al. *Cochrane Database Syst. Rev.* 2014, CD010820 (2014).
12. Zeiler, S.R. et al. *Neurorehabil. Neural Repair* 30, 794–800 (2016).
13. Li, S. et al. *Nat Neurosci* 13, 1496–1504 (2010).
14. Biernaskie, J. & Corbett, D. *J. Neurosci.* 21, 5272–5280 (2001).
15. Biernaskie, J., Chernenko, G. & Corbett, D. *J. Neurosci.* 24, 1245–1254 (2004).
16. Kwakkel, G., Kollen, B. & Lindeman, E. *Restor. Neurol. Neurosci.* 22, 281–99 (2004).
17. Taub, E. *Exerc. Sport Sci. Rev.* 4, 335–74 (1976).
18. Nudo, R.J., Wise, B.M., SiFuentes, F. & Milliken, G.W. *Science (80-.).* 272, 1791–1794 (1996).
19. Shadmehr, R. & Krakauer, J.W. *Exp. brain Res.* 185, 359–81 (2008).
20. Nowak, D.A., Timmann, D. & Hermsdörfer, J. *Neuropsychologia* 45, 696–703 (2007).
21. Wolpert, D.M., Miall, R.C. & Kawato, M. *Trends Cogn. Sci.* 2, 338–47 (1998).
22. Brooks, J.X., Carriot, J. & Cullen, K.E. *Nat. Neurosci.* 18, 1310–7 (2015).
23. Todorov, E. & Jordan, M.I. *Nat. Neurosci.* 5, 1226–35 (2002).

24. Kalman, R.E. *J. Basic Eng.* 82, 35 (1960).
25. Grafton, S.T. et al. *Nat. Neurosci.* 2, 563–567 (1999).
26. Gréa, H. et al. *Neuropsychologia* 40, 2471–80 (2002).
27. Rushworth, M.F., Nixon, P.D. & Passingham, R.E. *Exp. brain Res.* 117, 292–310 (1997).
28. Sirigu, A. et al. *Science* 273, 1564–8 (1996).
29. Packard, M.G. & Knowlton, B.J. *Annu. Rev. Neurosci.* 25, 563–93 (2002).
30. Packard, M.G. & McGaugh, J.L. *Behav. Neurosci.* 106, 439–46 (1992).
31. Kakei, S., Hoffman, D.S. & Strick, P.L. *Nat. Neurosci.* 4, 1020–1025 (2001).
32. Sergio, L.E. & Kalaska, J.F. *J. Neurophysiol.* 89, 212–228 (2002).
33. Gladstone, D.J., Danells, C.J. & Black, S.E. *Neurorehabil. Neural Repair* 16, 232–240 (2002).
34. Krakauer, J. & Marshall, R. *Ann. Neurol.* 78, 845–847 (2015).
35. Prabhakaran, S. et al. *Neurorehabil. Neural Repair* 22, 64–71 (2008).
36. Winters, C., van Wegen, E.E.H., Daffertshofer, A. & Kwakkel, G. *Neurorehabil. Neural Repair* 29, 614–622 (2015).
37. Feng, W. et al. *Ann. Neurol.* 78, 860–70 (2015).
38. Byblow, W.D., Stinear, C.M., Barber, P.A., Petoe, M.A. & Ackerley, S.J. *Ann. Neurol.* 78, 848–859 (2015).
39. Buch, E.R. et al. *Neurology* 86, 1924–1925 (2016).
40. Marshall, R.S. et al. *Ann. Neurol.* 65, 596–602 (2009).
41. Hope, T.M.H. et al. *Brain* 306514 (2018).doi:10.1093/brain/awy302
42. Hawe, R.L., Scott, S.H. & Dukelow, S.P. *Stroke* 50, 204–211 (2019).
43. Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E. & Donchin, E. *Psychol. Sci.* 4, 385–390 (1993).
44. Falkenstein, M., Hohnsbein, J., Hoormann, J. & Blanke, L. *Psychol. brain Res.* 192–195 (1990).
45. Miltner, W.H.R., Braun, C.H. & Coles, M.G.H. *J. Cogn. Neurosci.* 9, 788–798 (1997).
46. Dehaene, S., Posner, M.I. & Tucker, D.M. *Psychol. Sci.* 5, 303–305 (1994).
47. Vocat, R., Pourtois, G. & Vuilleumier, P. *Neuropsychologia* 49, 360–367 (2011).
48. Anguera, J.A., Seidler, R.D. & Gehring, W.J. *J. Neurophysiol.* 102, 1868–1879 (2009).
49. Torrecillos, F., Alayrangues, J., Kilavik, B.E. & Malfait, N. *J. Neurosci.* 35, 12753–65 (2015).
50. Cavanagh, J.F., Zambrano-Vazquez, L. & Allen, J.J.B. *Psychophysiology* 49, 220–238 (2012).
51. Poreisz, C., Boros, K., Antal, A. & Paulus, W. *Brain Res. Bull.* 72, 208–14 (2007).
52. Nitsche, M.A. et al. *Brain Stimul.* 1, 206–23 (2008).
53. Nitsche, M.A. & Paulus, W. *Restor. Neurol. Neurosci.* 29, 463–92 (2011).
54. Rahman, A. et al. *J. Physiol.* 591, 2563–78 (2013).
55. Radman, T., Ramos, R.L., Brumberg, J.C. & Bikson, M. *Brain Stimul.* 2, 215–28, 228.e1–3 (2009).
56. Nitsche, M.A. & Paulus, W. *J Physiol* 527 Pt 3, 633–639 (2000).
57. Nitsche, M.A. & Paulus, W. *Neurology* 57, 1899–1901 (2001).
58. Wiethoff, S., Hamada, M. & Rothwell, J.C. *Brain Stimul.* 7, 468–475 (2014).
59. Waters-Metenier, S., Husain, M., Wiestler, T. & Diedrichsen, J. *J. Neurosci.* 34, 1037–50 (2014).
60. Prichard, G., Weiller, C., Fritsch, B. & Reis, J. *Brain Stimul.* 7, 532–40 (2014).
61. Reis, J. et al. *Proc Natl Acad Sci U S A* 106, 1590–1595 (2009).
62. Fritsch, B. et al. *Neuron* 66, 198–204 (2010).
63. Zimerman, M. et al. *Ann. Neurol.* 73, 10–5 (2013).
64. Vines, B.W., Cerruti, C. & Schlaug, G. *BMC Neurosci.* 9, 103 (2008).
65. Sriraman, A., Oishi, T. & Madhavan, S. *Brain Res.* 1581, 23–9 (2014).
66. Zimerman, M. et al. *Stroke* 43, 2185–2191 (2012).
67. Lefebvre, S. et al. *Front. Hum. Neurosci.* 6, 343 (2012).
68. Lefebvre, S. et al. *Brain* 138, 149–63 (2015).
69. Khedr, E.M. et al. *Neurorehabil. Neural Repair* 27, 592–601 (2013).
70. Allman, C. et al. *Sci. Transl. Med.* 8, (2016).
71. Fusco, A. et al. *Restor. Neurol. Neurosci.* 32, 301–12 (2014).
72. Lindenberg, R., Renga, V., Zhu, L.L., Nair, D. & Schlaug, G. *Neurology* 75, 2176–84 (2010).
73. Nair, D.G., Renga, V., Lindenberg, R., Zhu, L. & Schlaug, G. *Restor Neurol Neurosci* 29, 411–420 (2011).
74. Stagg, C.J. et al. *J. Neurosci.* 29, 5202–6 (2009).
75. Bachtiar, V. et al. *Curr. Biol.* 4, 1023–1027 (2015).
76. Nitsche, M.A. et al. *J Physiol* 553, 293–301 (2003).
77. Stagg, C.J., Bachtiar, V. & Johansen-Berg, H. *Curr Biol* 21, 480–484 (2011).
78. Di Pino, G. et al. *Nat. Rev. Neurol.* 10, 597–608 (2014).
79. Di Lazzaro, V. et al. *Brain Stimul.* 7, 841–848 (2014).
80. Ward, N.S. & Cohen, L.G. *Arch. Neurol.* 61, 1844–8 (2004).

81. Rioult-Pedotti, M.S., Friedman, D., Hess, G. & Donoghue, J.P. *Nat. Neurosci.* 1, 230–4 (1998).
82. Rioult-Pedotti, M.S., Friedman, D. & Donoghue, J.P. *Science (80-.).* 290, 533–536 (2000).
83. Cargill, M. et al. *Nat. Genet.* 22, 231–8 (1999).
84. Shimizu, E., Hashimoto, K. & Iyo, M. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 126B, 122–3 (2004).
85. Egan, M.F. et al. *Cell* 112, 257–269 (2003).
86. McHughen, S.A. et al. *Cereb. Cortex* 20, 1254–62 (2010).
87. Kleim, J.A. et al. *Nat. Neurosci.* 9, 735–7 (2006).

Chapter 2. Optimal control models of movement

2.1 Individual differences in motor noise and adaptation rate are optimally related

Rick van der Vliet, Maarten A. Frens, Linda de Vreede, Zeb D. Jonker, Gerard M. Ribbers, Ruud W. Selles, Jos N. van der Geest and Opher Donchin

Abstract

Individual variations in motor adaptation rate were recently shown to correlate with movement variability or "motor noise" in a forcefield adaptation task. However, this finding could not be replicated in a meta-analysis of adaptation experiments. Possibly, this inconsistency stems from noise being composed of distinct components which relate to adaptation rate in different ways. Indeed, previous modeling and electrophysiological studies have suggested that motor noise can be factored into planning noise, originating from the brain, and execution noise, stemming from the periphery. Were the motor system optimally tuned to these noise sources, planning noise would correlate positively with adaptation rate and execution noise would correlate negatively with adaptation rate, a phenomenon familiar in Kalman filters. To test this prediction, we performed a visuomotor adaptation experiment in 69 subjects. Using a novel Bayesian fitting procedure, we succeeded in applying the well-established state-space model of adaptation to individual data. We found that adaptation rate correlates positively with planning noise (β = 0.44; 95%HDI=[0.27 0.59]) and negatively with execution noise (β = -0.39; 95%HDI=[-0.50 -0.30]). In addition, the steady-state Kalman gain calculated from planning and execution noise correlated positively with adaptation rate (r = 0.54; 95%HDI = [0.38 0.66]). These results suggest that motor adaptation is tuned to approximate optimal learning, consistent with the "optimal control" framework that has been used to explain motor control. Since motor adaptation is thought to be a largely cerebellar process, the results further suggest the sensitivity of the cerebellum to both planning noise and execution noise.

Introduction

As children we all learned: some of us move with effortless grace and others are frankly clumsy. Underlying these differences are natural variations in acquiring, calibrating and executing motor skill, which have been related to genetic [1–3] and structural factors [4]. Recently, it has been suggested that differences between individuals in the rate of motor adaptation (i.e. the component of motor learning responsible for calibrating acquired motor skills to changes in the body or environment [5]), correlate with movement variability, or motor noise [6]. However, this finding was not supported by a recent meta-analysis of adaptation experiments [7]. This inconsistency may arise because motor noise has multiple components with differing relations to adaptation rate. Our study characterizes the relationship between adaptation rate and motor noise and suggests that adaptation rate varies optimally between individuals in the face of multiple sources of motor variability.

Motor noise has many physiological sources such as motor preparation noise in (pre)motor networks, motor execution noise, and afferent sensory noise [8]. Modeling [9–11] and physiological studies [12,13] have divided the multiple sources of motor noise into planning noise and execution noise (see Figure 1A). Planning noise is believed to arise from variability in the neuronal processing of sensory information, as well as computations underlying adaptation and maintenance of the states in time [10,11]. Indeed, electrophysiological studies in macaques show that activity in (pre)motor areas of the brain is correlated with behavioral movement variability [12,13]. Similar results have also been seen in humans using fMRI [14]. In contrast, execution noise apparently originates in the sensorimotor pathway. In the motor pathway, noise stems from the recruitment of motor units [15–17]. Motor noise is believed to dominate complex reaching movements with reliable visual information [17]. In addition, sensory noise stems from the physical limits of the sensory organs and has been proposed to dictate comparably simpler smooth pursuit eye movements [18,19]. Planning and execution noise might affect motor adaptation rate in different ways.

Motor adaptation has long been suspected to be sensitive to planning noise and execution noise. Models of visuomotor adaptation incorporating both planning and execution noise have been shown to provide a better account of learning than single noise models [9–11]. In addition, manipulating the sensory reliability by blurring the error feedback, effectively increasing the execution noise, can lower the adaptation rate [20–23] whereas manipulating state estimation uncertainty by temporarily withholding error feedback, effectively increasing the planning noise, can elevate the adaptation rate [23]. These studies not only suggest that adaptation rate is tuned to multiple sources of noise, but also indicate that this tuning process is optimal and can therefore be likened to a Kalman filter [24]. Possibly, differences in adaptation rate between individuals correlate with planning noise and execution noise according to the same principle, predicting faster adaptation for people with more planning noise and slower adaptation for people with more execution noise [7] (Figure 1C and Figure 1D).

To test the relation between adaptation rate and planning noise and execution noise across individuals, we performed a visuomotor adaptation experiment in 69 healthy subjects. We fitted a state-space model of trial-to-trial behavior [10,11] using Bayesian statistics to extract planning noise, execution noise and adaptation rate for each subject. We show that the adaptation rate is sensitive to both types of noise and that this sensitivity matches predictions based on Kalman filter theory.

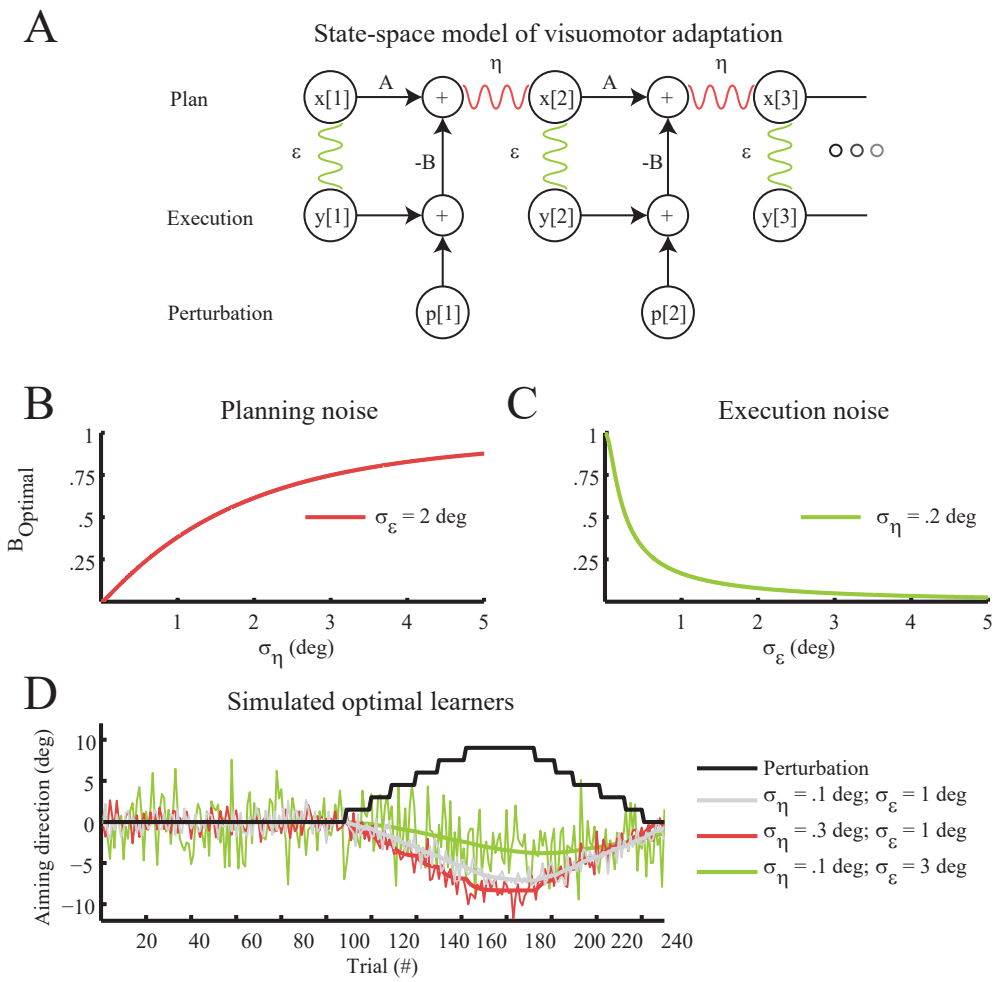


Figure 1. Planning and execution noise have opposing effects on visuomotor adaptation. A. State-space model of visuomotor adaptation. The aiming angle on trial 2 $x[2]$ is a linear combination of the aiming angle on the previous trial $x[1]$ multiplied by a retentive factor A minus the error $e[1]$ on the previous trial multiplied with adaptation rate B. In addition, the aiming angle is distorted by the random process η (planning noise). The actual movement angle $y[2]$ is the aiming angle$x[2]$ distorted by the random process ϵ (execution noise). The error $e[1]$ is the sum of the movement direction $y[1]$ and the external perturbation $p[1]$. **B.** Planning noise and optimal adaptation rate $B_{Optimal}$ (defined as the Kalman gain). The optimal adaptation rate increases with planning noise σ_η. In this figure, σ_ϵ was kept constant at 2°. **C.** Execution noise and optimal adaptation rate $B_{Optimal}$ (defined as the Kalman gain). The optimal adaptation rate decreases with execution noise σ_ϵ. In this figure, σ_η was kept constant at 0.2°. **D.** Simulated optimal learners. At trial 110, a perturbation (black line) is introduced that requires the optimal learners to adapt their movement. The gray learner has low planning noise $\sigma_\eta = 0.1°$ and execution noise $\sigma_\epsilon = 1°$. The red learner has a higher planning noise $\sigma_\eta = 0.3°$ than the gray learner $\sigma_\eta = 0.1°$. This causes the red learner to adapt faster. The green learner has a higher execution noise than the gray learner $\sigma_\epsilon = 3°$. This causes the green learner to adapt more slowly. For all learners, the thick line shows the average, thin line a single noisy realization.

Methods

Subjects

We included 69 right-handed subjects between October 2016 and December 2016, without any medical conditions that might interfere with motor performance (14 men and 55 women; age M=21 years, range 18 - 35 years; handedness score M=79; range 45 – 100). Subjects were recruited from the Erasmus MC University Medical Centre and received a small financial compensation. The study was performed in accordance with the Declaration of Helsinki and approved by the medical ethics committee of the Erasmus MC University Medical Centre.

Experimental procedure

Subjects were seated in front of a horizontal projection screen while holding a robotic handle in their dominant right hand (previously described in [25]). The projection screen displayed the location of the robotic handle ("the cursor"; yellow circle 5 mm radius), start location of the movement ("the origin", white circle 5 mm radius), and target location of the movement ("the target", white circle 5 mm radius) on a black background (see Figure 2A). Position of the origin on the screen was fixed throughout the experiment, approximately 40 cm in front of the subject at elbow height, while the target was placed 10 cm from the origin at an angle of -45°, 0° or 45°. To remove direct visual feedback of hand position, subjects wore an apron that was attached to the projection screen around their neck.

Subjects were instructed to make straight shooting movements from the origin towards the target and to decelerate only when they passed the target. A trial started with the presentation of the target and ended when the distance between the origin and cursor was at least 10 cm or when trial duration exceeded 2 seconds. At this point, movements were damped with a force cushion (damper constant 3.6 Ns/m, ramped up over 7.5 ms) and the cursor was displayed at its last position until the start of the next trial to provide position error feedback. Furthermore, timing feedback was given to keep trial duration (see definition below) in a tight range. The target dot turned blue if trial duration on a particular trial was too long (>600 ms), red if trial duration was too short (<400 ms) and remained white if trial duration was in the correct time range (400-600 ms). During presentation of position and velocity feedback, the robot pushed the handle back to the starting position. Forces were turned off when the handle was within 0.5 cm from the origin. Concurrently, the cursor was projected at the position of the handle again and subjects had to keep the cursor within 0.5 cm from the origin for 1 second to start the next trial.

The experiment included vision unperturbed, vision perturbed and no vision trials (see Figure 2B). In vision unperturbed trials, the cursor was shown at the position of the handle during the movement. The cursor was also visible in vision perturbed trials but at a predefined angle from the vector connecting the origin and the handle. In no vision trials, the cursor was turned off when movement onset was detected (see below) and was visible only at the start of the trial to help subjects keep the cursor at the origin.

The entire experiment lasted 900 trials with all three target directions (angle of -45°, 0° or 45°) occurring 300 times in random order. The three different trial types were used to build a baseline and a perturbation block (see Figure 2C). We designed the baseline block to obtain (1) reliable estimates of the noise parameters and (2) variance statistics (standard deviation and lag-1 autocorrelation of the movement angle) related to the noise parameters. Therefore, we included a large number of no vision trials (225 no vision trials) as well as vision unperturbed trials (225 vision unperturbed trials). The order of the vision unperturbed trials and no vision

trials was randomized except for trials 181-210 (no vision trials) and trials 241-270 (vision unperturbed trials). We designed the perturbation block to obtain (1) reliable estimates of the adaptation parameters and (2) variance statistics related to trial-to-trial adaptation (covariance between perturbation and movement angle). The perturbation block consisted of a large number of vision trials (400 vision trials) and a small number of no vision trials (50 no vision trials), with every block of nine trials containing one no vision trial. Every eight to twelve trials, the perturbation angle changed with an incremental 1.5° step. These steps started in the positive direction until reaching 9° and then switched sign to continue in the opposite direction until reaching -9°. This way, a perturbation signal was constructed with three "staircases" lasting 150 trials each (see Figure 2C). Design of the gradual perturbation was optimized to provide a "rich" input for system identification, without sacrificing the consistency of the signal too much as this has been shown to negatively affect the adaptation rate [26,27], and is similar to the perturbation used by Cheng and Sabes [11]. The experiment was briefly paused every 150 trials.

Data Collection

The experiment was controlled by a C++ program developed in-house. Position and velocity of the robot handle were recorded continuously at a rate of 500 Hz. Velocity data was smoothed with an exponential moving average filter (smoothing factor=0.18s). Trials were analyzed from movement start (defined as the time point when movement velocity exceeds 0.03 m/s) to movement end (defined as the time point when the distance from the origin is equal to or larger than 9.5 cm). Reaction time was defined as the time from trial start until movement start, movement duration as the time from movement start until trial end and trial duration as the time from trial start until trial end. Movement angle was calculated as the signed (+ or -) angle in degrees between the vector connecting origin and target and the vector connecting robot handle position at movement start and movement end. The clockwise direction was defined positive. Peak velocity was found by taking the maximum velocity in the trial interval. Trials with (1) a maximal displacement below 9.5 cm, (2) an absolute movement direction larger than 30° or (3) a duration longer than 1 second were removed from further analysis (2% of data).

Visuomotor adaptation model

Movement angle was modeled with the following state-space equation (see Figure 1A) [10,11]:

$$x[n+1] = Ax[n] - Be[n] + \eta \quad (1)$$

$$y[n] = x[n] + \epsilon \quad (2)$$

$$e[n] = y[n] + p[n] \quad (3)$$

$$\eta \sim N\left(0, \sigma_\eta^2\right), \epsilon \sim N(0, \sigma_\epsilon^2) \quad (4)$$

In this model, $x[n]$ is the aiming angle (the movement plan) and $y[n]$ the movement angle (the actually executed movement). Error e$[n]$ on a particular trial is the sum of $y[n]$ and the perturbation $p[n]$. The learning terms are A, which represents retention of the aiming angle over trials, and adaptation rate B, the fractional change from error $e[n]$. The movement angle is affected by planning noise process η, modeled as a zero-mean Gaussian with standard deviation σ_η, and execution noise process ϵ, modeled as a zero-mean Gaussian with standard deviation σ_ϵ.

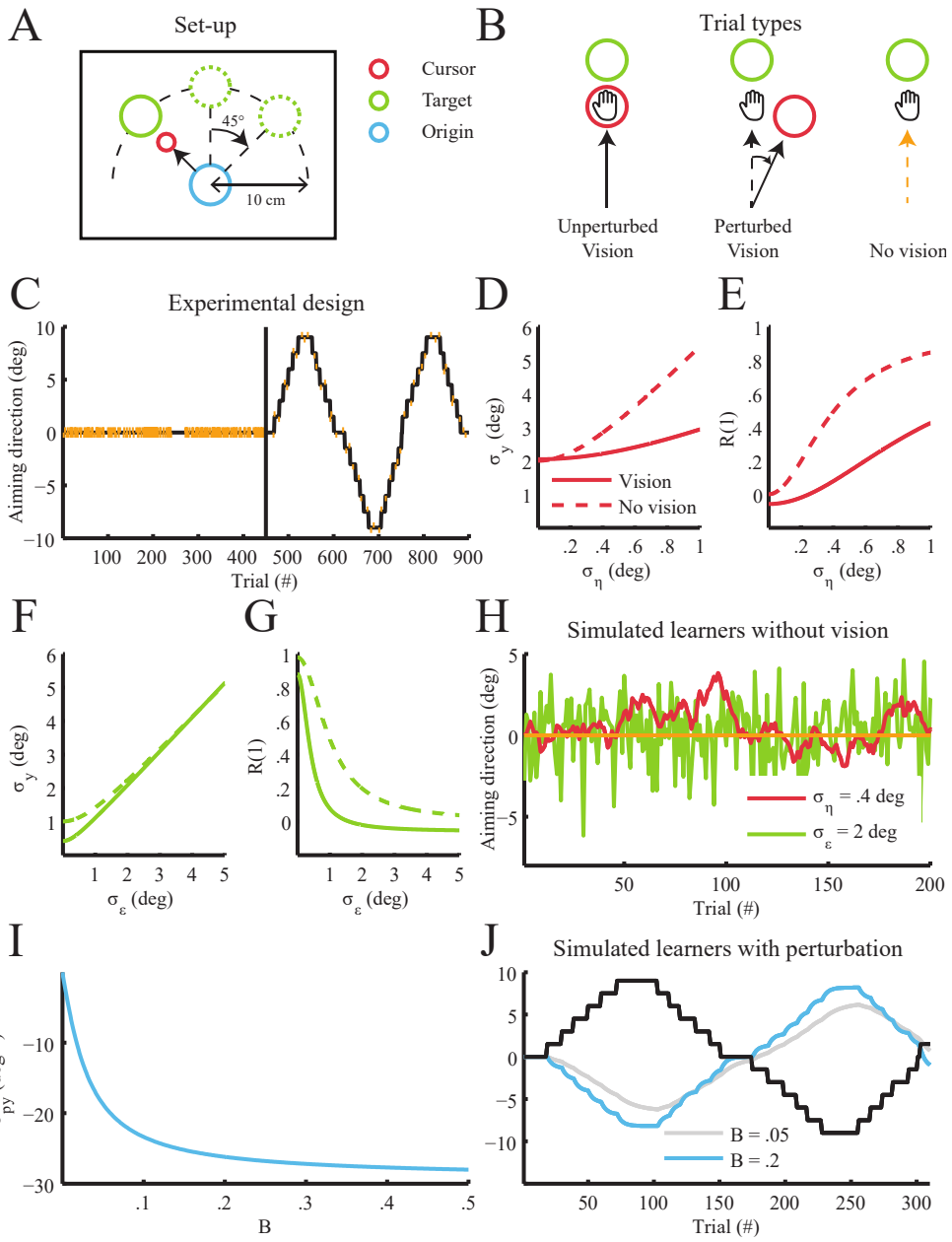


Figure 2. Measurements of planning and execution noise and adaptation rate in a visuomotor adaptation experiment. A. Set-up. The projection screen displayed the location of the robotic handle ("the cursor"), start location of the movement ("the origin"), and target of the movement ("the target") on a black background. The position of the origin on the screen was fixed throughout the experiment, while the target was placed 10 cm from the origin at an angle of -45°, 0° or 45°. **B.** Trial types. The experiment included vision unperturbed and perturbed trials and no vision trials. In vision unperturbed trials, the cursor was shown at the position of the handle during the movement. The cursor was also visible in vision perturbed trials but at a predefined angle from the vector connecting the origin and the handle. In no vision trials, the cursor was

turned off when movement onset was detected and therefore only visible at the start of movement to help subjects keep the cursor at the origin. **C.** Experimental design. The baseline block consisted of 225 vision unperturbed trials and 225 no vision trials (indicated by vertical red lines). The perturbation block had 50 no vision trials and 400 vision trials, with every block of nine trials containing one no vision trial. Most vision trials were perturbed vision trials whose perturbation magnitudes formed a staircase running from -9 to 9°. **D.** Simulation of planning noise σ_η and standard deviation σ_y of the movement angle. σ_y increases with σ_η. Calculated for $A = 0.98$ and $\sigma_\epsilon = 2°$ with $B = 0.2$ for the solid line and $B = 0$ for the dashed line. **E.** Simulation of planning noise σ_η and lag-1 autocorrelation $R(1)$ of the movement angle. $R(1)$ increases with σ_η. Calculated for $A = 0.98$ and $\sigma_\epsilon = 2°$ with $B = 0.2$ for the solid line and $B = 0$ for the dashed line. **F.** Simulation of execution noise σ_ϵ and standard deviation σ_y of the movement angle. σ_y increases with σ_ϵ Calculated for $A = 0.98$ and $\sigma_\eta = 0.2°$ with $B = 0.2$ for the solid line and $B = 0$ for the dashed line. **G.** Simulation of execution noise σ_ϵ and lag-1 autocorrelation $R(1)$ of the movement angle. $R(1)$ decreases with σ_ϵ Calculated for $A = 0.98$ and $\sigma_\eta = 0.2°$ with $B = 0.2$ for the solid line and $B = 0$ for the dashed line. **H.** Simulated learners without vision. The green and red traces show a single realization of two learners with either high planning noise (red learner $\sigma_\eta = 0.4°$ and $\sigma_\epsilon = 0°$) or high execution noise (green learner $\sigma_\eta = 0°$ and $\sigma_\epsilon = 2°$). Both sources increase the movement noise, but planning noise leads to correlated noise whereas execution noise leads to uncorrelated noise. This property can be seen from the relation between sequential trials. For the red learner sequential trials are often in the same (positive or negative) direction. For the green learner sequential trials are in random directions. This is captured by the lag-1 autocorrelation. **I.** Simulation of σ_{py} between the perturbation p and movement angle y, and adaptation rate B. σ_{py} gets more negative for increasing B (simulated with $A = 0.98$). **J.** Simulated learners with perturbation. The gray and blue lines show a simulated slow ($A = 0.98, B = 0.05$) and fast learner ($A = 0.98, B = 0.2$). The fast learner tracks the perturbation signal more closely than the slow learner. This property is captured by the covariance between the perturbation and the movement angle.

Statistics

Our statistical approach is a Bayesian approach (an excellent introduction to Bayesian statistics for a non-technical audience can be found in Kruschke [28]). We used this approach to fit the state-space model described in equations (1)-(4) because it offers a number of advantages over the expectation-maximization algorithm used in previous studies [10,11]. Perhaps the most important advantage of the Bayesian approach is that it naturally allows hierarchical modeling which shares data across subjects, allowing greater regularization of the parameter fits for each subject, as well as simultaneous estimates of the population distribution of the parameters [29,30]. In a classical approach, each subject's parameters are generally estimated independently and the uncertainty in those estimates is often not propagated forward when calculating population estimates. Indeed, the output of a Bayesian approach is not the best possible estimate of the parameter or even a maximum-likelihood estimate with a confidence interval, but rather a sampling from the parameter's probability distribution given the data [31]. This allows the analysis to naturally refocus on parameter uncertainty rather than focusing on point estimates [32–34]. The difficulty with point estimates has been a focus of much debate in the current discussion of the reproducibility crisis in science [35,36]. The Bayesian approach also estimates the hidden (state) variables simultaneously with the parameters, rather than creating a somewhat arbitrary distinction between imputation and estimation [37,38]. This allows analysis of how the state variable estimates change with the parameter estimates, an analysis that is tricky to do with an expectation-maximization approach. Finally, the Bayesian approach allows great flexibility in specifying the form of the model [31]. This can be useful in defining constraints on the model parameters or transforming variables to lie in more relevant parameter spaces, as defined below.

Modern Bayesian approaches rely on a family of algorithms called the Markov chain Monte Carlo (MCMC) algorithms [39]. These algorithms require definitions of the likelihood function (how the data would be generated if we knew the parameters) and the prior probability for the parameters (generally chosen to be broad and uninformative, but see below), and return samples from the posterior joint-probability function of the parameters. Thus, once the model and priors are specified, the output of the MCMC algorithm is a large matrix where each row is a sample and each column is one of the parameters in the model. These samples can be, then, summarized in different ways to generate parameter estimates (usually the mean of the samples but often the mode) and regions of uncertainty (very often a 95% region called the high density interval (HDI) which contains 95% of the posterior samples but also obeys the criterion that every sample in the HDI is more probable than every sample outside of it). They can also be used to assess asymmetry in the parameter distributions and covariance in the parameter estimates.

As outlined above, the Bayesian approach to state-space modeling we have taken, requires us to define priors on the model parameters. We will justify our choices in the following section. The adaptation parameters $B[s]$ and retention parameters $A[s]$ were sampled in the logistic space instead of the regular 0-1 space:

$$A[s] \sim \frac{1}{1+\exp\left(-N(\mu_A, \sigma_A^2)\right)}, \qquad B[s] \sim \frac{1}{1+\exp\left(-N(\mu_B, \sigma_B^2)\right)} \tag{5}$$

The logistic space spreads the range from 0-1 all the way from $-\infty$ to $+\infty$. This means that the distance between 0.1 and 0.01 and 0.001 are all similar in the logistic space, as are the distances between 0.9, 0.99 and 0.999. This space, thus, reflects much more accurately the real effects of changes in the parameter than we would have if we sampled in the untransformed space. This leads to much better sampling behavior and, thus, greater accuracy and less bias in the results. The priors for $A[s]$ and $B[s]$ were not actually specified in the description of the model. Only their shape was determined (normal in the logistic space). The actual prior was chosen by sampling hyperparameters for these normal distributions. For the hyperparameters, we did need to choose a specific prior, and here we choose highly uninformative priors in order to allow the posterior distribution to be influenced primarily by the data:

$$\mu_A \sim N(0, 10^3), \qquad \mu_B \sim N(0, 10^3) \tag{6}$$

$$\sigma_A^2 \sim \sigma_B^2 \sim 1/\Gamma(10^{-3}, 10^{-3}) \tag{7}$$

The sensitivity analysis (described below) showed that the choice to sample $A[s]$ and $B[s]$ from a normal distribution in the logistic space had no strong effect on the results. Following the standard Bayesian approach [28], we sampled the precision (inverse of the variance) and used a very broad gamma distribution as a prior for the precision.

$$\sigma_\eta^2[s] \sim 1/\Gamma(10^{-3}, 10^{-3}), \qquad \sigma_\epsilon^2[s] \sim 1/\Gamma(10^{-3}, 10^{-3}) \tag{8}$$

One reason the gamma distribution is a popular prior for the precision is that it is a conjugate prior which makes the algorithm more efficient. In any case, other choices of prior did not change our results in a meaningful way (see sensitivity analysis below).

MCMC sampling for the Bayesian state-space model was implemented in OpenBUGS (ver 3.2.3, OpenBUGS Foundation available from: http://www.openbugs.net/w/Downloads) with three 50,000 samples chains and 20,000 burn-in samples. A single estimate per subject s was made for $A[s]$ and $B[s]$, $\sigma_\eta^2[s]$ and $\sigma_\epsilon^2[s]$. We used all 150,000 MCMC samples that represent the posterior distribution of the model parameters $B[s]$, $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ given the data to calculate linear regressions and correlations between the model parameters across subjects. Results were presented as the mode of the effect size (either the correlation coefficients r or regression coefficient β) with 95%HDIs. Parameter estimates are plotted as the mode with 68% HDIs, similar to the standard deviation interval.

To demonstrate the test-retest properties of the Bayesian state-space model, we simulated two datasets with 50 learners on the visuomotor adaptation task outlined above. The first (optimal) dataset was simulated by drawing model parameters from the following distributions: $A[s] \sim N(0.97, 10^{-4})$, $\sigma_\eta[s] \sim N(0.6, 0.04)$, $\sigma_\epsilon[s] \sim N(3, 0.5625)$ and calculating $B[s]$ as the Kalman gain. The goal of this analysis was to determine the test-retest correlations of the model parameters $B[s]$, $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ and the ability to correctly estimate the relations between $B[s]$ and the noise parameters. For the second (permuted) dataset A[s], $\sigma_\eta[s]$, and $\sigma_\epsilon[s]$ were kept constant but $B[s]$ was permuted between learners. The motivation for this analysis was to show that our Bayesian state-space model does not introduce false relations between B and the noise parameters.

To evaluate the sensitivity of the main results to alternate prior distributions for the Bayesian state-space model, we repeated the entire analysis with (alternative priors 1) t-distributions with the hyperparameter for the degrees of freedom sampled from an exponential distribution (in line with recommendations from Kruschke [33]) as priors for $A[s]$ and $B[s]$, (alternative priors 2) t-distributions as priors for $A[s]$ and $B[s]$ and uniform distributions in the range $[0, 20]$ as priors for σ_η and σ_ϵ (in line with recommendations from Gelman [30]), and (alternative priors 3) beta distributions with hyperparameters sampled from gamma distributions as priors for $A[s]$ and $B[s]$ and uniform distributions as priors for σ_η and σ_ϵ. Finally, we addressed the concern that the between-subjects correlations of the model parameters might arise from within-subject correlations of the model parameters by permuting the MCMC samples differently for each parameter and recalculating the correlation and regression coefficients. The permuted distribution of the model parameters has the property that all correlations between the parameters within-subjects are zero.

Code Accessibility

BUGS/ JAGS code for the Bayesian state-space model can be accessed without restrictions at: https://github.com/rickvandervliet/Bayesian-state-space.

Results

Simulations

We designed a visuomotor adaptation task [40] to (1) fit the state-space model of adaptation and (2) investigate the validity of the parameter estimates $B[s]$, $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ by correlating the estimates with the variance statistics of the data (see Figure 2A-C).

The baseline block was designed to extract the standard deviation and the lag-1 autocorrelation of the movement direction and relate these measures to the parameter estimates

of $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$. The standard deviation and lag-1 autocorrelation in our baseline block are well-approximated by the following expressions:

$$\sigma_y = \sqrt{\left(\sigma_\epsilon^2 + \sum_{k=0}^{\infty}(A-B)^{2k}\sigma_\eta^2 + \sum_{k=0}^{\infty}(A-B)^{2k}B^2\sigma_\epsilon^2\right)} \tag{9}$$

$$R(1) = \frac{\sum_{k=0}^{\infty}(A-B)^{2k+1}\sigma_\eta^2 + B\sigma_\epsilon^2 + \sum_{k=0}^{\infty}(A-B)^{2k+1}B^2\sigma_\epsilon^2}{\sum_{k=0}^{\infty}A^k(A-B)^k\sigma_\eta^2 + \sigma_\epsilon^2 + \sum_{k=0}^{\infty}A^k(A-B)^kB^2\sigma_\epsilon^2} \tag{10}$$

In addition, we included a control segment of 30 trials without vision (B = 0), to calculate estimates of the standard deviation and lag-1 autocorrelation which are independent of the adaptation rate B:

$$\sigma_y = \sqrt{\left(\sigma_\epsilon^2 + \sum_{k=0}^{\infty}A^{2k}\sigma_\eta^2\right)} \tag{11}$$

$$R(1) = \frac{\sum_{k=0}^{\infty}(A^{2k+1}\sigma_\eta^2)}{\sigma_\epsilon^2 + \sum_{k=0}^{\infty}A^{2k}\sigma_\eta^2} \tag{12}$$

Both for the expressions with vision (9)-(10) (solid lines) and without vision (11)-(12) (dashed lines), standard deviation σ_y increases with planning noise σ_η (see simulations in Figure 2D) and

	$\boldsymbol{\sigma_\eta[s]}$ $(\boldsymbol{\beta})$	$\boldsymbol{\sigma_\epsilon[s]}$ $(\boldsymbol{\beta})$	$\boldsymbol{K[s]}$ $(\boldsymbol{r})$
Main analysis	0.44 [0.27 0.59]	-0.39 [-0.50 -0.30]	0.54 [0.38 0.66]
Alternative priors 1	0.44 [0.26 0.60]	-0.40 [-0.50 -0.29]	0.53 [0.38 0.66]
Alternative priors 2	0.45 [0.27 0.61]	-0.40 [-0.51 -0.30]	0.53 [0.37 0.66]
Alternative priors 3	0.44 [0.28 0.60]	-0.40 [-0.51 -0.30]	0.53 [0.38 0.66]
Permuted samples	0.29 [0.10 0.45]	-0.38 [-0.50 -0.24]	0.38 [0.21 0.66]

Table 1. Sensitivity and control analyses. For the main analysis, we used logistic normal distributions with hyperparameters sampled from normal and gamma distributions as priors for $A[s]$ and $B[s]$ and inverse gamma distributions as priors for $\sigma_\eta^2[s]$ and $\sigma_\epsilon^2[s]$. For the sensitivity analysis, we used (alternative priors 1) t-distributions with the hyperparameter for the degrees of freedom sampled from an exponential distribution as priors for $A[s]$ and $B[s]$, (alternative priors 2) t-distributions as priors for $A[s]$ and $B[s]$ and uniform distributions in the range $[0, 20]$ as priors for σ_η and σ_ϵ, and (alternative priors 3) beta distributions with hyperparameters sampled from gamma distributions as priors for $A[s]$ and $B[s]$ and uniform distributions as priors for σ_η and σ_ϵ. Finally, as a control analysis for within-subjects correlations of the model parameters, we recalculated the correlation and regressions coefficients after permuting the samples of the main analysis differently for each parameter.

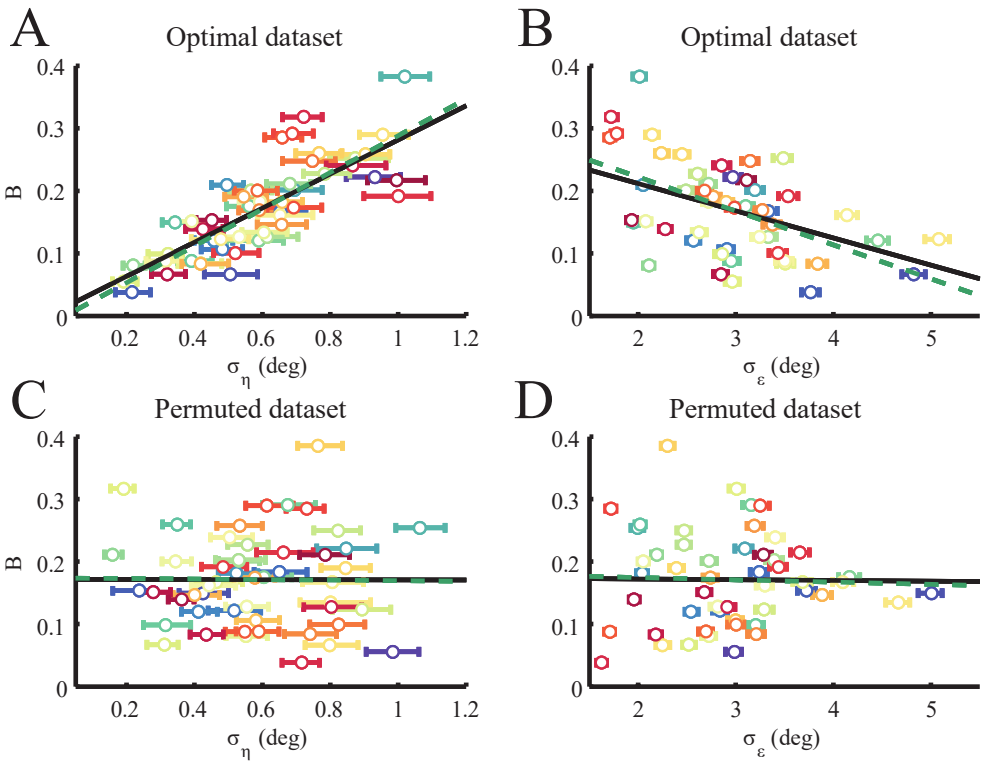


Figure 3. Test-retest properties of the Bayesian state-space model. A-B. Regression of $\boldsymbol{B}[\boldsymbol{s}]$ onto (A) $\boldsymbol{\sigma_\eta}[\boldsymbol{s}]$ and (B) $\boldsymbol{\sigma_\epsilon}[\boldsymbol{s}]$ for the simulated optimal dataset. **C-D.** Regression of $\boldsymbol{B}[\boldsymbol{s}]$ onto (C) $\boldsymbol{\sigma_\eta}[\boldsymbol{s}]$ and (D) $\boldsymbol{\sigma_\epsilon}[\boldsymbol{s}]$ for the simulated permuted dataset. Parameter estimates with 68% HDIs are shown for every simulated learner as a dot with error bars. The black solid line shows the regression on the model parameters estimated with the Bayesian state-space model, the green dashed line the regression on the original model parameters.

execution noise σ_ϵ (see simulations in Figure 2F) whereas lag-1 autocorrelation $R(1)$ increases with planning noise σ_η (see simulations in Figure 2E) but decreases with execution noise σ_ϵ (see simulations in Figure 2G), with the strongest correlations between σ_y and σ_ϵ, and $R(1)$ and σ_η. We therefore expected similar relations between the noise parameters $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$, and the standard deviation $\sigma_{y,baseline}[s]$ and lag-1 autocorrelation $R_{Baseline}(1)[s]$ of the baseline block (see simulations of planning and execution noise in the baseline block in Figure 2H).

The perturbation block was designed to extract the covariance σ_{py} between the perturbation and the movement angle from the data and relate this parameter to the adaptation rate B. The covariance σ_{py} depends solely on the learning parameters A and B and becomes increasingly negative for higher adaptation rates because learning is compensatory (see simulations in Figure 2I). Therefore, we expected a similar relation between the covariance $\sigma_{py}[s]$ and adaptation rate $B[s]$ in the perturbation block of our experiment (see simulations of two learners with a low or high adaptation rate in Figure 2J).

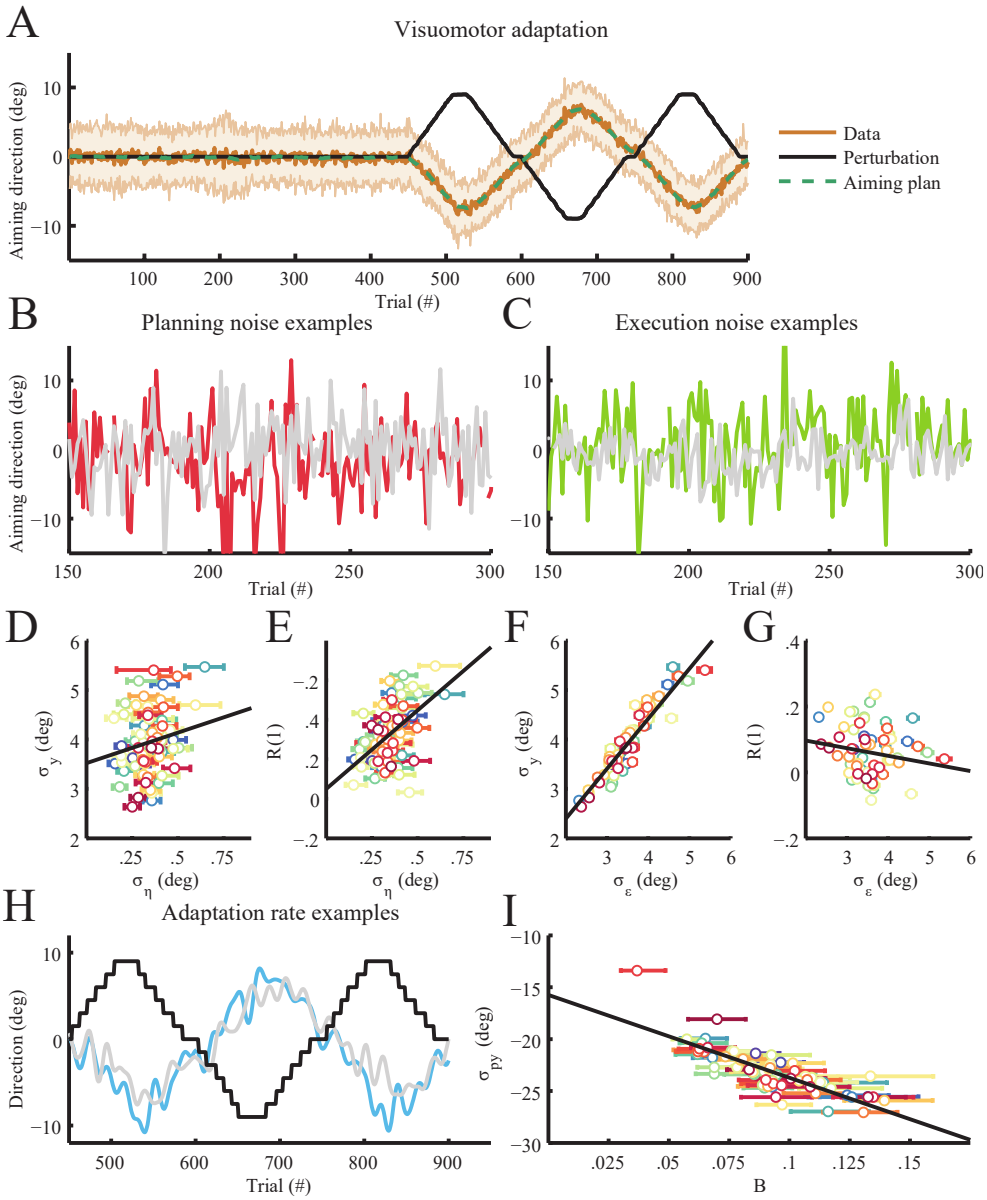


Figure 4. State-space model of visuomotor adaptation. A. Visuomotor adaptation. Average movement angle of the 69 subjects with standard deviations are shown in brown tone colors. The black line indicates the average perturbation signal, the green line the average posterior estimate of the aiming angle. **B.** Planning noise examples. The gray line shows a subject with low planning noise ($\sigma_\eta = 0.15°$ $\sigma_\epsilon = 4.6°$), the red line a subject with high planning noise ($\sigma_\eta = 0.65°$ $\sigma_\epsilon = 4.6°$). **C.** Execution noise examples. The gray line shows a subject with low execution noise ($\sigma_\eta = 0.36°$ $\sigma_\epsilon = 2.3°$), the green line a subject with high execution noise ($\sigma_\eta = 0.29°$ $\sigma_\epsilon = 5.0°$). **D.** Relation between the parameter estimate σ_η and baseline measure $\sigma_{y,baseline}$. The black line is a linear regression of $\sigma_{y,baseline}[s]$ onto $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ for average $\sigma_\epsilon[s]$. **E.** Relation between the parameter estimate σ_η and baseline measure $R(1)_{baseline}$. The black line is a linear regression of

$R(1)_{baseline}[s]$ onto $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ for average $\sigma_\epsilon[s]$. **F.** Relation between the parameter estimate σ_ϵ and baseline measure $\sigma_{y,baseline}$. The black line is a linear regression of $\sigma_{y,baseline}[s]$ onto $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ for average $\sigma_\eta[s]$. **G.** Relation between the parameter estimate σ_ϵ and baseline measure $R(1)_{baseline}$. The black line is a linear regression of $R(1)_{baseline}[s]$ onto $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ for average $\sigma_\eta[s]$. **H.** Adaptation rate examples. The thick lines show a slow (gray, $B = 0.055$) and fast subject (blue, $B = 0.14$) smoothened with a 6th order Butterworth filter. The black shows the perturbation signal for the fast subject. **I.** Relation between the parameter estimate $B[s]$ and perturbation block estimate $\sigma_{py}[s]$. Parameter estimates and 68% HDIs are shown for every subject as a dot with error bars.

Next, we designed a Bayesian state-space model to estimate the model parameters. To demonstrate the test-retest properties of this approach, we simulated one dataset with optimal learners and one dataset wherein the adaptation rate of the optimal dataset was permuted across learners. Excellent test-retest correlation were found both in the optimal dataset ($B[s]$ r = 1.00; 95%HDI = [1.00 1.00], $\sigma_\eta[s]$ r = 0.89; 95%HDI = [0.85 0.93] and $\sigma_\epsilon[s]$ r = 0.99; 95%HDI = [0.98 0.99]) and in the permuted dataset ($B[s]$ r = 1.00; 95%HDI = [1.00 1.00], $\sigma_\eta[s]$ r = 0.90; 95%HDI = [0.86 0.93] and $\sigma_\epsilon[s]$ r = 0.99; 95%HDI = [0.98 0.99]). In the optimal dataset, the Bayesian state-space model was able to uncover the relations between $B[s]$ and the noise parameters $\sigma_\eta[s]$ β = 0.73; 95%HDI = [0.68 0.77] (see Figure 3A) and $\sigma_\epsilon[s]$ β = -0.44; 95%HDI = [-0.51 -0.38]), which were 0.81 and -0.53 in the simulated data (see Figure 3B). In the permuted dataset, the Bayesian state-space model did not falsely introduce relations between $B[s]$ and the noise parameters $\sigma_\eta[s]$ β = 0; 95%HDI = [-0.09 0.08] (see Figure 3C) and $\sigma_\epsilon[s]$ β = -0.01; 95%HDI = [-0.04 0.02]), as they were -0.01 and -0.04 in the original dataset (see Figure 3D). Therefore, the Bayesian state-space model can reliably estimate the model parameters and the regression coefficients between the noise terms and the adaptation rate.

Experimental results

Sixty-nine subjects performed the visuomotor adaptation task outlined above. Overall, participants started moving 230ms IQR = [211 254]ms after target presentation and completed the movement in 290ms IQR = [251 320]ms, resulting in a trial duration of 520ms IQR = [500 534]ms with 87% of trials IQR = [84 95]% in the correct time window between 400ms and 600ms. Standard deviation of movement angle calculated across the 69 subjects illustrates the differences in movement behavior between people (Figure 4A). The group average aiming angle $x[n]$, calculated from 1,000 samples of the posterior distribution using the model (green dotted line), shows good agreement with the group average movement angle calculated directly from the data (brown solid line).

Figures 4B and 4C show example subjects with low or high planning noise $\sigma_\eta[s]$ (see Figure 4B) and low or high execution noise $\sigma_\epsilon[s]$ (see Figure 4C). We calculated the standard deviation and lag-1 autocorrelation using all trials in the baseline block and regressed these estimates onto $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$. Agreeing with our group level predictions (see Figures 2D-G), we found a positive relation between planning noise $\sigma_\eta[s]$ and standard deviation $\sigma_{y,baseline}[s]$ (β = 0.18; 95%HDI = [0.11 0.24]; see Figure 4D), between planning noise $\sigma_\eta[s]$ and lag-1 autocorrelation $R_{Baseline}(1)[s]$ (β = 0.42; 95%HDI = [0.29 0.55]; see Figure 4E) and between execution noise σ_ϵ[s] and standard deviation $\sigma_{y,baseline}[s]$ (β = 0.91; 95%HDI = [0.87 0.94]; see Figure 4F) and a negative relation between execution noise $\sigma_\epsilon[s]$ and lag-1 autocorrelation $R_{Baseline}(1)[s]$ (β = -0.14; 95%HDI = [-0.24 -0.07]; see Figure 4G). Next, we calculated the standard deviation and lag-1 autocorrelation of trials 181-210 only, which are no vision trials

where adaptation rate B = 0. Here, we found similar correlations between (1) planning noise $\sigma_\eta[s]$ and standard deviation$\sigma_{y,novision}[s]$ (β = 0.12; 95%HDI = [-0.04 0.27]), (2) planning noise $\sigma_\eta[s]$ and lag-1 autocorrelation $R_{Novision}(1)[s]$ (β = 0.22; 95%HDI = [0.07 0.35]), (3) execution noise σ_ϵ[s] and standard deviation $\sigma_{y,novision}[s]$ (β = 0.44; 95%HDI = [0.39 0.49]), and (4) execution noise $\sigma_\epsilon[s]$ and lag-1 autocorrelation $R_{Novision}(1)[s]$ (β = -0.04; 95%HDI = [-0.10 -0.01]). Example subjects with a low and high adaptation rate are shown in Figure 4H. Again, according to the model prediction (see Figure 2I), we found a negative relation between adaptation rate $B[s]$ and covariance $\sigma_{py}[s]$ on a group level (r = -0.69; 95%HDI = [-0.78 -0.60]; see Figure 4I).

Next, we investigated the relation between adaptation rate and the noise terms. The results are illustrated with scatterplots of the parameter estimates for individual subjects (Figure 5 left column), heatmaps of the parameter estimate distributions for the entire population (Figure 5 middle column) and line plots of the regression and correlation coefficient densities (Figure 5 right column). We regressed $B[s]$ onto $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ and found a positive relation between $\sigma_\eta[s]$ and $B[s]$ (β = 0.44 95%HDI=[0.27 0.59]) (see Figure 5A-C) and a negative relation between $\sigma_\epsilon[s]$ and $B[s]$ (β = -0.39 95%HDI = [-0.50 -0.30]) (see Figure 5D-F) with a variance explained of 0.32 [0.19 0.45]. This finding indicates that a significant proportion of the difference in adaptation rate between individuals can be explained from differences in their planning and execution noise with the direction of the correlations in agreement with Kalman filter theory (see Figure 1B-1C). In addition, we determined the steady-state Kalman gain for every subject from $A[s]$, $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ and correlated the steady-state Kalman gain with $B[s]$. Steady-state Kalman gain was calculated by solving the Riccati equation for the steady-state covariance $P_\infty[s]$:

$$A[s]^T P_\infty[s] A[s] - P_\infty[s] - A[s]^T P_\infty[s](P_\infty[s] + \sigma_\epsilon[s]^2)^{-1} P_\infty[s] A[s] + \sigma_\eta[s]^2 = 0 \quad (13)$$

$$K[s] = P_\infty[s]/(P_\infty[s] + \sigma_\epsilon[s]^2) \quad (14)$$

On a group level, the Kalman gain was a good approximation for the adaptation rate as the difference between the mean $K[s]$ and the mean $B[s]$ normalized with respect to the mean $B[s]$ was 10% [6.6 14]%. On an individual level, we found a positive correlation between steady-state Kalman gain $K[s]$ and $B[s]$ (r = 0.54; 95%HDI = [0.38 0.66]; see Figure 5G-I), adding support to the claim that individual differences in adaptation rate can be explained from differences in noise according to an optimal learning rule. To assess the robustness of our findings, we performed a sensitivity analysis for the model priors (see Table 1: alternative priors 1-3) and a control analysis for within-subject correlations (see Table 1: permuted samples) and found consistent results.

Finally, we investigated how planning and execution noise correlated with movement peak velocity. Execution noise originates from muscle activity and should increase with vigorous contraction when larger motor units are recruited which fire at a lower frequency and produce more unfused twitches [15,16]. Indeed, by regressing peak velocity onto the noise terms we found a negligible correlation between peak velocity and planning noise β = -0.12; 95%HDI = [-0.27 0.02] and a small positive correlation between peak velocity and execution noise β = 0.22; 95%HDI = [0.18 0.28].

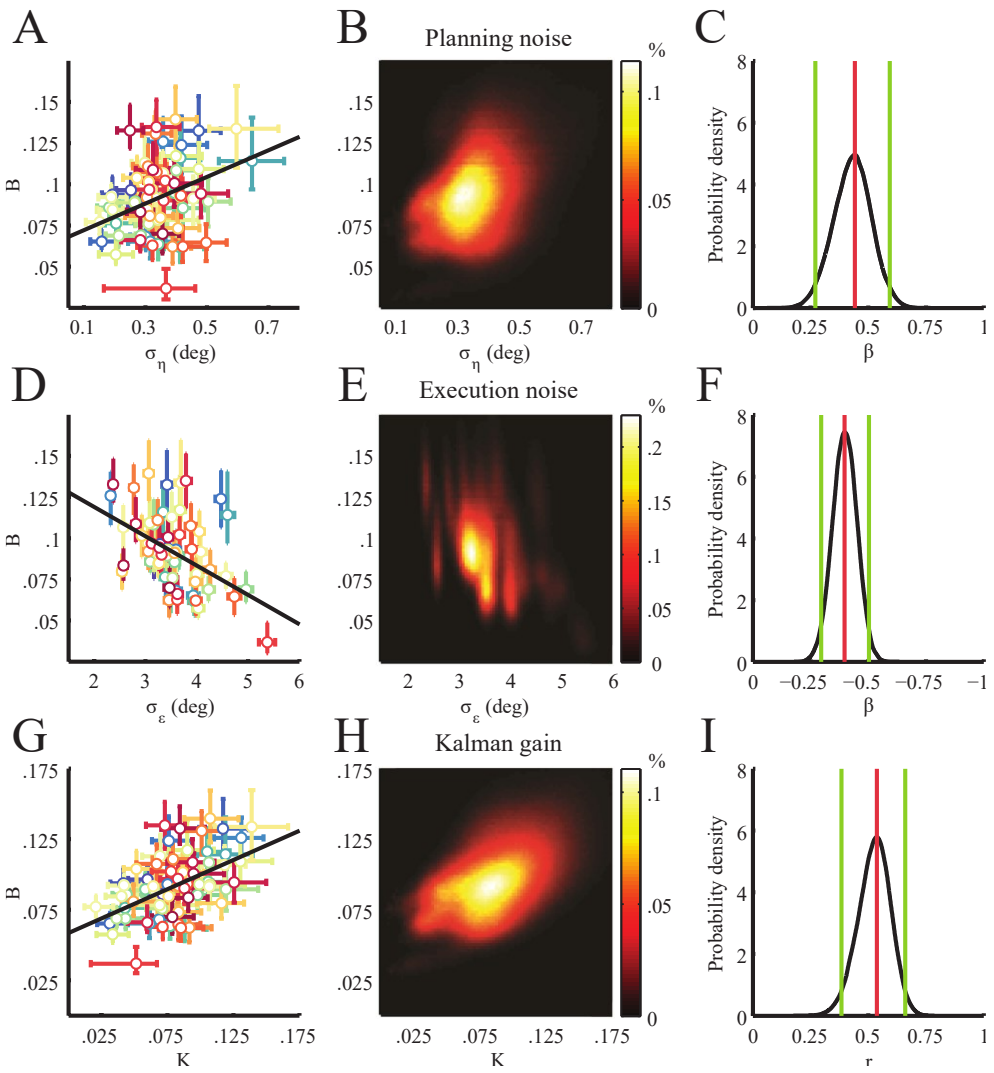


Figure 5. Relation between noise and adaptation rate. A, D, G. Scatter plots of individual parameter estimates. Parameter estimates and 68% HDIs are shown for every subject as a dot with error bars. The black line is (A) a linear regression of $B[s]$ onto $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ for average $\sigma_\epsilon[s]$, (D) a linear regression of $B[s]$ onto $\sigma_\eta[s]$ and $\sigma_\epsilon[s]$ for average $\sigma_\eta[s]$ and (G) the correlation between $K[s]$ and $B[s]$. **B, E, H.** Heatmaps of the parameter estimate distributions. The heatmaps illustrate the distribution of the parameter estimates for the entire population of 69 subjects. The intensity represents the percentage of samples in a specific range for (B) $\sigma_\eta[s]$ and $B[s]$, (E) $\sigma_\epsilon[s]$ and $B[s]$ (H) $K[s]$ and $B[s]$. **C, F, I.** Effect size densities. The black line represents the probability density of (C) the regression coefficient for $B[s]$ and $\sigma_\eta[s]$, (F) the regression coefficient for $B[s]$ and $\sigma_\epsilon[s]$, (I) the correlation coefficient for $B[s]$ and $K[s]$. The green lines indicate the 95% HDIs. The red line shows the mode.

Discussion

We investigated the relation between components of motor noise and visuomotor adaptation rate across individuals. If adaptation approximates optimal learning from movement error, it can be predicted from Kalman filter theory that planning noise correlates positively and execution noise negatively with adaptation rate[24]. To test this hypothesis, we performed a visuomotor adaptation experiment in 69 subjects and extracted planning noise, execution noise and adaptation rate using a state-space model of trial-to-trial behavior. Indeed, we found that adaptation rate correlates positively with planning noise (β = 0.44 95%HDI = [0.27 0.59]) and negatively with execution noise (β = -0.39 95%HDI = [-0.50 -0.30]). In addition, the steady-state Kalman gain calculated from planning and execution noise correlated positively with adaptation rate (r = 0.54; 95%HDI = [0.38 0.66]). We discuss implications of our findings for the optimal control model of movement and cerebellar models of adaptation and identify future applications of Bayesian state-space model fitting.

Optimal control model of movement

The optimal control model of movement has been successful in providing a unified explanation of motor control and motor learning [41]. In this framework, the motor system sets a motor goal (possibly in the prefrontal cortex) and judges its value based on expected costs and rewards in the basal ganglia [42]. Selected movements are executed in a feedback control loop involving the motor cortex and the muscles which runs on an estimate of the system's states [42]. Both the feedback controller and the state estimator are optimal in a mathematical sense. The feedback controller because it calculates optimal feedback parameters for minimizing motor costs and maximizing performance, given prescribed weighting of these two criteria [43]. The state estimator because it optimally combines sensory predictions from a forward model (cerebellum) with sensory feedback from the periphery (parietal cortex), similar to a Kalman filter [24,44]. In the optimal control model of movement, motor adaptation is defined as calibrating the forward model, which is optimal in the same sense as the state estimator [5].

Wu et al. [6], is one of the first studies to suggest that there may be a positive relationship between motor noise and motor adaptation. They outlined two apparent challenges of their findings to the optimal control approach: first, they claimed that optimal motor control is inconsistent with a positive relation between motor noise and adaptation rate; second, they claimed that optimal motor control does not account for the possibility that the motor system shapes motor noise to optimize adaptation. We take a different view. Because we find that only the planning component correlates positively with adaptation rate, our results are predicted by Kalman filter theory[24] and consistent with optimal control models of movement [41,43]. However, we do agree that the mathematical structure used to express the optimal control approach does not provide a clear way to discuss shaping noise to optimize adaptation. While this may be a technical difficulty from the point of view of optimal feedback approaches, it is apparent that there is electrophysiological evidence that some animals do shape noise to optimize adaptation. This evidence can be found in monkeys [45]. In addition, studies in Bengalese finches show that a basal ganglia-premotor loop learns a melody from reward [46] by injecting noise [47] to promote exploration [48] during training [49] and development [50]. We suggest that a similar mechanism operates in humans during adaptation. This additional tuning mechanism could be an interesting topic of future studies into optimal control models of movement.

Cerebellar model of motor adaptation

Motor adaptation is the learning process which fine tunes the forward model and is believed to take place in the olivocerebellar system [51]. How could this learning process be sensitive to planning noise and execution noise on a neuronal level?

Central to the forward model is the cerebellar Purkinje cell, which responds to selected sensory [52] and motor [53] parallel fiber input with a firing pattern reflecting kinematic properties of upcoming movements [54,55]. When Purkinje cell predictions of the upcoming kinematic properties are inaccurate, activity of neurons in the cerebellar nuclei is proportional to the prediction error. This is apparently because inhibitory Purkinje cell input cannot cancel the excitatory input from mossy fibers and the inferior olive [56]. The sensory prediction error calculated by the cerebellar nuclei could be used to update either (1) motor commands in a feedback loop with (pre)motor areas [53] or (2) state estimates of the limb in the parietal cortex [57,58]. During adaptation, parallel fiber to Purkinje cell synapses associated with predictive signals are strengthened and parallel fiber to Purkinje cell synapses associated with non-predictive signals are silenced [59]. These plasticity mechanisms are affected by climbing fibers originating from the inferior olive, which integrate input from the sensorimotor system and the cerebellar nuclei and act as a teaching signal in the olivocerebellar system [60,61].

No previous experimental or modeling work has considered how planning or execution noise might be conveyed to the cerebellum or how they might influence plasticity. We speculate that planning noise is reflected in synaptic variability of the parallel fiber to Purkinje cell synapse. Electrophysiological studies of CA1 hippocampal neurons have shown that synaptic noise can improve detection of weak signals through stochastic resonance [62]. Such a mechanism might help form appropriate connections at the parallel fiber to Purkinje cell synapse during adaptation. In addition, theoretical studies on deep learning networks have shown that gradient descent algorithms, which can be likened to error-based learning, benefit from adding noise to the gradient at every training step [63]. Furthermore, we speculate that execution noise affects adaptation through climbing fiber firing modulation. Execution noise will decrease reliability of sensory prediction errors because (1) the motor plan is not executed faithfully (motor noise) [17] and (2) the sensory feedback is inaccurate (sensory nose) (Osborne et al., 2005). Therefore, when sensory information for a specific movement plan has been unreliable in the past the olivocerebellar system might decrease its response to sensory prediction error, for example by decreasing climbing fiber firing in the inferior olive [61], which would lower the adaptation rate. The existence of such a mechanism has also been suggested by a recent behavioral study which showed a specific decline in adaptation rate for movement perturbations that had been inconsistent in the past[27].

Two-rate models of adaptation

Our results are based on a one-rate learning model of adaptation [9–11]. However, recent studies have suggested that a two-rate model composed of a slow but retentive and a fast but forgetting learning system provides a better explanation for learning phenomena such as savings and anterograde interference [64]. The fast learning system might represent an explicit process, which could be located in the cortex and the slow learning system an implicit process, which could be located in subcortical areas such as the cerebellum [65–67]. How could we interpret our results in light of these two-rate models? In a two-rate state-space model, the two systems will add to produce the movement output [64]. That is, the total adaptation rate is equal to the sum of the adaptation rates of the two systems and the same goes for the planning noise. Of course, a two-

rate model will still include only one term for execution noise. Therefore, a two-rate model can reproduce our results either if both systems are optimally tuned or if only one system is optimally tuned but is relatively dominant. With our current experimental design, we cannot differentiate between these two options. Future studies combining reporting-based approaches to discern the contributions of the implicit and explicit processes and the Bayesian statistical approach to state-space modeling presented in this paper could further unravel this question.

References

1. Frank, M.J., Doll, B.B., Oas-Terpstra, J. & Moreno, F. *Nat. Neurosci.* 12, 1062–8 (2009).
2. McHughen, S.A. et al. *Cereb. Cortex* 20, 1254–62 (2010).
3. Fritsch, B. et al. *Neuron* 66, 198–204 (2010).
4. Tomassini, V. et al. *Hum. Brain Mapp.* 32, 494–508 (2011).
5. Shadmehr, R., Smith, M.A. & Krakauer, J.W. *Annu. Rev. Neurosci.* 33, 89–108 (2010).
6. Wu, H.G., Miyamoto, Y.R., Gonzalez Castro, L.N., Ölveczky, B.P. & Smith, M.A. *Nat. Neurosci.* 17, 312–21 (2014).
7. He, K. et al. *PLOS Comput. Biol.* 12, e1005023 (2016).
8. Faisal, A.A., Selen, L.P.J. & Wolpert, D.M. *Nat. Rev. Neurosci.* 9, 292–303 (2008).
9. van Beers, R.J. *Neuron* 63, 406–17 (2009).
10. Cheng, S. & Sabes, P.N. *Neural Comput.* 18, 760–793 (2006).
11. Cheng, S. & Sabes, P.N. *J. Neurophysiol.* 97, 3057–69 (2007).
12. Chaisanguanthum, K.S., Shen, H.H. & Sabes, P.N. *J. Neurosci.* 34, 12071–80 (2014).
13. Churchland, M.M., Afshar, A. & Shenoy, K. V. *Neuron* 52, 1085–1096 (2006).
14. Haar, S., Donchin, O. & Dinstein, I. *J. Neurosci.* 37, 9076–9085 (2017).
15. Harris, C.M. & Wolpert, D.M. *Nature* 394, 780–4 (1998).
16. Jones, K.E., Hamilton, A.F. & Wolpert, D.M. *J. Neurophysiol.* 88, 1533–44 (2002).
17. van Beers, R.J., Haggard, P. & Wolpert, D.M. *J. Neurophysiol.* 91, 1050–63 (2004).
18. Osborne, L.C., Lisberger, S.G. & Bialek, W. *Nature* 437, 412–6 (2005).
19. Bialek, W. *Annu. Rev. Biophys. Biophys. Chem.* 16, 455–478 (1987).
20. van Beers, R.J., Sabes, P., Gon, J.D. van der, Ghez, C. & Brenner, E. *PLoS One* 7, e49373 (2012).
21. Baddeley, R.J., Ingram, H.A. & Miall, R.C. *J. Neurosci.* 23, 3066–75 (2003).
22. Burge, J., Ernst, M.O. & Banks, M.S. *J. Vis.* 8, 20 (2008).
23. Wei, K. & Körding, K. *Front. Comput. Neurosci.* 4, 11 (2010).
24. Kalman, R.E. *J. Basic Eng.* 82, 35 (1960).
25. Donchin, O. et al. *J. Neurophysiol.* 107, 134–47 (2012).
26. Gonzalez Castro, L.N., Hadjiosif, A.M., Hemphill, M.A. & Smith, M.A. *Curr. Biol.* 24, 1050–61 (2014).
27. Herzfeld, D.J., Vaswani, P.A., Marko, M.K. & Shadmehr, R. *Science (80-.).* 345, 1349–1353 (2014).
28. Kruschke, J.K.. (Academic Press: 2010).
29. Browne, W.J. & Draper, D. *Bayesian Anal.* 1, 473–514 (2006).
30. Gelman, A. *Bayesian Anal.* 1, 515–534 (2006).
31. Kruschke, J.K. & Liddell, T.M. *Psychon. Bull. Rev.* 25, 178–206 (2018).
32. Kruschke, J.K. & Liddell, T.M. *Psychon. Bull. Rev.* 25, 155–177 (2018).
33. Kruschke, J.K. *J. Exp. Psychol. Gen.* 142, 573–603 (2013).
34. Wagenmakers, E.-J. et al. *Behav. Res. Methods* (2014).doi:10.3758/s13428-014-0517-4
35. Ioannidis, J.P.A. *PLoS Med.* 2, e124 (2005).
36. Cumming, G. *Psychol. Sci.* 25, 7–29 (2014).
37. Carlin, B.P., Polson, N.G. & Stoffer, D.S. *J. Am. Stat. Assoc.* 87, 493 (1992).
38. Carter, C.K. & Kohn, R. *Biometrika* 81, 541–553 (1994).
39. Andrieu, C., de Freitas, N., Doucet, A. & Jordan, M.I. *Mach. Learn.* 50, 5–43 (2003).
40. Tseng, Y.-W.W., Diedrichsen, J., Krakauer, J.W., Shadmehr, R. & Bastian, A.J. *J Neurophysiol* 98, 54–62 (2007).
41. Todorov, E. & Jordan, M.I. *Nat. Neurosci.* 5, 1226–35 (2002).

42. Shadmehr, R. & Krakauer, J.W. *Exp. brain Res.* 185, 359–81 (2008).
43. Åström, K.J. (Karl J. & Murray, R.M. (Princeton University Press: 2008).
44. Wolpert, D.M., Ghahramani, Z. & Jordan, M.I. *Science* 269, 1880–2 (1995).
45. Mandelblat-Cerf, Y., Paz, R. & Vaadia, E. *J. Neurosci.* 29, 15053–62 (2009).
46. Charlesworth, J.D., Warren, T.L. & Brainard, M.S. *Nature* 486, 251–5 (2012).
47. Kao, M.H., Doupe, A.J. & Brainard, M.S. *Nature* 433, 638–43 (2005).
48. Tumer, E.C. & Brainard, M.S. *Nature* 450, 1240–4 (2007).
49. Stepanek, L. & Doupe, A.J. *J. Neurophysiol.* 104, 2474–86 (2010).
50. Olveczky, B.P., Andalman, A.S. & Fee, M.S. *PLoS Biol.* 3, e153 (2005).
51. De Zeeuw, C.I. et al. *Nat. Rev. Neurosci.* 12, 327–344 (2011).
52. Chabrol, F.P., Arenz, A., Wiechert, M.T., Margrie, T.W. & DiGregorio, D.A. *Nat. Neurosci.* 18, 718–727 (2015).
53. Kelly, R.M. & Strick, P.L. *J. Neurosci.* 23, (2003).
54. Herzfeld, D.J., Kojima, Y., Soetedjo, R. & Shadmehr, R. *Nature* 526, 439–442 (2015).
55. Pasalar, S., Roitman, A. V, Durfee, W.K. & Ebner, T.J. *Nat. Neurosci.* 9, 1404–1411 (2006).
56. Brooks, J.X., Carriot, J. & Cullen, K.E. *Nat. Neurosci.* 18, 1310–7 (2015).
57. Clower, D.M., West, R.A., Lynch, J.C. & Strick, P.L. *J. Neurosci.* 21, 6283–91 (2001).
58. Grafton, S.T. et al. *Nat. Neurosci.* 2, 563–567 (1999).
59. Dean, P., Porrill, J., Ekerot, C.-F. & Jörntell, H. *Nat. Rev. Neurosci.* 11, 30–43 (2010).
60. Ohmae, S. & Medina, J.F. *Nat. Neurosci.* 18, 1798–1803 (2015).
61. De Zeeuw, C.I. et al. *Trends Neurosci.* 21, 391–400 (1998).
62. Stacey, W.C. & Durand, D.M. *J. Neurophysiol.* 83, 1394–402 (2000).
63. Neelakantan, A. et al. (2015).at <http://arxiv.org/abs/1511.06807>
64. Smith, M.A., Ghazizadeh, A. & Shadmehr, R. *PLoS Biol.* 4, e179 (2006).
65. Mazzoni, P. & Krakauer, J.W. *J. Neurosci.* 26, (2006).
66. Taylor, J.A., Krakauer, J.W. & Ivry, R.B. *J. Neurosci.* 34, 3023–32 (2014).
67. McDougle, S.D., Bond, K.M. & Taylor, J.A. *J. Neurosci.* 35, 9568–79 (2015).

2.2 Frontal midline theta activity acts as a bottom-up alarm signal and not as a top-down teaching signal in the context of motor adaptation

Zeb D. Jonker, Rick van der Vliet, Guido Maquelin, Joris van der Cruijsen, Gerard M. Ribbers, Ruud W. Selles, Opher Donchin and Maarten A. Frens

Abstract

Feedback-related negativity and the underlying theta activity (4-8Hz) have been predominantly studied in cognitive decision-making tasks with a binary outcome (failure or success). Recently, feedback-related FMθ has also been found in the context of motor adaptation tasks. However, whether this FMθ is actively involved in trial-to-trial visuomotor adaptation is still an unanswered question. To answer this question, 60 healthy participants (19 men and 41 women) performed a trial-to-trial visuomotor experiment with a continuous outcome, while their frontal midline theta activity was measured with EEG. On the trial level, we investigated whether FMθ serves as a 'top-down teaching signal' or as a 'bottom-up alarm signal'. Additionally, on the participant level, we explored if the relation between error size and subsequent FMθ-power (EEG-error sensitivity) differs between individuals and whether these differences are associated with individual differences in motor adaptation parameters: planning noise, execution noise and adaptation rate. We found that the frontal midline theta activity in each trial was best explained by the absolute error in the corresponding trial and not by the correction in the following trial. This result indicates that frontal midline theta activity acts as a 'bottom-up alarm signal', and not as a 'top-down teaching signal'. Furthermore, we found that individual differences in EEG-error sensitivity were negatively related execution noise and positively related to adaptation rate. This result indicates that the EEG-error sensitivity of individuals is tuned to the distribution of errors that an individual naturally produces.

Introduction

Since the discovery of feedback-related negativity (FRN) [1] in electro-encephalography (EEG) recordings over the anterior cingulate cortex [2], the FRN and the underlying theta activity [3] have been predominantly investigated in cognitive decision-making tasks with a binary outcome (failure or success). In the context of these decision-making tasks, frontal midline theta activity (FMθ) is thought to represent a surprise signal [4]. However, whether this surprise signal serves as a 'top-down teaching signal' or as a 'bottom-up alarm signal' is still an unanswered question. As a teaching signal, FMθ may indicate the weight that an error in the current prediction should have on a future prediction, whereas, as an alarm signal, FMθ may indicate the need for cognitive control without necessarily defining how this control should be asserted [4].

Recently, feedback-related FMθ has also been found in the context of motor adaptation tasks [5–7]. By using perturbations of different sizes, these studies showed a positive relation between the absolute size of the error feedback and the subsequent EEG activity (EEG-error sensitivity). Analogous to a 'top-down teaching signal', Arrighi et al., (2016) suggested that FMθ might represent a strengthening mechanism that boosts the visuomotor remapping in downstream brain regions, such as the olivo-cerebellar system [8,9]. However, Arrighi et al., (2016) noted that their study, which allowed for online movement correction, was not designed to disentangle the involvement of FMθ in error detection, error correction or both.

Therefore, the goal of this study was to investigate the involvement of FMθ in trial-to-trial visuomotor adaptation. On the trial level, we tested whether FMθ is related to the absolute error in the corresponding trial (bottom-up), the error correction in the following trial (top-down) or both. Additionally, on the participant level, we explored if the relation between error size and the subsequent FMθ-power (EEG-error sensitivity) differs between individuals and whether these differences are associated with individual differences in motor adaptation parameters: planning noise, execution noise and adaptation rate [10–13].

Methods

Participants

We included 60 right-handed [14] participants without any medical conditions that might interfere with motor performance (19 men and 41 women; mean age = 25.6 years, range = 18-61). Participants received a small financial token for travelling and time compensation. The study was performed in accordance with the Declaration of Helsinki and was approved by the medical ethics committee of the Erasmus MC University Medical Centre (study identification number: MEC-2019-0042).

Experimental procedure

The experimental procedure was adapted from the procedure described in Van der Vliet et al., (2018). Participants were seated in front of a horizontal projection screen with a robotic handle underneath. This handle was situated at elbow height and could be moved in the horizontal plane. The position of the handle was projected on top of the screen as a cursor (green circle 5mm diameter). Furthermore, the screen displayed an origin (white circle 10mm diameter) and a target (red circle 10mm diameter) at fixed positions. The origin was located approximately 40cm straight in front of the participant and the target was projected exactly 10cm behind the origin, approximately 50cm in front of the participant. Participants were instructed to hold the handle in their dominant right hand and move the cursor from the origin through the target in one

smooth reaching movement. To prevent direct feedback of hand position underneath the screen, participants were wearing an apron which was secured to the top of the screen.

The experiment consisted of three types of trials: no-vision, unperturbed, and perturbed trials (Figure 1A). At the start of each trial, participants held the cursor in the origin. The target appeared after one second, indicating that the participant should start the movement. In all trial types, the cursor disappeared when the handle left the origin. In no-vision trials, the cursor did not reappear during the entire movement. In unperturbed and perturbed trials, the cursor reappeared when the handle distance from the origin exceeded 5cm. However, in perturbed trials, the cursor was projected at a predefined angle relative to the actual handle position.

The movement was damped with a force cushion (3.6Ns/m, ramped up over 7.5ms) when the handle position exceeded 10cm distance from the origin, and thus exceeded the target distance. We defined this time point as movement offset. In perturbed and unperturbed trials, the cursor froze at this time point to provide feedback on the endpoint error. Furthermore, the target changed color. We instructed participants to move through the target in 400-600ms after the target appeared, to prevent adjustments during the movement. If the movement duration was too short (<400ms) the target stayed red, if the movement duration was correct (400-600ms) the target turned white, and if the duration was too long (>600ms) the target turned blue. One second after onset of the force cushion, the robot pushed the handle back to the starting position. When the handle was at the origin, the cursor reappeared at the handle position.

The experiment consisted of 900 trials, divided in two blocks of 450 trials: a baseline block followed by a perturbation block (Figure 1B). The exact order of the trial types was randomized for each participant separately. The baseline block contained 225 unperturbed trials and 225 no-vision trials in a completely randomized order. In contrast, the perturbation block contained 400 perturbation and 50 no-vision trials in a pseudorandomized order: in every epoch of 9 trials, there was 1 randomly interspersed no-vision trial. Furthermore, the perturbation angle changed from 0° to 9° to -9° and back to 0° with increments of 1.5° every 8 to 12 trials. The experiment was paused after every 150 trials for approximately 2 minutes.

Movement recording and preprocessing

The experiment was controlled by a custom C++ program. The position and velocity signals of the robot handle were sampled at 500Hz. The velocity signal was filtered with an exponential moving average filter (smoothing factor = 0.18s). Movement start time was defined as the time point when movement velocity exceeded 0.03 m/s and movement end time was defined as the moment that the hand location exceeded 10cm distance from the starting position. The hand angle was defined as the angle between the vector connecting the origin and the hand at the end of the movement, relative to the vector connecting the origin and the target. The clockwise direction was defined as positive. Trials were removed if movement duration exceeded 1s or if the hand angle exceeded 30 degrees. On average, 2.6% (range [0 24]) of the 900 movement traces were excluded per participant.

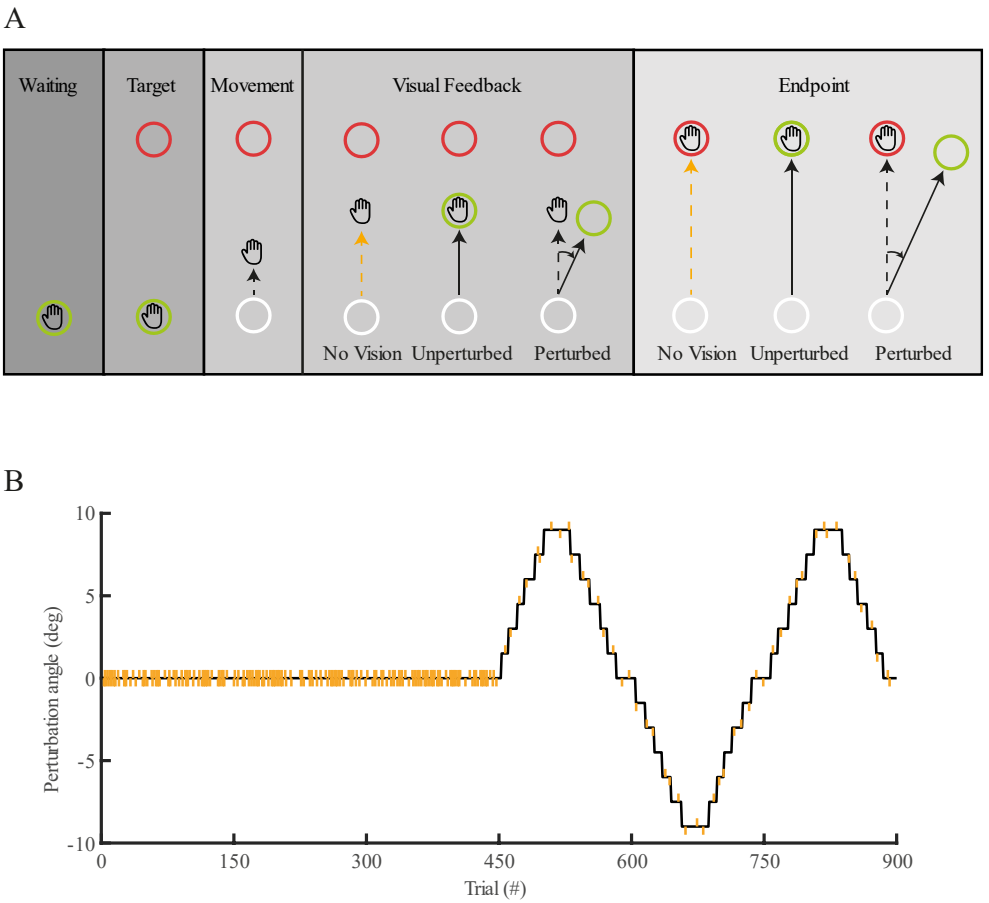


Figure 1. Design of the experiment. A. Trial design. The projection screen displayed the location of the robotic handle ("the cursor", green circle), start location of the movement ("the origin", white circle), and target of the movement ("the target", red circle) on a black background. The position of the origin and the target were fixed throughout the experiment. The experiment included three types of trials: no-vision, unperturbed and perturbed trials. In all trials the cursor was turned off when movement onset was detected. In no-vision trials, the cursor was not shown during the entire movement. In unperturbed and perturbed trials, the cursor reappeared halfway the movement. However, in perturbed trials the cursor was projected at a predefined angle. **B.** Experimental design. The experiment consisted of a baseline part and a perturbation part. The baseline part consisted of 225 unperturbed trials and 225 no-vision trials (indicated by vertical yellow lines). The perturbation part had 50 no-vision trials and 400 vision trials, with every block of nine trials containing 1 no-vision trial. Most vision trials were perturbed vision trials whose perturbation magnitudes formed a staircase running from –9° to 9°.

Movement analysis

In the movement analysis we estimated the execution noise, planning noise and adaptation rate of each participant. The analysis was performed with Bayesian Markov Chain Monte Carlo simulations in JAGS [15] and is extensively described in van der Vliet et al., (2018). In short, we fitted a state-space model of trial-to-trial behavior adapted from Cheng and Sabes, (2006):

$$x[s,n+1] = A[s] * x[s,n] - B[s] * e[s,n] + \eta[s,n] \quad (1)$$

$$y[s,n] = x[s,n] + \varepsilon[s,n] \quad (2)$$

$$e[s,n] = y[s,n] + p[s,n] \quad (3)$$

$$\eta[s,n] \sim N(0, \sigma_\eta^2[s]) \quad (4)$$

$$\varepsilon[s,n] \sim N(0, \sigma_\varepsilon^2[s]) \quad (5)$$

For each trial n of participant s, $x[s,n]$ is the movement plan, $y[s,n]$ the angle of the hand relative to the target at the endpoint, $p[s,n]$ the perturbation angle and $e[s,n]$ the error angle of the cursor on the screen relative to the target. Participant-specific motor adaptation parameters estimated with this model are the retention rate $A[s]$, which is the fractional retention of the movement plan $x[s,n]$ in the previous trial, and the adaptation rate $B[s]$, which is the fractional change caused by the error $e[s,n]$ in the previous trials. The noise terms include planning noise $\eta[s,n]$, and execution noise $\varepsilon[s,n]$. Planning noise and execution noise are modeled as zero-mean Gaussians. The standard deviations of these Gaussians $\sigma_\eta[s]$ and $\sigma_\varepsilon[s]$ represent the magnitude of planning and execution noise for each participant. The participant-specific motor learning parameters were estimated hierarchically with uninformative priors. Subsequently, we calculated the error distribution of each participant: the standard deviation of the distribution of signed errors.

EEG recording and preprocessing

Participants were wearing a 128 channel EEG cap connected to a 136 channel REFA system (TMSi, Oldenzaal, The Netherlands). The EEG data were recorded with a sampling rate of 2048Hz from 64 channels: FP1, FPz, FP2, AF7, AF3, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7 FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO5, PO3, POz, PO4, PO6, PO8, O1, Oz, O2, and 2 EOG electrodes. The recording of each channel was referenced to the average signal across all channels.

After recording, the EEG signals were preprocessed with the EEGLAB toolbox (version 14.1.2b; Swartz Center for Computational Neuroscience, La Jolla, USA) and custom scripts in MATLAB version 2018b (Mathworks, Natick, USA). The raw average referenced EEG signals were digitally filtered between 2-40Hz with a 6th order Butterworth filter, cut into trial epochs of 2000ms centered around the onset of visual feedback, and down-sampled to 128Hz. Non-stereotypical artifacts were automatically removed by excluding trials in which the average log power was an outlier compared to the other trials within the same channel (absolute z-score >3.5). Similarly, stereotypical artifacts were automatically removed by performing a fast-independent component analysis [17] and excluding components in which the average log power was an outlier (absolute z-score >3.5). The other components were projected back to the channel space. On average 38.2 (range [13 79]) of the 900 EEG traces were excluded and 1.1 (range [1 2]) of the 64 components were removed.

The EEG traces were filtered with a surface Laplacian to improve the spatial resolution The Legendre polynomial order was 50 and the smoothing parameter was 1E-5 [18]. Subsequently, the traces were decomposed into their time-frequency representations using Morlet wavelet

convolution. The wavelet frequencies ranged from 2 to 30 Hz in 29 steps of 1Hz and the number of cycles ranged from 4 to 10 in 29 logarithmically-spaced steps [19]. After convolution, the traces were trimmed to 1800ms around the onset of visual feedback to remove edge artifacts introduced by the concatenation. Consecutively, the log power at each time-frequency point within each trial was normalized as a percentage change relative to the average log power of that frequency in the predefined baseline window (from 900ms until 600ms before visual feedback) across all trials. Frontal midline theta activity was defined as the normalized log power in channel FCz averaged over the 4-8Hz frequencies [3]. The processing window was defined as the 600ms after group average movement offset (78-678ms after the onset visual feedback).

EEG analysis

On the trial level, we investigated whether the frontocentral theta activity (FMθ) in the processing window is related to the absolute error in the corresponding trial, the correction in the following trial, or both. Absolute errors were defined as the absolute error angle between the cursor on the screen relative to the target. Corrections were defined as the hand angle in the following trial minus the hand angle in the corresponding trial, divided by the sign (1 or -1) of the error angle in the corresponding trial. Thus, negative values represent corrections in the direction of the target, whereas positive values represent corrections in the opposite direction. While estimating the relation of FMθ with error detection and error correction, we controlled for other factors that might influence FMθ, such as the trial number (duration of the experiment) and the feedback on the trial duration (color change of the target). We used the linear mixed model function in MATLAB version 2018b (Mathworks, Natick, USA) to estimate within-participant effects across participants.

$$FM\theta[s,n] \sim \alpha[s] + D[s] * |e|[s,n] + C[s] * c[s,n] + \\ + T[s] * t[s,n] + F_{short}[s] * f_{short}[s,n] + F_{long}[s] * f_{long}[s,n] \quad (6)$$

For each trial n of participant s, $FM\theta[s,n]$ is the frontal midline theta activity in the processing window, $|e|[s,n]$ is the absolute error in the corresponding trial, $c[s,n]$ is the correction in the following trial, $t[s,n]$ is the original trial number and $f_{short}[s,n]$ and $f_{long}[s,n]$ are binary variables indicating if the trial was "too short" ($f_{short}[s,n] = 1$) or "too long" ($f_{long}[s,n] = 1$). In order to compare the strength of the relations, all variables were z-score normalized. The intercept $\alpha[s]$, the relation with absolute error size (EEG-error sensitivity) $D[s]$, the relation with error correction $C[s]$, the effect of trial number $T[s]$, and the effect of the color feedback $F_{short}[s]$ and $F_{long}[s]$ were all modelled as random effects.

We used a backward elimination method to determine which variables best explain the frontocentral theta activity. Stepwise removal of variables with the smalles effect size was continued as long as the Bayesian Information Criterion decreased, and the last model for which this criterion held was selected. To check the assumption of a linear relation, we plotted $FM\theta$ against the absolute error and the error correction. Based on visual inspection, we included all trials with visual feedback (unperturbed and unperturbed), an absolute error smaller than 5 degrees, and an error correction between -7 and +3 degrees. On average 478, range [264 587], artifact-free trials per participant were included. To check for collinearity, we calculated the relation between the absolute error and the error correction in the following trial with a separate hierarchical model:

$$c[s,n] \sim \alpha[s] + \beta[s] * |e|[s,n] \quad (7)$$

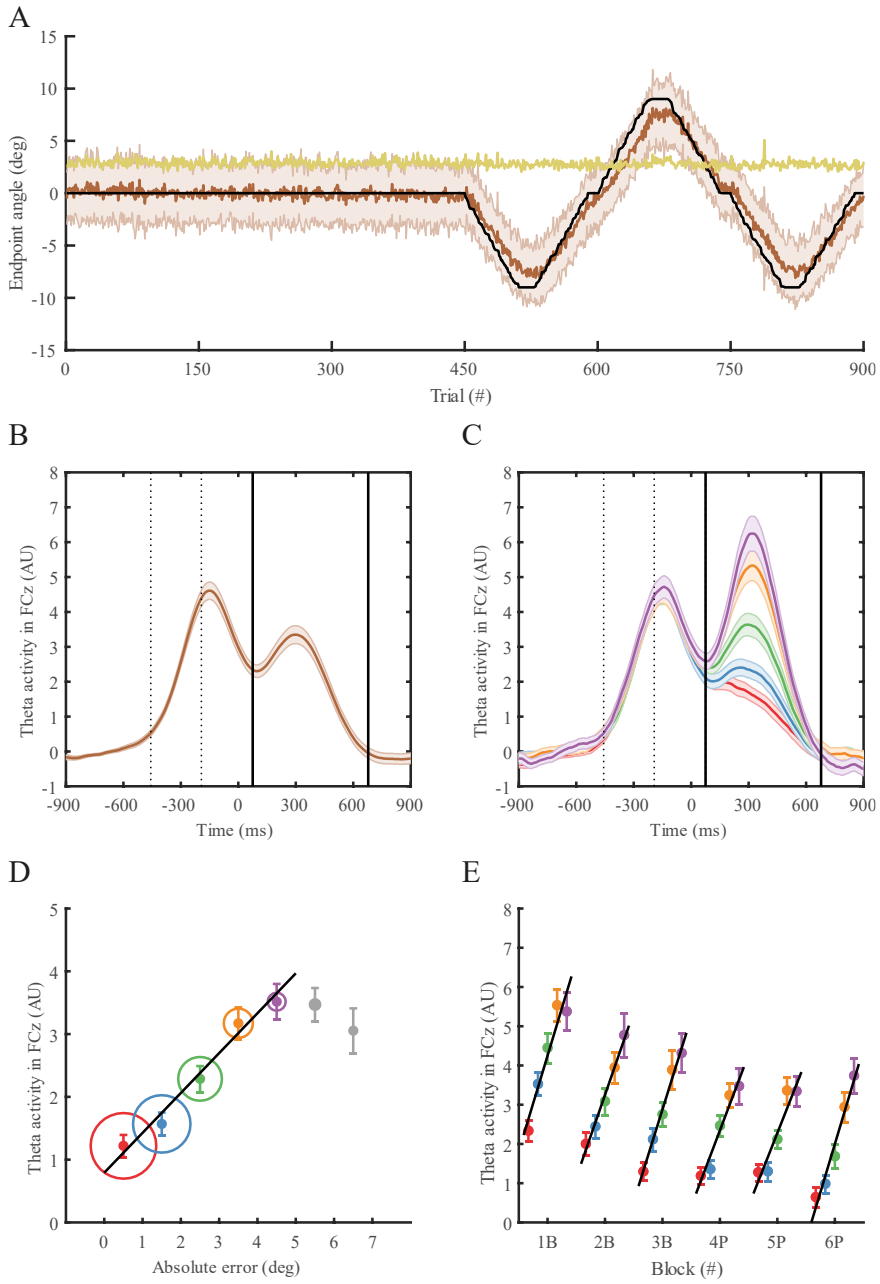


Figure 2. Group average data. A. Motor learning throughout the experiment. The brown line indicates the average hand angle of all 60 participants. The error bar represents the standard deviation. The black line indicates the negative of the average perturbation signal. The yellow line indicates the standard deviation of the signed errors. B-D: EEG-error sensitivity. Error bars represent the standard error of the mean **B.** Grand average frontal midline theta activity of all trials with visual feedback (perturbed and unperturbed). The dotted lines indicate the average target onset (-387±24ms) and the average movement onset (-179±24ms). The long vertical lines represent the borders of the 600ms processing window, starting at the average

movement offset (78±15ms). **C.** Grand average frontal midline theta activity of trials with visual feedback sorted on absolute error. Red is [0 1], blue is [1 2], green is [2 3], orange is [3 4] and purple is [4 5] degrees. **D.** Average EEG-error sensitivity. Colored dots indicate the average frontal midline theta activity in the processing window (panel C). The slope of the black line represents the linear relation between absolute error size and frontal midline theta activity in the processing window i.e. EEG-error sensitivity. The diameters of the open circles represent the relative number of trials in a trial bin. **E.** Average EEG-error sensitivity in each subblock of 150 trials.

Individual differences in EEG-error sensitivity and motor adaptation

On the participant level, we explored whether individual differences in EEG-error sensitivity are related to individual differences in the motor adaptation paramters by visualizing the data and by calculating the Spearman correlation coefficients. For the motor adaptation parameters, we used the point estimates from the movement analysis. For the EEG-error sensitivity, we used the point estimates ($D[s]$) from a compact version of the linear mixed model described in the EEG analysis:

$$FM\theta[s,n] \sim \alpha[s] + D[s] * |e|[s,n] + T[s] * t[s,n] \quad (8)$$

In this compact version, we included all trials with visual feedback (unperturbed and unperturbed), and an absolute error smaller than 5 degrees, regardless of the error correction in the following trial (Average = 539, range [395 602])

Finally, we used the point estimates from the movement analysis to investigate the relation between the individual differences in adaptation rate and the two noise terms (planning noise and execution noise) with a multiple linear regression model [13].

$$B[s] \sim \alpha + \beta 1 * \sigma_\eta[s] + \beta 2 * \sigma_\varepsilon[s] \quad (9)$$

Step	Absolute error D	Trial number T	Error correction C	Trial duration F_{Short}	Trial duration F_{long}	BIC
1	0.13 [0.10 0.16]	-0.10 [-0.13 -0.08]	-0.02 [-0.03 -0.01]	0.01 [-0.00 0.01]	0.00 [-0.01 0.0]	78595
2	0.13 [0.10 0.16]	-0.10 [-0.13 -0.08]	-0.02 [-0.03 -0.01]	0.01 [-0.00 0.02]	-	78575
3	0.13 [0.10 0.16]	-0.10 [-0.13 -0.08]	-0.02 [-0.03 -0.01]	-	-	78557
4	0.14 [0.11 0.17]	-0.10 [-0.13 -0.08]	-	-	-	78553
5	0.14 [0.10 0.16]	-	-	-	-	79030

Table 1. Results from EEG analysis. A backward elimination method was applied to select the model which best explained the frontal medial theta activity in each trial. The full model (Equation 6) included the relation with absolute error size (EEG-error sensitivity) $D[s]$, the relation with error correction $C[s]$, the effect of trial number $T[s]$, and the effect of the color feedback $F_{short}[s]$ and $F_{long}[s]$. The model with the lowest Bayesian Information Criterion (BIC) included only the absolute error in the corresponding trial and the trial number.

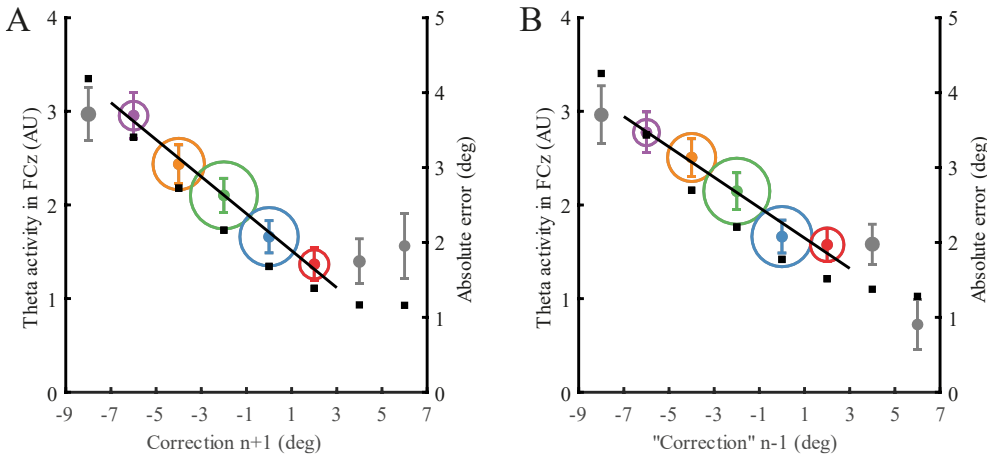


Figure 3. Uncorrected association between frontal midline theta activity (FMθ) and the correction in the following trial. Trials with visual feedback are sorted based on the error correction in the following trial. Purple is [-7 -5], orange is [-5 -3], green is [-3 -1], blue is [-1 1] and purple is [1 3] degrees. Dots indicate the average frontal midline theta activity in the processing window and the diameters. Error bars represent the standard error of the mean (n=60). The open circles represent the relative number of trials in a bin. The slope of the black line represents the linear association between frontal midline theta activity and error correction. Black squares indicate the average abolute error in each bin. **A.** Uncorrected association between FMθ and error correction in the following trial. **B.** Uncorrected association between FMθ and 'error correction' erroneously calculated back in time, using the preceding trial instead of the following trial.

Results

Participants adapted their reaching movements to the perturbation, as illustrated by the average aiming direction in Figure 2A. Furthermore, this figure illustrates the similarity in the average error distribution between the baseline block and the perturbation block.

EEG-error sensitivity

The frontal midline theta activity (FMθ) in each trial was best explained by the absolute error in the corresponding trial and the trial number and not by the correction in the following trial nor the feedback on the trial duration (Table 1). Post-feedback peak in frontal midline theta activity started around the movement offset and ended approximately 600ms later (see Figure 2B). Larger absolute errors resulted in higher frontal midline theta activity (EEG-error sensitivity) with an approximately linear relation for absolute errors up to 5 degrees (Figure 2C). EEG-error sensitivity was present in both the baseline and the perturbation blocks (see Figure 2D). Furthermore, the overall frontal midline theta activity in the processing window decreased throughout the experiment (see Figure 2D).

The uncorrected association between FMθ and error correction is a reflection of the relation between FMθ and the absolute error size (see Figure 3). The association between FMθ and error correction is approximately linear for corrections between -7 and +3 degrees (see Figure 3A). However, the error correction in the following trial is correlated to the absolute error in the corresponding trial ($\boldsymbol{\beta}$ = -0.32) as illustrated by the black squares. Furthermore, the association between FMθ and error correction in the following trial is comparable to the association between FMθ a hypothetical error correction in the trial preceding trial. This argues against a causal role for FMθ in error correction.

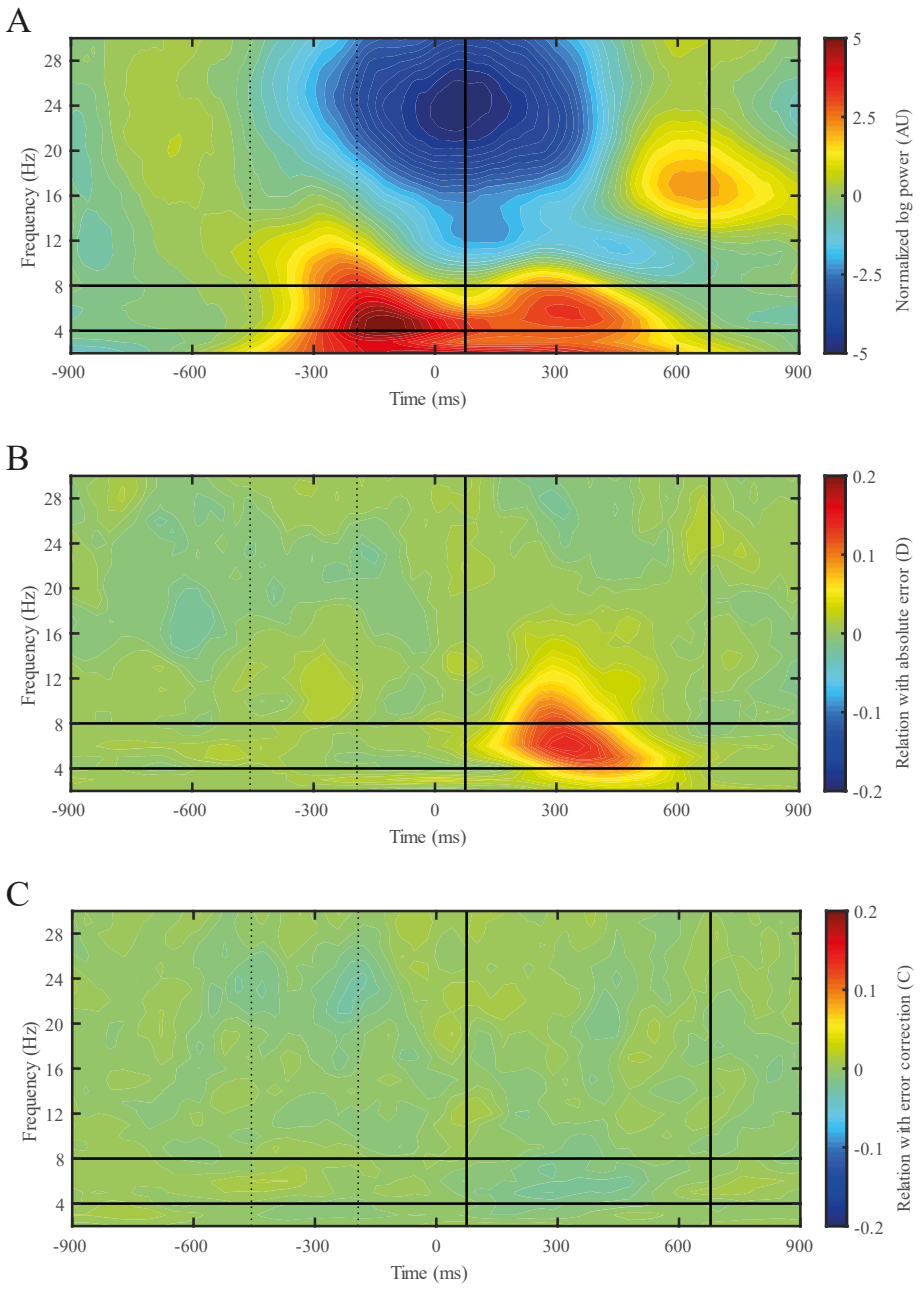


Figure 4. Time-frequency plot of channel FCz. Equation 6 was calculated for each time-frequency point. **A.** Average normalized log power in channel FCz. The broken lines indicate the average target onset and the average movement offset. The solid lines indicate the borders of the processing window and the borders of the theta frequency band. **B.** Group average relation between EEG activity and absolute error size in the corresponding trial (EEG-error sensitivity) for each time-frequency point in channel FCz. **C.** Group average relation between EEG activity and correction in the following trial for each time-frequency point in channel FCz.

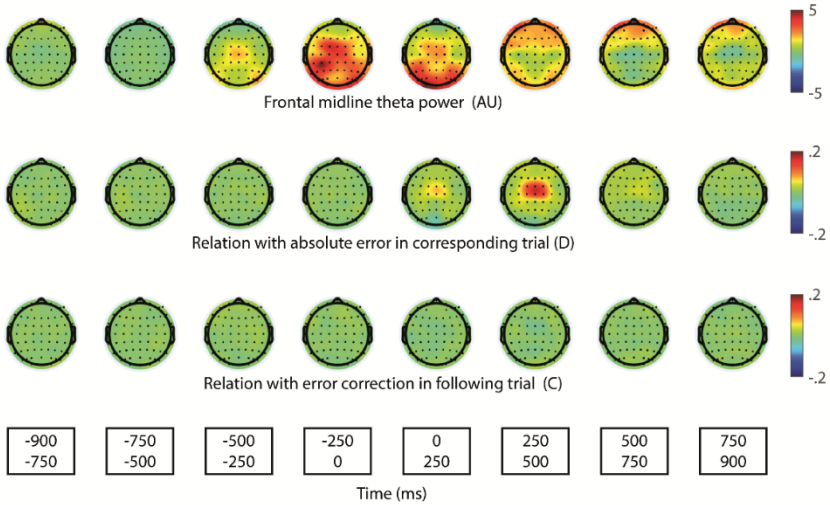


Figure 5. Topographic plot of theta activity (4-8Hz). Each topographic plot represents the average of a time window. Equation 6 was calculated for each channel in each timewindow. First row: Group average log normalized theta activity over the scalp. Second row: Group average relation between theta activity and absolute error size in the corresponding trial (EEG-error sensitivity) Third row: Group average relation between theta activity and correction in the following trial.

Our pre-defined frequency window (4-8 Hz) and pre-defined electrode channel (FCz) to quantify EEG-error sensitivity were both found to be appropriate (see Figures 4 and 5). EEG-error sensitivity is most prominent in the theta band and does not become apparent in any of the other frequency bands (see Figure 4). With regards to electrode selection, EEG-error sensitivity in the theta band is indeed most prominent in the channel FCz and comparatively unremarkable in other brain regions (see Figure 5).

Individual differences in EEG-error sensitivity and motor learning

The group average execution noise was 2.52±0.58 degrees; the group average planning noise was 0.48±0.13 degrees; the group average adaptation rate was 0.13±0.04 and the group average error distribution was 2.95±0.60 degrees. Figure 6 shows the associations between the individual differences in EEG-error sensitivity and the individual differences in the motor learning parameters. Individual differences in EEG-error sensitivity were weakly correlated with adaptation rate (Figure 6C) and weakly anticorrelated with execution noise (Figure 6B) and the error distribution (Figure 6D). Figure 7 shows the data of four participants with different levels of EEG error sensitivity in more detail. Finally, individual differences in adaptation rate were positively related to planning noise and negatively related to execution noise (see Figure 8).

Discussion

This study demonstrates that, in the context of visuomotor adaptation, FMθ does not act as a 'top-down teaching signal', but rather as 'bottom-up alarm signal'. The EEG analysis showed that the feedback-related frontal midline theta activity (FMθ) in each trial was better explained by the absolute error size in the corresponding trial, than by the correction in the following trial, or a

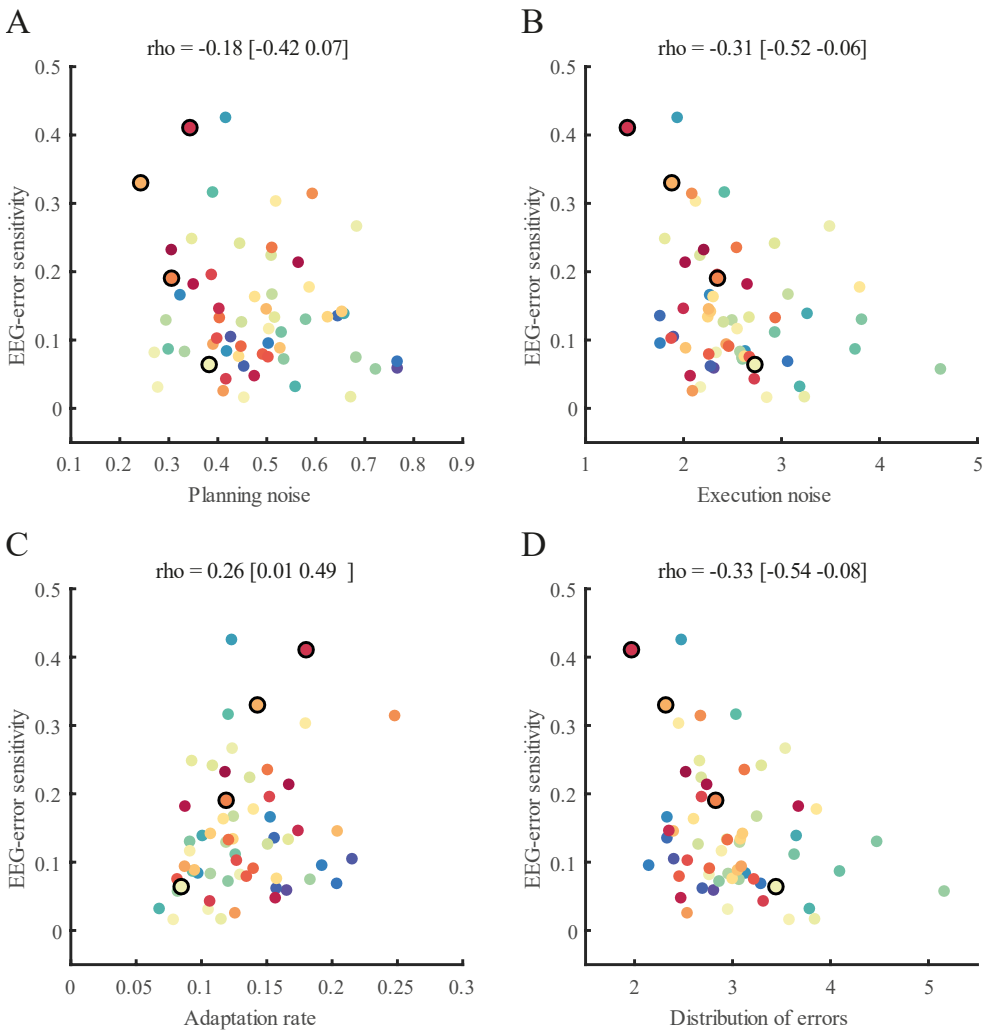


Figure 6. Relation between the individual differences in adaptation rate, and planning noise and execution noise. Each colored dot represents a participant. The black lines depict the regression lines of the multiple linear regression. **A.** Relation between adaptation rate and planning noise, corrected for execution noise. Each dot represents the residual adaptation rate, after correction for the estimated effect of execution noise. The other end of the grey lines represents the adaptation rate before the correction. If the adaptation rate of the participants would be perfectly described by their execution noise, all colored dots would have an y-value of zero. **B.** Relation between adaptation rate and planning noise, corrected for planning noise.

combination of both variables. The positive relation between frontal midline EEG activity and the absolute error size (EEG-error sensitivity) corroborates earlier work by Anguera (2009), orrecillos (2014) and (Arrighi 2016). This study expands their results in two ways. First of all, this study shows that EEG-error sensitivity is also present in the absence of external perturbations i.e. in response to small self-made errors during natural movements. Second, this study shows that FMθ is directly involved in error detection, but not actively involved in error correction.

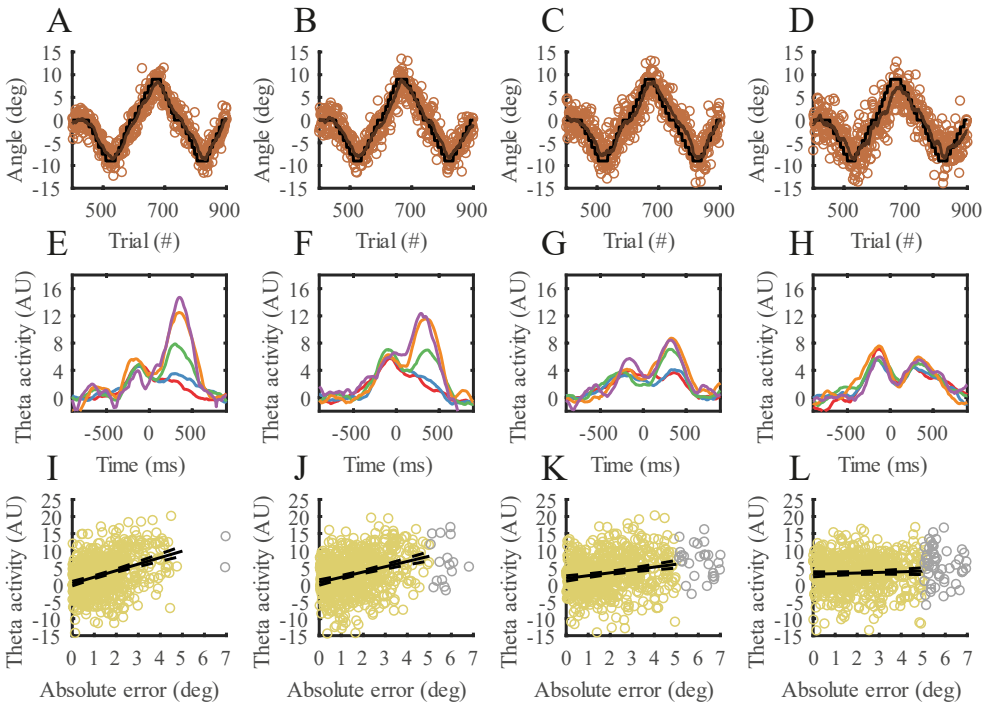


Figure 7. Individual participant data. Each column represents a participant. **A-D.** Motor learning throughout the experiment. The black line is the negative of the perturbation signal. Brown circles represent the hand angle in individual trials and the brown line is the walking average over 20 trials. **E-H**. Average frontal midline theta activity of trials with visual feedback sorted on absolute error. Red is [0 1], blue is [1 2], green is [2 3], orange is [3 4] and purple is [4 5] degrees. **I-L.** EEG error sensitivity. Each circle indicates the average frontal midline theta activity in the processing window of a single trial. Yellow circels represent trials with an absolute error lower than or equal to 5 degrees. The slope of the black line represents the linear relation between absolute error size and frontal midline theta activity in the processing window i.e. EEG-error sensitivity.

Furthermore, on the participant level, we found that the individual differences in EEG-error sensitivity were positively associated with adaptation rate and negatively associated with execution noise. These associations are in line with the notion that FMθ represents a surprise signal [4]: an individual with a low execution noise and a high adaptation rate naturally produces a small distribution of errors. Therefore, for this individual the same absolute error size is more surprising than for someone who naturally moves more variably.

In agreement with our previous visuomotor experiment [13], we found that individual differences in adaptation rate are positively related to planning noise and negatively related to execution noise. This confirms that the results of our previous visuomotor experiment can be reproduced and thus strengthens the evidence that the adaptation rate of individuals is tuned to approximate the optimal learning rate according to Kalman filter theory [10–12].

Now that we have shown that FMθ acts as a 'bottom-up alarm signal', the most logical next step is to investigate what the role of this alarm signal is in the context of motor adaptation. We propose that when FMθ, becomes strong enough it may cause individuals to become aware of a change in the environment and apply an explicit strategy to overcome this change. It is

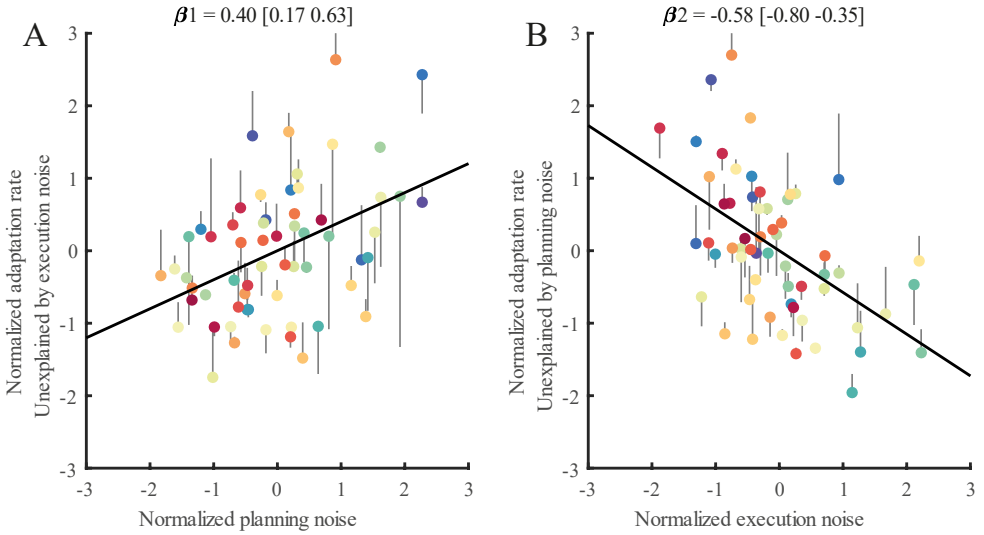


Figure 8. Uncorrected associations between the individual differences in EEG-error sensitivity and the motor learning parameters. Each colored dot represents a participant. The black open circles indicate the participants that are highlighted in Figure 8. The Spearman correlations are shown above each panel. **A.** Relation between EEG-error sensitivity and planning noise. **B.** Relation between EEG-error sensitivity and execution noise. **C.** Relation between EEG-error sensitivity and adaptation rate. **D.** Relation between EEG-error sensitivity and the distribution of signed errors.

thought that motor adaptation results from a combination of implicit learning and explicit learning [20]. Werner et al., (2019) reported that a small gradual perturbation signal induces predominantly implicit learning, whereas a large continuous perturbation signal induces a combination of implicit and explicit learning. Savoie et al., (2018), reported that FMθ during reaching movements was higher when participants were made aware of a perturbation and instructed how to counteract it. However, whether conversely FMθ after the movement leads to awareness has not yet been studied. Therefore, an interesting follow up experiment would be to investigate whether feedback related FMθ predicts the use of explicit learning in a motor adaptation task with a large continuous perturbation signal.

Another interesting follow up experiment would be to investigate the relation between EEG-error sensitivity in the theta (4-8Hz) the beta band (15-30Hz). Multiple studies have shown EEG-error sensitivity in the post movement beta synchronization over the sensorimotor area [23–26]. However, in the current study we were not able to explore post movement beta synchronization in great detail, because the exact moment participants stopped moving was obscured by the force cushion and the pushback to the starting position.

In summary, this study demonstrates that, in the context of motor adaptation, FMθ does not act as a 'top-down teaching signal', but rather as 'bottom-up alarm signal'. Furthermore, this study shows that individual differences in EEG-error sensitivity (the sensitivity of FMθ to error feedback) are negatively associated with the distribution of errors that each individual produces. Finally, this study confirmed that individual differences in adaptation rate are negatively related to planning noise and positively related to execution noise.

References

1. Miltner, W.H.R., Braun, C.H. & Coles, M.G.H. *J. Cogn. Neurosci.* 9, 788–798 (1997).
2. Dehaene, S., Posner, M.I. & Tucker, D.M. *Psychol. Sci.* 5, 303–305 (1994).
3. Cavanagh, J.F., Zambrano-Vazquez, L. & Allen, J.J.B. *Psychophysiology* 49, 220–238 (2012).
4. Cavanagh, J.F. & Frank, M.J. *Trends Cogn. Sci.* 18, 414–421 (2014).
5. Anguera, J.A., Seidler, R.D. & Gehring, W.J. *J. Neurophysiol.* 102, 1868–1879 (2009).
6. Torrecillos, F., Albouy, P., Brochier, T. & Malfait, N. *J. Neurosci.* 34, 4845–4856 (2014).
7. Arrighi, P. et al. *PLoS One* 11, 1–27 (2016).
8. Shadmehr, R. & Krakauer, J.W. *Exp. brain Res.* 185, 359–81 (2008).
9. De Zeeuw, C.I. et al. *Nat. Rev. Neurosci.* 12, 327–344 (2011).
10. Kalman, R.E. *J. Basic Eng.* 82, 35 (1960).
11. Todorov, E. & Jordan, M.I. *Nat. Neurosci.* 5, 1226–35 (2002).
12. van Beers, R.J. *Neuron* 63, 406–17 (2009).
13. van der Vliet, R. et al. *eNeuro* 5, ENEURO.0170-18.2018 (2018).
14. Oldfield, R.C. *Neuropsychologia* 9, 97–113 (1971).
15. Plummer, M.0–41 (2017).
16. Cheng, S. & Sabes, P.N. *Neural Comput.* 18, 760–793 (2006).
17. Delorme, A., Sejnowski, T. & Makeig, S. *Neuroimage* 34, 1443–1449 (2007).
18. Perrin, F., Pernier, J., Bertrand, O. & Echallier, J.F. *… Clin. Neurophysiol.* 184–187 (1989).
19. Cohen, M.X. *MIT Press* (2014).
20. Mazzoni, P. & Krakauer, J.W. *J. Neurosci.* 26, (2006).
21. Werner, S., Strüder, H.K. & Donchin, O. *PLoS One* 14, e0220748 (2019).
22. Savoie, F.-A., Thénault, F., Whittingstall, K. & Bernier, P.-M. *Sci. Rep.* 8, 1–16 (2018).
23. Tan, H., Jenkinson, N. & Brown, P. *J. Neurosci.* 34, 5678–5688 (2014).
24. Torrecillos, F., Alayrangues, J., Kilavik, B.E. & Malfait, N. *J. Neurosci.* 35, 12753–65 (2015).
25. Tan, H., Wade, C. & Brown, P. *J. Neurosci.* 36, 1516–1528 (2016).
26. Alayrangues, J., Torrecillos, F., Jahani, A. & Malfait, N. *Neuroimage* 184, 10–24 (2019).

Chapter 3. Proportional recovery models of stroke

3.1 Predicting upper limb motor impairment recovery after stroke: a mixture model

Rick van der Vliet, Ruud W. Selles, Eleni-Rosalina Andrinopoulou, Rinske Nijland, Gerard M. Ribbers, Maarten A. Frens, Carel Meskers and Gert Kwakkel

Abstract

Objective: Spontaneous recovery is an important determinant of upper extremity recovery after stroke, and has been described by the 70% proportional recovery rule for the Fugl-Meyer motor upper extremity (FM-UE) scale. However, this rule is criticized for overestimating the predictability of FM-UE recovery. Our objectives were to (1) develop a longitudinal mixture model of FM-UE recovery, (2) identify FM-UE recovery subgroups, and (3) internally validate the model predictions.

Methods: We developed an exponential recovery function with the following parameters: subgroup assignment probability, proportional recovery coefficient r_k, time constant in weeks τ_k, and distribution of the initial FM-UE scores. We fitted the model to FM-UE measurements of 412 first-ever ischemic stroke patients and cross-validated endpoint predictions and FM-UE recovery cluster assignment.

Results: The model distinguished five subgroups with different recovery parameters ($r_1 = 0.09$, $\tau_1 = 5.3$, $r_2 = 0.46$, $\tau_2 = 10.1$, $r_3 = 0.86$, $\tau_3 = 9.8$, $r_4 = 0.89$, $\tau_4 = 2.7$, $r_5 = 0.93$, $\tau_5 = 1.2$). Endpoint FM-UE was predicted with a median absolute error of 4.8 IQR=[1.3 12.8] at one week poststroke and 4.2 IQR=[1.3 9.8] at two weeks. Overall accuracy of assignment to the poor (subgroup one), moderate (subgroups two and three) and good (subgroups four and five) FM-UE recovery clusters was 0.79 95%ETI=[0.78 0.80] at one week poststroke and 0.81 95%ETI=[0.80 0.82] at two weeks.

Interpretation: FM-UE recovery reflects different subgroups, each with its own recovery profile. Cross-validation indicates that FM-UE endpoints and FM-UE recovery clusters can be well predicted. Results will contribute to the understanding of upper limb recovery patterns in the first six months after stroke.

Introduction

Longitudinal studies have repeatedly demonstrated the time-dependency of neurological recovery after stroke, including upper[1,2] and lower limb motor function,[3,4] visuo-spatial neglect,[5] and speech.[6] This suggests that recovery follows a predictable pattern, which is often described as spontaneous neurological recovery.[7,8] Understanding the mechanisms and individual dynamics that drive stroke recovery is vital for developing better prognostic models and more effective, personalized therapeutic interventions.[9–12]

The proportional recovery rule has been instrumental in modeling spontaneous upper extremity recovery by linking baseline motor impairment, measured with the Fugl-Meyer assessment of the upper extremity (FM-UE),[13] to the observed motor recovery, defined as the difference between the measurements early and three to six months after stroke (ΔFM-UE).[14] More specifically, the proportional recovery rule states that in three to six months (1) the majority of patients (recoverers) gain a fixed proportion, estimated between 0.55 and 0.85,[2] of their potential recovery, calculated as the difference between baseline FM-UE and the scale's maximum score of 66, while (2) the minority of patients (non-recoverers) show only very moderate improvement which cannot be linked to potential recovery.[1,2,14] Mechanistically, the key underlying difference between recoverers and non-recoverers is currently understood as the intactness of the corticospinal tract early after stroke.[15–18]

The proportional recovery rule has been criticized for a number of reasons. Recent analyses indicated that a strong correlation between baseline FM-UE and recovery can emerge even when baseline FM-UE is completely uncorrelated to endpoint FM-UE.[19,20] Therefore, even though the proportional recovery rule is not wrong,[19] it probably overstates the predictability of endpoint FM-UE.[19,20] In addition, the proportional recovery rule does not model the time course of recovery early poststroke, which means it cannot model the rate of recovery nor update predictions with repeated measurements in time. Finally, predictions of endpoint FM-UE based on the 70% proportional recovery rule for individual patients has not previously been reported.

To increase our understanding of upper extremity recovery after stroke, we need a model that (1) relates the FM-UE to potential recovery as a function of time after stroke, with (2) separate sets of parameters for different subgroups, including those that show no improvement early poststroke.[21] In this study, we developed and cross-validated a new longitudinal mixture model of FM-UE recovery, which describes different patterns of recovery over time using exponential functions, and identifies subgroups based on: (1) the degree of recovery as a fraction of potential recovery, (2) the rate of recovery, and (3) the initial FM-UE score. Our goals were to estimate (1) the number of recovery subgroups, (2) the recovery parameters for each subgroup, and (3) the predictability of endpoint FM-UE at 3-6 months poststroke, as well as subgroup assignment as a function of time poststroke. Results will contribute to the understanding and prediction of upper limb recovery patterns in the first six months after stroke.

Materials and methods

Study population

We combined FM-UE data of first-ever ischemic stroke patients collected in four different prospective cohort studies: the EXPLICIT[22], EPOS,[23] 4D-EEG[24] and EXPLORE studies. These datasets contain repeated measurements of the FM-UE scores and the exact measurement dates in days poststroke, which also differ between patients assigned to the same follow-up scheme for practical reasons. Data collection and patient characteristics of the EXPLICIT and EPOS cohorts have been described extensively elsewhere.[22,23] The 4D-EEG and EXPLORE cohorts recruited

patients with a first-ever ischemic stroke within three weeks poststroke. In the 4D-EEG study, patients were measured weekly during the first five weeks poststroke and after 8, 12 and 26 weeks. In the EXPLORE study, patients were measured 1, 2, 3, 5, 12 and 26 weeks poststroke. Inclusion criteria were comparable to the EPOS cohort. The majority of patients received standard rehabilitation treatment according to the Dutch rehabilitation guidelines, which are in agreement with current international rehabilitation guidelines.[25,26] In the EXPLICIT study, half of the patients with an unfavorable prognosis received electromyography-triggered neuromuscular stimulation and half of the patients with a favorable prognosis received modified constrained-induced movement therapy.[22] Since both of these interventions did not affect the FM-UE at any time poststroke,[22] we disregarded therapeutic intervention as a factor in the analysis. The 4D-EEG and EXPLORE studies have been approved by the medical ethics committees of the VU University Medical Center (NL 47079 029 14, for 37 patients measured) and the Leiden University Medical Center (NL39323.058.12, for 11 patients measured), respectively.

We included a patient if (1) at least two repeated measurements were available, and (2) the first and last measurement were at least 12 weeks apart. This way, we maximized the number of included patients while still being able to cross-validate predictions of endpoint FM-UE. Additional patient data were: age; gender; handedness; dominant side affected; Bamford scale (LACI/PACI/TACI)[27]; administration of alteplase (rt-PA); NIHSS (range 0-42) with item 11, extinction and inattention (range 0-2), reported separately;[28] motricity index (range 0-99) with the shoulder abduction item listed separately (dichotomized as no shoulder abduction (0) and at least some shoulder abduction (1));[29,30] and finger extension (dichotomized as no finger extension (0) and at least some finger extension (1)) as a separate item of the FM-UE (range 0-66).[31]

Longitudinal mixture model of FM-UE recovery

We designed a longitudinal model of FM-UE recovery after stroke based on the principles of proportional recovery, which are (1) a proportional relation between observed recovery over time and potential recovery at baseline (longitudinal), and (2) the existence of clinically distinct subgroups of FM-UE recovery (mixture). Longitudinal, therefore, refers to the ability of the model to handle repeated measurements over time and mixture to the ability of the model to identify different subgroups. Since FM-UE recovery follows an exponential pattern,[7] we chose an exponential function as the time-dependent element of the model, with the asymptote defined as a proportion of the potential recovery and the time constant expressed in weeks. In addition, we included an intercept which represents the FM-UE early after stroke. The mathematical expression of our model is:

$$\mu_{ij|k} = \alpha_{i|k} + r_k * \left(66 - \alpha_{i|k}\right) * \left(1 - \mathrm{e}^{-t_{ij}/\tau_k}\right) \quad (15)$$

$$y_{ij|k} \sim N\left(\mu_{ij|k}, \sigma_\epsilon^2\right) \quad (16)$$

With i the patient identification number [1 I], j the measurement identification number [1 J], and k the subgroup identification number [1 K]. The equation describes how the FM-UE ($y_{ij|k}$) for a particular patient i and measurement j is determined by the (estimated) baseline FM-UE ($\alpha_{i|k}$) plus an exponential term $r_k * \left(66 - \alpha_{i|k}\right) * \left(1 - \mathrm{e}^{-t_{ij}/\tau_k}\right)$ which increases over time t_{ij} as the patient recovers. We chose to express measurement dates poststroke in weeks by dividing the number of days poststroke by seven. The asymptote of the exponential term is determined by the potential recovery $\left(66 - \alpha_{i|k}\right)$ multiplied by the recovery coefficient r_k [0 1], which describes

how much of the potential recovery is achieved. The rate of the exponential term (i.e. how quickly the patient recovers) is defined by time constant τ_k in weeks $[1/7\ \infty)$, which signifies the time point when recovery has reached a proportion of $1 - \mathrm{e}^{-1} \approx 0.63$ of the asymptotic value. Finally, σ_ϵ^2 is the residual error variance.

Model fitting

We chose a Bayesian approach to mixture modeling rather than expectation-maximization, as Bayesian data analysis (1) focuses on parameter uncertainty rather than on point estimates, (2) estimates hidden variables (for example the subgroup identification number k) simultaneously with the parameters, and (3) offers flexibility in specifying the form of the model (for example to constrain the recovery coefficient r_k between 0 and 1).[32] Modern Bayesian approaches rely on a family of algorithms called the Markov-chain Monte-Carlo (MCMC) algorithms.[32] These algorithms require defining a likelihood function (how the data would be generated if we knew the parameters) and the prior probability distributions for the parameters, and they return samples from the posterior joint-probability function of the parameters. We chose the following prior probability distributions for the model parameters:

$$\alpha_{i|k} \sim \frac{66}{1 + \exp\left(-\mathrm{N}\left(\mu_{\alpha,k}, \sigma_{\alpha,k}^2\right)\right)} \quad (17)$$

$$\mu_{\alpha,k} \sim \mathrm{N}(0, 10^3) \quad (18)$$

$$\sigma_{\alpha,k}^2 \sim 1/\Gamma(10^{-3}, 10^{-3}) \quad (19)$$

$$r_k \sim \frac{1}{1 + \exp\left(-\mathrm{N}(0, 10^3)\right)} \quad (20)$$

$$\tau_k \sim \Gamma(10^{-3}, 10^{-3}) + 1/7 \quad (21)$$

$$k \sim Cat(K, p_k), p_k \sim Dirichlet(K, \gamma) \quad (22)$$

$$1/\sigma_\epsilon^2 \sim \Gamma(10^{-3}, 10^{-3}) \quad (23)$$

For the patient-specific baseline FM-UE $\alpha_{i|k}$, we defined a logistic normal prior distribution with the hyperparameters sampled from weakly informative normal and gamma distributions. This means that each subgroup is characterized by a specific distribution of the FM-UE early after stroke, which can be close to 0 or to the maximum of 66, or span the entire range with almost equal probability. The subgroup-specific prior distribution for the recovery coefficient r_k is also a logistic normal distribution, which spans the 0 to 1 range. Time constant τ_k, specified separately for each subgroup, has a weakly informative gamma prior distribution, shifted by $1/7$ to set the lower limit at one day. Subgroup labels k have a categorical prior distribution with hyperparameters for the subgroup assignment probability vector p_k, sampled from a Dirichlet distribution with concentration parameter γ. Finally, the precision $1/\sigma_\epsilon^2$ has a weakly informative gamma prior distribution.

MCMC sampling was used to simultaneously calculate (1) the number of subgroups in the data, and (2) the model parameters. We used the Rousseau and Mengersen criterion[33,34] as implemented by Nasserinejad et al.[35] to select the number of subgroups $K_{Optimal}$, setting the overfitted number of subgroups K at 10, the concentration parameter γ at $0.9 * d/2$ (equal to 1.8 for our study), the cut-off value for the subgroup size at 5% of the number of patients and the number of parallel chains to 10. From the parallel chains, we selected the solution which minimized the number of subgroups and maximized the total subgroup assignment probability.

Subgroup assignment probabilities were normalized to 1. The subgroups were arranged according to the recovery coefficient r_k, making r_1 the lowest and $r_{K,optimal}$ the highest recovery coefficient. The 'optimal FM-UE recovery cluster' was determined as the FM-UE recovery cluster a patient was assigned to most by the model. Goodness of fit was evaluated with the explained variance, which we calculated as one minus the residual error variance (σ_ϵ^2) divided by the total FM-UE variance across patients and measurements.

Cross-validation

Predictability of ΔFM-UE (the difference between the first and last measurements available for a particular patient) and endpoint FM-UE (last measurement available for a particular patient), as well as FM-UE recovery cluster assignment (poor, moderate or good recovery, see RESULTS section for the definitions), was estimated using the proposed model. We used cross-validation, which is a method for internal validation, to obtain correct estimates of the predictions. The study population was divided *n* (total number of patients) times into a prediction dataset containing data from only one patient and a fitting dataset containing data from all other patients. For all *n*-folds, we first ran the fitting dataset with settings $K = K_{Optimal}$ and $\gamma = 1.8$, and randomly selected 100 samples from the posterior distribution of the model parameters. In addition, we paired the five subgroups with one of the three FM-UE recovery clusters using a 1-nearest neighbor algorithm trained on the model parameters r_k, τ_k, $\mu_{\alpha,k}$ and $1/\sigma_\epsilon^2$. Next, MCMC sampling was performed for all 100 model parameter sets using the measurements available from the prediction dataset in the first one to 12 weeks poststroke (12 time points). Only patients who had at least one measurement available were included in the analysis for a specific time interval. Therefore, the number of patients available for cross-validation increased with time poststroke. Outcome measures were (1) the predicted ΔFM-UE between the first and last measurements of a patient, (2) the predicted FM-UE at the last measurement of a patient, and (3) the 'predicted FM-UE recovery cluster', defined as the FM-UE recovery cluster a patient was assigned to most by the model.

To evaluate prediction accuracy, we (1) calculated the absolute difference between the predicted and observed values, (2) correlated the predicted and observed ΔFM-UE and FM-UE, and (3) determined the accuracy of the FM-UE recovery cluster assignment (proportion of patients in the study population who were correctly assigned), the positive predictive value (proportion of patients in one of the three 'predicted FM-UE recovery clusters' who were correctly assigned), and the miss rate (proportion of patients in one of the three 'optimal FM-UE recovery clusters' who were incorrectly assigned). Note that accuracy is only defined for the entire study population whereas the positive predictive value and miss rate are defined for the three FM-UE recovery clusters separately.

Covariate model

We compared the predictive accuracy of the model presented above to a model incorporating a set of static (not changing over time) covariates: age at stroke onset, gender, Bamford classification, and alteplase treatment. The static covariates did not include right-handedness and dominant side affected as these were biased by the inclusion criteria of the cohort studies. We modeled age as a normal distribution with hyperparameters sampled from normal and gamma distributions, gender and alteplase treatment as binomial distributions with the hyperparameters sampled from beta distributions, and Bamford classification as a categorical distribution with hyperparameters sampled from Dirichlet distributions. The cross-validated

primary outcomes (absolute median error in endpoint FM-UE and ΔFM-UE, correlations between actual and observed endpoint FM-UE and ΔFM-UE, and mean accuracy of FM-UE recovery cluster assignment) of the models with and without covariates differed less than 10% at every time point poststroke. Therefore, we decided to present a simpler model without covariates.

Markov-chain Monte-Carlo sampling

MCMC sampling was implemented in JAGS 4.3.0 (available from: https://sourceforge.net/projects/mcmc-jags/). Matlab 2015a (MathWorks, Natick, Massachusetts, United States) and Matjags (available from: http://psiexp.ss.uci.edu/research/programs_data/jags/) were used for data and sample processing. Settings for determining the number of subgroups and calculating the model parameters were: $2.5x10^4$ burn-in samples and $2.5x10^4$ posterior distribution samples, 10 parallel chains, and initial guesses for the model parameters. Settings for cross-validation were: 10^3 burn-in samples, 10^4 posterior distribution samples, 1 parallel chain, and the mean model parameters estimated in step 1 as initial values for model fitting. All scripts can be accessed at: https://github.com/rickvandervliet/Bayesian-Proportional-Recovery. This website also hosts scripts which can prospectively predict FM-UE recovery profiles for individual patients based on the model presented in this paper. In addition, we have created an online application offering the same functionality in a user-friendly format: https://emcbiostatistics.shinyapps.io/LongitudinalMixtureModelFMUE/.

Results

Out of a total 479 patients in all four cohorts, we included data from 412 patients whose FM-UE had been measured at least two times, with the first and last measurements spaced at least 12 weeks apart. The 412 included patients were found to have a mean 6.1 (SD=1.9) measurements per patient, with an interval of 26.2 (SD=2.0) weeks between the first and last measurements.

FM-UE recovery cluster	**Poor**	**Moderate**		**Good**	
Subgroup	1	2	3	4	5
p_k	0.27	0.14	0.11	0.18	0.30
	[0.22 0.31]	[0.10 0.18]	[0.08 0.15]	[0.12 0.24]	[0.24 0.37]
r_k	0.09	0.46	0.86	0.89	0.93
	[0.07 0.11]	[0.43 0.50]	[0.83 0.90]	[0.87 0.90]	[0.92 0.94]
τ_k	5.3	10.1	9.8	2.7	1.2
	[2.8 9.2]	[8.4 12.3]	[8.9 10.8]	[2.5 2.8]	[1.1 1.3]
$\mu_{\alpha,k}$	-3.2	-2.1	-2.8	-1.3	0.0
	[-4.0 -2.8]	[-2.9 -1.2]	[-4.1 -1.3]	[-2.6 -0.1]	[-0.6 0.6]
$\sigma_{\alpha,k}$	0.6	2.2	3.0	2.9	2.4
	[0.3 1.5]	[1.5 3.3]	[1.7 4.8]	[2.0 4.0]	[1.9 3.0]

Table 1. Model parameters. Subgroup mean model parameters with 95%ETIs calculated over all samples. p_k subgroup assignment probability, r_k recovery coefficient, τ_k time constant in weeks, $\mu_{\alpha,k}$ mean of the initial distribution of the FM-UE in the logistic space, $\sigma_{\alpha,k}$ standard deviation of the initial distribution of the FM-UE in the logistic space.

FM-UE recovery cluster	Poor	Moderate		Good	
Subgroup	1	2	3	4	5
Patients (#)	111 [97 120]	56 [49 66]	44 [37 57]	72 [54 94]	126 [104 146]
Age (y)	63 [42 93]	65 [43 86]	60 [28 85]	64 [38 85]	66 [33 86]
Male (%)	56	58	53	53	47
Right-handed (%)	90	89	92	90	95
Dominant hand affected (%)	27	46	52	49	43
Bamford LACI-PACI-TACI (%)	28 47 25	50 37 13	55 31 14	70 22 8	64 26 9
Alteplase treatment (%)	29	18	24	15	18
NIHSS	13 [6 21]	8 [2 18]	9 [2 18]	5 [1 18]	5 [0 14]
Motricity index	5 [0 34]	28 [0 84]	23 [0 92]	55 [0 100]	66 [0 100]
Shoulder abduction (%)	23	69	51	94	95
Finger extension (%)	2	24	24	69	88

Table 2. Baseline patient clinimetric scores. Subgroup mean clinimetric scores with 95%ETIs calculated per subgroup over all samples. LACI = lacunar anterior circulation infarction, PACI = partial anterior circulation infarction, TACI = total anterior circulation infarction.

The first FM-UE had been measured within the first 72 hours for 53% of patients, within the first week for 76% of patients and within the first two weeks for 93% of patients.

The longitudinal mixture model of FM-UE recovery identified five different subgroups, with a residual error standard deviation σ_ϵ of 3.9 95%ETI=[3.7 4.0] points on the FM-UE, corresponding to a variance explained of 0.97 95%ETI=[0.97 0.98] (See Figure 1 and Table 1). Patient characteristics (age, gender, and handedness) were comparable between subgroups. Baseline clinimetric scores correlated with the recovery coefficient as expected (Table 2), that is, more favorable clinimetric scores were associated with higher recovery coefficients. For example, subgroup five, with the highest recovery coefficient, had the lowest score on the NIHSS, and the highest scores on the motricity index and the finger extension item of the FM-UE, while the opposite was true for subgroup one.

The number of patients included in the cross-validation increased with time poststroke as more patients with a baseline FM-UE became available (see Figure 2A). In addition, the median number of measurements per patient increased from two measurements at one week poststroke until five measurements eight weeks poststroke (see Figure 2B). Median future recovery, defined as endpoint FM-UE – last available FM-UE for each patient, decreased with time poststroke from 10.0 IQR=[3.0 26.3] until 2.0 IQR=[0.0 8.0] at 12 weeks poststroke (see Figure 2C). Reliability of endpoint FM-UE and ΔFM-UE predictions increased with time poststroke and was higher for endpoint FM-UE than for ΔFM-UE. The median absolute error for the predicted endpoint FM-UE was 4.8 IQR=[1.3 12.8] at one week poststroke and 4.2 IQR=[1.3 9.8] at two weeks poststroke (see Figure 2D), while the mean correlation between predicted and observed FM-UE was 0.84 95%ETI=[0.83 0.84] at one week poststroke and 0.86 95%ETI=[0.86 0.87] at two weeks poststroke (see Figure 2F). The median absolute error for the predicted ΔFM-UE was 5.2 IQR=[1.7 12.9] at one week poststroke and 4.8 IQR=[1.7 11.0] at two weeks poststroke (see Figure 2E), while the mean correlation between predicted and observed ΔFM-UE was 0.68 95%ETI=[0.67

0.69] at one week poststroke and 0.71 95%ETI=[0.71 0.72] at two weeks poststroke (see Figure 2F).

Based on the recovery coefficients (r_k), time constants (τ_k), and initial distributions ($\mu_{\alpha,k}$ and $\sigma_{\alpha,k}$), we organized the five subgroups into three main FM-UE recovery clusters with poor (subgroup one), moderate (subgroups two and three) and good (subgroups four and five) recovery profiles (See also Tables 1-2) Mean accuracy of the FM-UE recovery cluster assignment was 0.79 95%ETI=[0.78 0.80] at one week poststroke and 0.81 95%ETI=[0.80 0.82] at two weeks (see Figure 2G). Positive predictive value was high (> 0.9) for the good FM-UE recovery cluster as early as one week poststroke (see Figure 2H) and low to modest for the poor and moderate FM-UE recovery cluster at week one (0.66 95%ETI=[0.63 0.68] and 0.50 95%ETI=[0.42 0.57], respectively) and week two (0.72 95%ETI=[0.70 0.73] and 0.61 95%ETI=[0.57 0.64], respectively). The miss rate was lower than 0.1 for the poor and moderate FM-UE recovery cluster from week one onwards, while the miss rate for the moderate cluster was much higher at one week (0.74 95%ETI=[0.68 0.79]) and two weeks poststroke (0.63 95%ETI=[0.60 0.65]) (see Figure 2I).

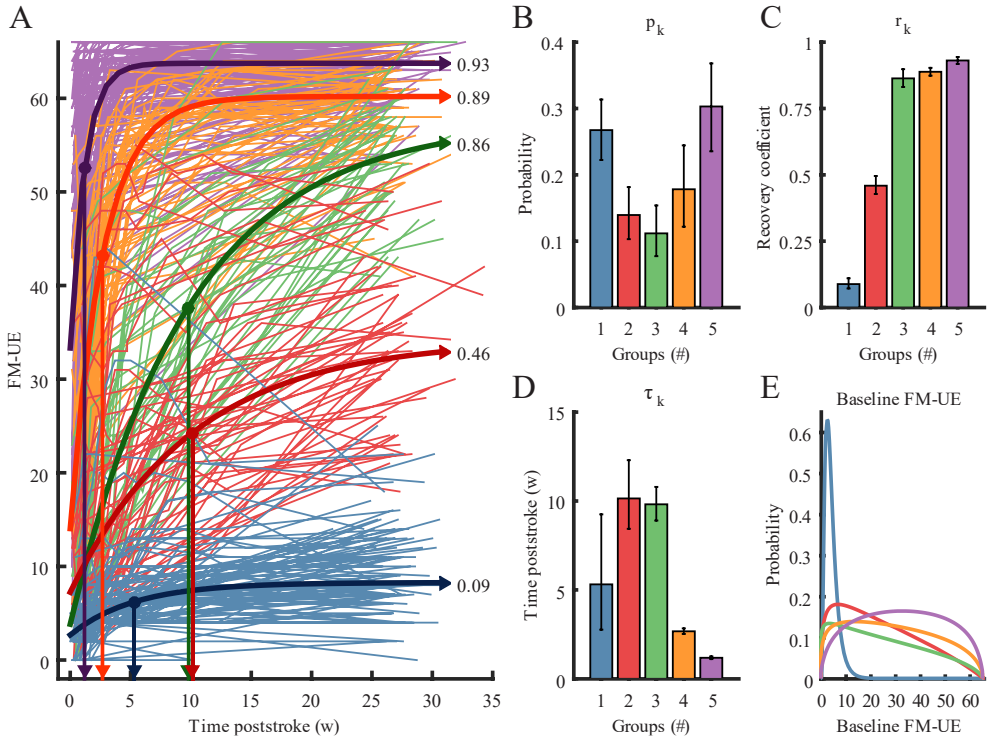


Figure 1. Longitudinal mixture model of FM-UE recovery. A. FM-UE recovery data of the 412 ischemic stroke patients in our dataset. Individual patients are color-coded according to the subgroup they were assigned to most by the longitudinal mixture model of FM-UE recovery. The average subgroup recovery patterns are shown in bold. **B-D.** Estimated model parameters for the five different subgroups: subgroup assignment probability (B), recovery coefficient (C), time constant (D), initial distribution of the FM-UE (E). Whiskers indicate 95%ETIs.

FM-UE data for typical patients with model-based predictions of FM-UE recovery and FM-UE recovery clusters are shown in Figure 3. This figure illustrates how the credible intervals of the predictions decrease as more measurements become available and how individuals can initially be misclassified in terms of their FM-UE recovery cluster. Our prediction algorithm is available through a web-based application (Shiny App) which can be accessed on: https://emcbiostatistics.shinyapps.io/LongitudinalMixtureModelFMUE/. This web-based application requires FM-UE scores and measurement dates and outputs predicted FM-UE profiles with credible intervals as well as the most likely FM-UE recovery cluster.

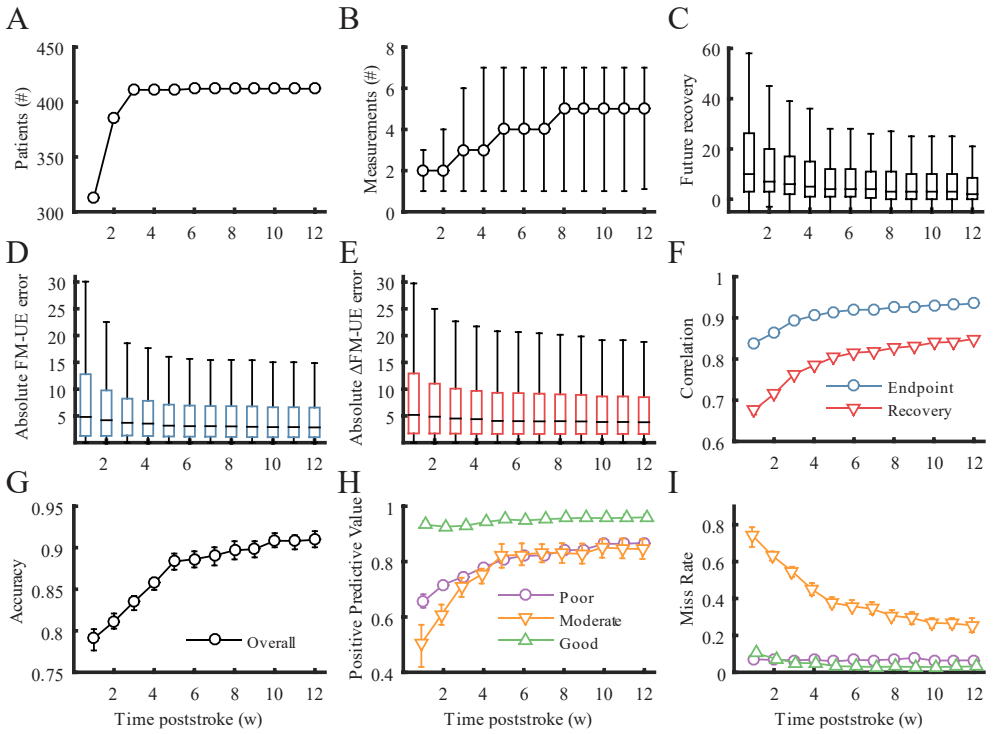


Figure 2. Cross-validation of model predictions. A. Number of patients who had at least one measurement at a specific time post stroke and were therefore included in the cross-validation. **B.** Median number of measurements per patient available for cross-validation at a specific time post stroke. Error bars indicate 95%ETI's across patients with at least one measurement. Whiskers represent 1.5 times the interquartile range; outliers not shown. **C.** Future recovery, defined as endpoint FM-UE – last available FM-UE for each patient at a specific time post stroke. **D-E.** Boxplot of the absolute error across all 412 patients times 100 samplings of the endpoint FM-UE (**D**) and the ΔFM-UE (**E**). Whiskers represent 1.5 times the interquartile range; outliers not shown. **F.** Correlation between predicted and observed FM-UE (blue circles) and ΔFM-UE (red triangles) with error bars indicating the 95%ETIs over the 100 samplings. **G-I.** FM-UE recovery cluster assignment accuracy (**G**), positive predictive value (**H**) and miss rate (**I**) with error bars indicating the 95%ETIs across the 100 samplings.

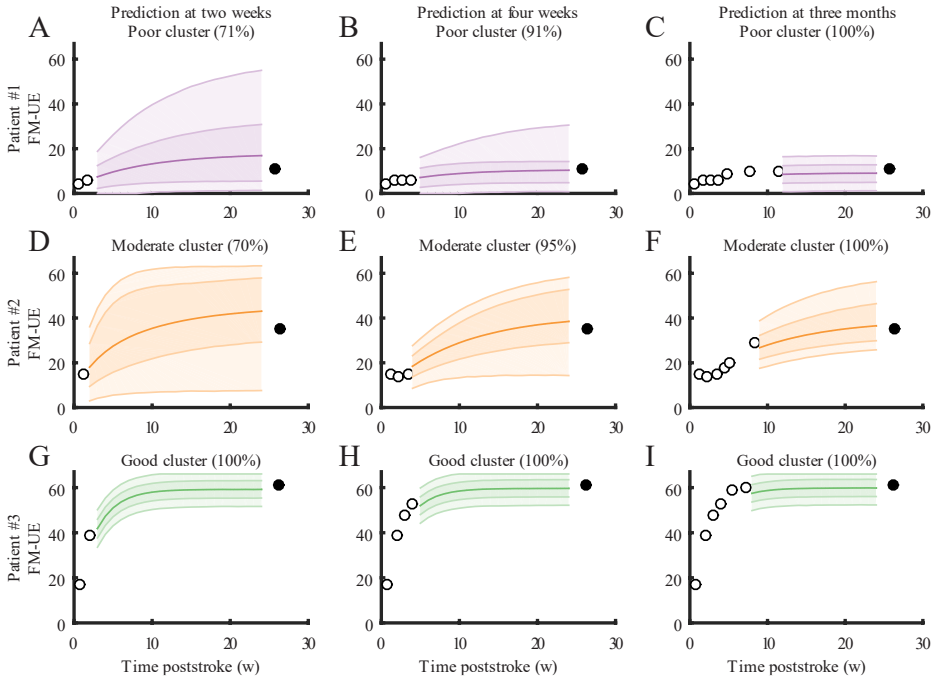


Figure 3. Model FM-UE predictions for three typical patients. Model FM-UE predictions for example patients from the optimal (given all FM-UE data) poor (**A-C**), moderate (**D-F**) or good (**G-I**) FM-UE recovery cluster. The left column illustrates predictions made using data available at two weeks post stroke, the second column at four weeks post stroke and the final column at three months post stroke. Open circles represent data used for prediction modeling. Filled markers indicate the actual endpoint FM-UE. The prediction is shown as the mean profile (dark line) with 68% credible intervals (dark shaded area) and 95% credible intervals (light shaded area). The figure titles and the colors of the credible intervals (poor (purple), moderate (orange) or good (green)) indicate the predicted FM-UE clusters as well as the probability of cluster assignment.

Discussion

We have developed a longitudinal mixture model of FM-UE recovery which describes the time course of FM-UE recovery after a first-ever ischemic stroke and does not suffer from mathematical coupling.[19,20] Based on this model, we analyzed a large FM-UE dataset of 412 first-ever ischemic stroke patients collected in prospective cohorts. Subsequently, we identified five subgroups, which we organized in three clinically relevant clusters of poor, moderate, and good FM-UE recovery. Based on a cross-validation, our paper provides first-ever estimates of predictability of endpoint FM-UE between 3 and 6 months poststroke, as well as subgroup assignment as a function of time poststroke. These results contribute to the understanding of recovery patterns in the first six months after stroke.

Our current longitudinal mixture model of FM-UE recovery, as opposed to the proportional recovery model, cannot be confounded by mathematical coupling. Hope et al. showed that the correlations between baseline FM-UE score (distribution X) and the amount of recovery defined as endpoint FM-UE – baseline FM-UE (distribution Y-X) found in proportional recovery research could be inflated by mathematical coupling.[19] However, since mathematical coupling applies to correlations of data points (baseline and endpoint FM-UE) and not to models

of longitudinal data, the recovery coefficients in our paper represent non-confounded measures of recovery as a proportion of potential recovery. In addition, mathematical coupling does not apply to the outcomes of the cross-validation as we report correlations between the model predictions and the observed values for endpoint FM-UE and ΔFM-UE rather than correlations of the form X and Y-X.

In contrast to studies relying on the proportional recovery rule, which have identified two subgroups of recoverers (fitters) and non-recoverers (non-fitters),[1,2,15–17,36] our model distinguishes five subgroups, differing in the amount and rate of recovery as well as the distribution of the FM-UE early after stroke. Patients in subgroup one, containing approximately 30% of patients, have a low baseline FM-UE and a small recovery coefficient resulting in a poor outcome. These patients seem to overlap with the non-recovers from the proportional recovery rule. Subgroups two to five refine the recovers in a more granular pattern. The majority of the recoverers (subgroups 4 and 5) regain close to 0.9 of their potential recovery in the first weeks after stroke, which is on the high end of previous estimates 0.55-0.85.[1,2,15–17,36], whereas the recoverers in subgroups 2 (0.45) and 3 (0.86) also regain a fair amount of their potential recovery but over a much longer time frame. Since previous studies have identified disruption of the corticospinal tract as the essential difference between recoverers and non-recoverers,[15–18] we expect a similar contrast between patients from subgroup one and patients from subgroups two to five. Indeed, the baseline Bamford classification shows a strikingly higher percentage of total anterior circulation infarctions (TACI) in subgroup one compared with the other four subgroups. Further definition of the structural and possibly also the genetic characteristics of the five subgroups might lead to a better understanding of FM-UE recovery.

Our study provides first-ever cross-validated estimates of individual endpoint FM-UE and ΔFM-UE prediction errors. Theoretically, it is possible to predict endpoint FM-UE at baseline using the proportional recovery rule as well. One approach could be to first identify recoverers and non-recoverers using measurements of corticospinal intactness (TMS[16] and DTI[15]) and then estimate endpoint FM-UE for the recoverers as the baseline FM-UE plus a proportional recovery term and for the non-recovers as just the baseline FM-UE. However, even though TMS[37,38] and DTI[17,39] have been validated as markers of recovers and non-recoverers, the absolute error of predicted FM-UE or ΔFM-UE scores for a population of first-ever ischemic stroke patients based on this combined approach has never been cross-validated. We found the median absolute error of endpoint FM-UE to be 4.8 at the first week poststroke and 4.2 at the second week poststroke, which is at the low end of what is deemed to be a clinically important difference (4.25 to 7.25)[40]. Therefore, our model can provide a satisfactory prognosis to patients as early as one week poststroke. In the future, further reduction in prediction errors may be achieved by adding (1) dynamical covariates (covariates that also change over time) such as the NIHSS and (2) biomarkers of corticospinal integrity (e.g., TMS[16] or DTI[15]) to improve the accuracy of subgroup assignment early after stroke. Interested researchers can apply our model to predict FM-UE recovery and the FM-UE recovery cluster by accessing a web-based application at https://emcbiostatistics.shinyapps.io/LongitudinalMixtureModelFMUE/. This application requires one or multiple FM-UE measurements (dates and scores) from a single patient to predict upper limb recovery within the first six months. Predictions are presented as the expected recovery with 68% and 95% credibility intervals to express uncertainly.

Currently, we do not yet recommend clinicians to implement our model in clinical practice nor to provide FM-UE recovery predictions based on our model. First, future studies should externally validate the model with different stroke rehabilitation datasets. Outcome of

these studies might also be that the precision of the model needs to be increased (using some of the recommendations listed above) before clinical implementation is realistic. Second, guidelines need to be developed on responsible communication of stroke recovery prognoses to patients and health care professionals, with special emphasis on the uncertainly of the model predictions. Finally, it is necessary to investigate whether knowledge of the FM-UE prognosis actually improves rehabilitation efficiency or outcome.

Based on the five subgroups identified by the model, we defined poor, moderate and good FM-UE recovery clusters (similar to the lower, middle and upper band groups identified in the classic descriptive cohort study of Garraway and colleagues almost 40 years ago[41]). These clusters could, in the future, be relevant for personalizing therapeutic interventions as well as supporting decisions on the discharge policy after admission to acute and subacute stroke units. For example, patients in the poor FM-UE recovery cluster (subgroup one) will show very limited motor recovery and might, therefore, benefit from learning compensation strategies[7] or early started neuropharmacological interventions[42] aimed at promoting neural repair.[9] In contrast, patients in the moderate FM-UE recovery cluster (subgroups two and three) recovery reasonably well over an extended period, and might benefit from early started intensive therapeutic interventions aimed at behavioral restitution.[7] Patients in the good FM-UE recovery cluster (subgroups four and five) are expected to require support in regaining advanced skills such as writing.[7]

In a research setting, the present model can be used to select patients for interventions designed for a specific FM-UE recovery cluster (e.g., interventions designed specifically for the moderate recovery cluster). Patient selection can be achieved by predicting the cluster for a patient based on the patient's early FM-UE score(s) using the web-based application of our model (see link above). The efficiency of this approach depends critically on the positive predictive value and the miss rate of cluster assignment. Positive predictive value in this context indicates the proportion of patients from a predicted cluster who have been assigned to their optimal cluster and will therefore receive the personalized intervention specifically designed for their cluster. In the current model, the positive predictive value is high for the good FM-UE recovery cluster, fair for the poor cluster and relatively low for the moderate cluster. We therefore expect that an intervention designed for good recovers will be regularly offered to the right patients, while an intervention designed for moderate recovers will be regularly offered to the wrong patients. The miss rate is the proportion of incorrectly assigned patients from an optimal cluster who therefore receive a personalized intervention designed for another cluster. We found that the miss rate is low for the good and poor cluster, and therefore expect that patients from these clusters will often get the intervention designed for their cluster, and high for the moderate cluster, and therefore expect that patients from this cluster will often an intervention designed for another cluster. Identification of patients in the poor and moderate FM-UE recovery cluster might benefit from additional repeated FM-UE measurements over time. Of particular interest would be to design a decision algorithm which identifies patients in whom the cluster prediction is uncertain, and advises on specific measurements to achieve sufficient accuracy. An additional option to increase assignment accuracy would be to incorporate additional clinical markers as explained above.

Another future application of the longitudinal mixture model of FM-UE recovery could be to detect intervention effects in recovery and rehabilitation trials with more statistical power.[14] To estimate an intervention effect, the model would need to be amended with an additional term to capture differences in the extent or possibly also the rate of recovery. This amended model could be fitted to serially collected clinical data to establish the added value of

an innovative therapeutic intervention above usual care either for the entire study population or for the three FM-UE recovery clusters separately. Given that all serial measurements are analyzed, rather than just the baseline and endpoint FM-UE, as is true for stroke recovery and rehabilitation trials, this approach could significantly promote study power. Studies investigating therapies specifically designed for either poor, moderate, or good stroke recoverers could additionally use the model's predicted FM-UE recovery cluster early after stroke to select patients, as explained above. This way, the proportion of patients from a certain FM-UE recovery cluster and the power to detect an intervention effect in that FM-UE recovery cluster will both increase, with the positive predictive value determining study homogeneity and the miss rate the study inclusion efficiency.[43,44] Further quantification of these approaches will be one of the main targets of our future work.

Limitations of the present study include (1) the lack of severely affected patients with a hemiparesis in the dominant hand, (2) the restricted generalization to patients with upper limb motor impairment after a first-ever ischemic stroke, (3) the focus on stroke recovery rather than deterioration. The language center is localized in the left hemisphere for most left-handed and right-handed people.[45] Therefore, severely affected patients with left hemisphere lesions often have language impairments that hinder providing informed consent and therefore participating in a clinical study. This explains the low percentage of patients with a hemiparesis on the dominant side in subgroup one (severely affected patients). In addition, we cannot conclude whether hemorrhagic stroke patients have similar recovery patterns, or investigate how spontaneous neurological recovery is affected by recurrent stroke. Finally, our model is not equipped to predict FM-UE deterioration after stroke. As recently emphasized, the next step is to start an international collaboration for building datasets large enough to address these questions and move recovery and rehabilitation studies forward.[43] These databases could also be used to model recovery of lower limb impairment[3,4] as well as other non-motor modalities such as speech[6] and visuo-spatial neglect[5] after stroke.[21]

References

1. Prabhakaran, S. et al. *Neurorehabil. Neural Repair* **22**, 64–71 (2008).
2. Winters, C., van Wegen, E.E.H., Daffertshofer, A. & Kwakkel, G. *Neurorehabil. Neural Repair* **29**, 614–622 (2015).
3. Smith, M.-C., Byblow, W.D., Barber, P.A. & Stinear, C.M. *Stroke* **48**, 1400–1403 (2017).
4. Veerbeek, J.M., Winters, C., van Wegen, E.E.H. & Kwakkel, G. *PLoS One* **13**, e0189279 (2018).
5. Nijboer, T.C.W., Kollen, B.J. & Kwakkel, G. *Cortex* **49**, 2021–2027 (2013).
6. Lazar, R.M., Speizer, A.E., Festa, J.R., Krakauer, J.W. & Marshall, R.S. *J. Neurol. Neurosurg. Psychiatry* **79**, 530–534 (2008).
7. Kwakkel, G., Kollen, B. & Lindeman, E. *Restor. Neurol. Neurosci.* **22**, 281–99 (2004).
8. Gresham, G.E. *Stroke* **17**, 358–60
9. Ward, N.S. *Nat. Rev. Neurol.* **13**, 244–255 (2017).
10. Boyd, L.A. et al. *Int. J. Stroke* **12**, 480–493 (2017).
11. Bernhardt, J., Godecke, E., Johnson, L. & Langhorne, P. *Curr. Opin. Neurol.* **30**, 48–54 (2017).
12. Byblow, W., Schlaug, G. & Wittenberg, G. *Ann. Neurol.* **80**, 339–341 (2016).
13. Gladstone, D.J., Danells, C.J. & Black, S.E. *Neurorehabil. Neural Repair* **16**, 232–240 (2002).
14. Krakauer, J. & Marshall, R. *Ann. Neurol.* **78**, 845–847 (2015).
15. Feng, W. et al. *Ann. Neurol.* **78**, 860–70 (2015).
16. Byblow, W.D., Stinear, C.M., Barber, P.A., Petoe, M.A. & Ackerley, S.J. *Ann. Neurol.* **78**, 848–859 (2015).
17. Buch, E.R. et al. *Neurology* **86**, 1924–1925 (2016).
18. Marshall, R.S. et al. *Ann. Neurol.* **65**, 596–602 (2009).

19. Hope, T.M.H. et al. *Brain* 306514 (2018).doi:10.1093/brain/awy302
20. Hawe, R.L., Scott, S.H. & Dukelow, S.P. *Stroke* **50**, 204–211 (2019).
21. Kundert, R., Goldsmith, J., Veerbeek, J.M., Krakauer, J.W. & Luft, A.R. *Neurorehabil. Neural Repair* 154596831987299 (2019).doi:10.1177/1545968319872996
22. Kwakkel, G. et al. *Neurorehabil. Neural Repair* **30**, 804–816 (2016).
23. Nijland, R.H.M.M., van Wegen, E.E.H.H., Harmeling-van der Wel, B.C. & Kwakkel, G. *Stroke.* **41**, 745–50 (2010).
24. Kwakkel, G. (2013).at <http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4221>
25. Duncan, P.W. et al. *Stroke* **36**, e100-43 (2005).
26. Quinn, T. et al. *J. Rehabil. Med.* **41**, 99–111 (2009).
27. Bamford, J., Sandercock, P., Dennis, M., Burn, J. & Warlow, C. *Lancet (London, England)* **337**, 1521–6 (1991).
28. Meyer, B.C., Hemmen, T.M., Jackson, C.M. & Lyden, P.D. *Stroke* **33**, 1261–6 (2002).
29. W Bohannon, R. *J. Phys. Ther. Sci.* **11**, 59–61 (1999).
30. Collin, C. & Wade, D. *J. Neurol. Neurosurg. Psychiatry* **53**, 576–9 (1990).
31. Sanford, J., Moreland, J., Swanson, L.R., Stratford, P.W. & Gowland, C. *Phys. Ther.* **73**, 447–54 (1993).
32. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D.B. (Chapman and Hall/CRC: 2013).
33. Malsiner-Walli, G., Frühwirth-Schnatter, S. & Grün, B. *Stat. Comput.* **26**, 303–324 (2016).
34. Rousseau, J. & Mengersen, K. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **73**, 689–710 (2011).
35. Nasserinejad, K., van Rosmalen, J., de Kort, W. & Lesaffre, E. *PLoS One* **12**, e0168838 (2017).
36. Zarahn, E. et al. *Cereb. Cortex* **21**, 2712–21 (2011).
37. Stinear, C.M. et al. *Stroke* **48**, 795–798 (2017).
38. Schambra, H.M. et al. *Neurorehabil. Neural Repair* **33**, 568–580 (2019).
39. Guggisberg, A.G., Nicolo, P., Cohen, L.G., Schnider, A. & Buch, E.R. *Neurorehabil. Neural Repair* **31**, 1029–1041 (2017).
40. Page, S.J., Fulk, G.D. & Boyne, P. *Phys. Ther.* **92**, 791–8 (2012).
41. Garraway, W.M., Akhtar, A.J., Smith, D.L. & Smith, M.E. *J. Epidemiol. Community Health* **35**, 39–44 (1981).
42. Pekna, M., Pekny, M. & Nilsson, M. *Stroke* **43**, 2819–2828 (2012).
43. Bernhardt, J. et al. *Int. J. Stroke* **11**, 454–458 (2016).
44. Winters, C., Heymans, M.W., van Wegen, E.E.H. & Kwakkel, G. *Trials* **17**, 468 (2016).
45. Knecht, S. et al. *Brain* **123**, 2512–2518 (2000).

3.2 Improving statistical power of subacute upper limb motor rehabilitation trials

Rick van der Vliet, Gert Kwakkel, Eleni-Rosalina Andrinopoulou, Rinske Nijland, Gerard M. Ribbers, Maarten A. Frens, E.E.H. van Wegen, Carel G.M. Meskers and Ruud W. Selles

Abstract

Objective: Powering trials is a key challenge considering current weak evidence for the effectiveness of interventions to improve motor recovery early after stroke. The goal of this paper was to compare the power to detect a difference on the Fugl-Meyer assessment of the upper extremity (FM-UE) between a longitudinal mixture model and a cross-sectional non-parametric (Mann-Whitney U) test, and investigate the effect of measurement time and intervention onset variability, number of repeated measurements, and patient selection based on predicted recovery.

Methods: We amended a longitudinal mixture model of stroke recovery to account for participation in a stroke rehabilitation trial. Using this amended model, we simulated FM-UE data of patients in the first six months poststroke and calculated the statistical power to detect a 4.25 point FM-UE difference.

Results: The number of patients needed to obtain 90% power dropped approximately sevenfold from 510 patients for the cross-sectional test to 70 patients for the longitudinal mixture model. Between-patient variability in measurement dates and intervention onset did not influence study power. Inclusion of more repeated measurements was associated with higher study power. Patient selection based on the predicted FM-UE recovery cluster increases the power to detect an intervention effect in that specific cluster, at the cost of falsely excluding patients when the predicted cluster is wrong.

Interpretation: A longitudinal mixture model can significantly reduce the minimum sample size of stroke rehabilitation trials and will, therefore, be useful for designing future stroke rehabilitation trials and for re-analyzing already completed stroke rehabilitation trials.

Introduction

The evidence for effectiveness of therapeutic interventions aimed at enhancing motor recovery early poststroke is still weak.[1–3]A major reason is that 97% of all trials in stroke rehabilitation are underpowered to detect effect sizes of 5 to 10%.[1] For example, around 300 patients with subacute stroke are required to obtain 80% power for an intervention effect of 6.6 points on the Fugl Meyer Upper Extremity score (FM-UE).[4] However, since (1) spontaneous biological recovery is the main driver of improvement on the FM-UE in the subacute phase after stroke,[5] and (2) spontaneous biological recovery is known to vary from poor to good between groups of stroke patients,[6–9] the cross-sectional approach of comparing endpoint FM-UE or FM-UE recovery from baseline to endpoint (ΔFM-UE) between intervention and usual care groups may not be optimal. Therefore, it could be more appropriate to investigate (1) longitudinal modeling approaches, which better capture non-linear upper extremity recovery after stroke[8] and (2) selection techniques for creating patient populations with similar recovery potential, as also recently suggested by the Stroke Recovery and Rehabilitation Roundtable group.[4,10–12]

Recently, we have developed a longitudinal mixture model of FM-UE recovery,[9] based on the principles of the proportional recovery rule,[6–8] which could be used both for estimating intervention effects on FM-UE recovery as well as selecting patients based on their predicted FM-UE recovery cluster (poor, moderate or good recovery). This model describes the course of FM-UE recovery in the first six months after a first-ever ischemic stroke based on (1) a proportional relation between observed recovery over time and potential recovery at baseline (longitudinal), and (2) the existence of clinically distinct subgroups of FM-UE recovery (mixture). Both concepts were implemented in an exponential recovery function with subgroup-specific model parameters for the amount and rate of recovery, and the distribution of the baseline FM-UE. In a FM-UE dataset of 412 first-ever ischemic stroke patients, five subgroups were identified with distinct recovery profiles. The five subgroups were grouped in three FM-UE recovery clusters of poor (~30% of patients, almost no recovery), moderate (~25% of patients, moderate recovery over the first three months) and good recovery (~45% of patients, almost full recovery in the first month). With a residual error as low as 3.9 points on the FM-UE, this model may improve the precision of effect size estimates. In addition, cluster assignment accuracy of the poor, moderate and good clusters was as high as 0.79 at one week poststroke, implying that model-based selection of patients could help select stroke patients in future stroke rehabilitation trials to further promote study power.[5,13]

The goals of this paper are to study the effects on study power of (1) FM-UE analysis with either the longitudinal mixture model or a cross-sectional non-parametric (Mann-Whitney U) test (2) between-patient variability in measurement dates and intervention onset, (3) the number of repeated measurements, and (4) patient selection based on the predicted FM-UE recovery cluster (poor, moderate or good) for estimating cluster-specific intervention effects. To this end, we amended the longitudinal mixture model of FM-UE recovery to account for participation in a stroke rehabilitation trial and simulated FM-UE data of patients in the first six months poststroke. Using simulated patient data, we calculated the power to detect the lower limit for a clinically important difference on the FM-UE (4.25 points).[14]

Materials and methods

We first amended our previously-reported longitudinal mixture model of FM-UE recovery[9] (see *Longitudinal mixture model of FM-UE recovery*) for application in the design and statistical analysis of a stroke rehabilitation trial (see: *Amended longitudinal mixture model of FM-UE*

recovery). Using this amended model, we simulated 1000 longitudinal datasets of patients participating in a stroke rehabilitation trial by assuming values for, amongst others, the size of the intervention effect, the number of repeated measurements, the timing of measurements poststroke, and the number of patients (see: *Patient data simulation*). These datasets were used to calculate the power of different simulated study designs (see Table 1) based on (1) a cross-sectional nonparametric (Mann-Whitney U) test, (2) a longitudinal model analysis without patient selection or (3) a longitudinal model analysis with patient selection based on the predicted FM-UE recovery cluster (poor, moderate or good) (see: *Power analyses*). These different steps are outlined in more detail below.

Longitudinal mixture model of FM-UE recovery

Details of the longitudinal mixture model of FM-UE recovery have been reported extensively.[9] In brief, we included FM-UE data from 412 first-ever ischemic stroke patients collected in four different prospective cohort studies: the EXPLICIT, [15] EPOS,[16] 4D-EEG[17] and EXPLORE studies. We developed a longitudinal model of FM-UE recovery after stroke, based on the principles of proportional recovery, which are (1) a proportional relation between observed recovery over time and potential recovery at baseline and (2) the presence of distinct subgroups of FM-UE recovery (mixture).[6–8] Since FM-UE recovery follows an exponential pattern,[5] we chose an exponential function as the time-dependent element of the model, with the asymptote defined as a proportion of the potential recovery and the time constant expressed in weeks poststroke. In addition, we included an intercept which represents the baseline FM-UE immediately after stroke:

$$\mu_{ij|k} = \alpha_{i|k} + r_k * \left(66 - \alpha_{i,k}\right) * \left(1 - e^{-\frac{t_{ij}}{\tau_k}}\right) \quad (1)$$

$$y_{ij|k} \sim N\left(\mu_{ij|k}, \sigma_\epsilon^2\right) \quad (2)$$

$$\alpha_{i|k} = \frac{66}{1 + exp\left(-N\left(\mu_{\alpha,k}, \sigma_{\alpha,k}^2\right)\right)} \quad (3)$$

$$k \sim Cat(K, p_k) \quad (4)$$

With i the patient identification number $[1 - I]$, j the measurement identification number $[1 - J]$, k the subgroup identification number $[1 - K]$ and t_{ij} the measurement date in weeks poststroke. The subgroup-specific model parameters are the recovery coefficient r_k (reflecting the amount of recovery), the time constant τ_k (reflecting the rate of recovery), the mean and standard deviation of the logistic normal distribution ($\mu_{\alpha,k}$ and $\sigma_{\alpha,k}^2$) of the patient-specific baseline FM-UE $\alpha_{i|k}$, and the subgroup assignment probability p_k. σ_ϵ^2 is the standard deviation of the residual error. As reported earlier,[9] we identified five distinct subgroups with different recovery parameters ($r_1 = 0.09$, $\tau_1 = 5.3$; $r_2 = 0.46$, $\tau_2 = 10.1$; $r_3 = 0.86$, $\tau_3 = 9.8$; $r_4 = 0.89$, $\tau_4 = 2.7$; $r_5 = 0.93$, $\tau_5 = 1.2$), with a residual error standard deviation of 3.9 points on the FM-UE (95%ETI = [3.7 4.0]). We defined poor (subgroup 1), moderate (subgroups 2 and 3) and good (subgroups 4 and 5) FM-UE recovery clusters for further analysis.

Amended longitudinal mixture model of FM-UE recovery

We amended the longitudinal mixture model of FM-UE recovery with a study-specific exponential function to estimate study group effects on the amount of recovery θ_{g_i} for patients in the usual care group θ_1 (covering all differences between the usual care group and the original population, including a placebo effect) and for patients in the intervention group θ_2 (covering all differences between the intervention group and the original population, including a placebo and an intervention effect). The intervention effect $\Delta\theta$ is calculated by subtracting the usual care group effect θ_1 from the intervention group effect θ_2. For the intervention group effect θ_2, we either calculated an estimate for all three clusters combined ($\theta_{2,All}$) or for the three clusters separately (θ_{2,C_k}): $\theta_{2,Poor}$, $\theta_{2,Moderate}$ and $\theta_{2,Good}$. For the time constants of the study-specific exponential function, we used the same time constants τ_k describing the spontaneous recovery, which assumes that the study group effects manifest with different rates for the different subgroups. Sensitivity analyses were performed to test the impact of this assumption (see *Patient data simulation*). In addition, we corrected for the intervention start date poststroke by shifting the study-specific exponential function with a patient-specific (i) intervention start date in weeks Δt_i: $\left(t_{ij} - \Delta t_i, 0\right)$. The mathematical expression for the mean of the model $\mu_{ij|k}$ becomes:

$$\mu_{ij|k} = \alpha_{i|k} + r_k * \left(66 - \alpha_{i|k}\right) * \left(1 - e^{-\frac{t_{ij}}{\tau_k}}\right) + \theta_{g_i} * \left(1 - e^{-\frac{(t_{ij} - \Delta t_i, 0)}{\tau_k}}\right) \quad (5)$$

With g_i the study group [1 2] for patient i and θ_{g_i} the study group effect. Furthermore, we modeled $y_{ij|k}$ with a t-distribution rather than a normal distribution to increase the robustness of the analysis in small datasets:

$$y_{ij|k} \sim T\left(\mu_{ij|k}, \sigma_\epsilon^2, df\right) \quad (6)$$

With df the degrees of freedom.

Patient data simulation

We generated 1000 simulated datasets, each containing 1000 patients with similar characteristics as our initial dataset of 412 first-ever ischemic stroke patients. To do so, we inserted the means of the model parameters (r_k, τ_k, $\mu_{\alpha,k}$, $\sigma_{\alpha,k}$, σ_ϵ) estimated on the original cohort[9] in equations 3-6, with the degrees of freedom k set to 10[3]. We limited the analysis to patients with a clinically-relevant paresis just after stroke, by restricting the baseline FM-UE $\alpha_{i|k}$ to scores below 53,[18] and adjusted the subgroup assignment probabilities to account for this cut-off value. The adjusted subgroup assignment probabilities were: $p_1 = 0.31, p_2 = 0.15, p_3 = 0.12$, $p_4 = 0.17$ and $p_5 = 0.25$. Furthermore, we limited $\mu_{ij|k}$ and $y_{ij|k}$ to the range of the FM-UE (0-66).

The overall ($\Delta\theta_{All}$) and cluster-specific intervention effects ($\Delta\theta_{Poor}$, $\Delta\theta_{Moderate}$ and $\Delta\theta_{Good}$) were set to 4.25, which is the lower limit for a clinically important difference on the FM-UE.[14] Furthermore, the intervention start date Δt_i was set to 3 weeks poststroke and the measurement dates t_{ij} to 1, 2, 4, 12 and 26 weeks poststroke. For some study designs, we added between-patient variability to the intervention onset Δt_i and the measurement dates t_{ij}, by randomly shifting the dates maximally one week forward or backward with uniform probability: $U(-1,1)$. Finally, we performed sensitivity analyses for the assumptions on (1) the recovery rate of the intervention effect by setting the time constants used in data simulation to fixed values

(two weeks or ten weeks) rather than subgroup-specific values and (2) the residual error of the model σ_ϵ by simulating data with 25% more residual error than found in the original dataset. Details of the parameters used for each study design are presented in Table 1.

For the longitudinal model analyses, we randomly included 10 to 140 patients (half from the usual care group and half from the intervention group) with increments of 10 patients from the pool of 1000 patients for each of the 1000 datasets. In the patient-selection designs, patient inclusion was limited to those patients who had been assigned to a certain FM-UE recovery cluster (poor, moderate or good) based on their FM-UE measurements at one or two weeks poststroke (see below). For the cross-sectional analysis, the maximum number of patients was set to 700 patients.

Power analyses

We compared the effects of different ways of analyzing the simulated FM-UE trial data on study

Comparison	Fig	Design	Model	Cluster specific	Patient select	Timing var. t_{ij}	Timing var. Δt_i	τ	σ_ϵ	t_{ij} 1	2	4	12	26
Reference	1-4	1	L	No	No	No	No	τ_k	3.9					
Model	1	2	C	-	-	No	No	τ_k	3.9					
Between-patient timing variability	2	3	L	No	No	Yes	No	τ_k	3.9					
		4	L	No	No	Yes	Yes	τ_k	3.9					
Repeated measurement	2	5	L	No	No	No	No	τ_k	3.9					
Patient selection	3	6	L	Yes	No	No	No	τ_k	3.9					
		7	L	Yes	Yes	No	No	τ_k	3.9					
		8	L	Yes	Yes	Yes	No	τ_k	3.9					
		9	L	Yes	Yes	Yes	No	τ_k	3.9					
Assumptions	4	10	L	No	No	No	No	2	3.9					
		11	L	No	No	No	No	10	3.9					
		12	L	No	No	No	No	2	3.9					
		13	L	No	No	No	No	10	3.9					
		14	L	No	No	No	No	τ_k	4.8					

Table 1. Simulated study designs. Details of the simulated study designs. The comparison column indicates which study designs are used to investigate the influence of (1) the model, (2) between-patient variability in measurement dates and intervention onset, (3) repeated measurements, (4) patient selection based on the predicted stroke recovery cluster, and (5) assumptions on the recovery rate of the intervention and on the model residual error on the power to detect a 4.25 point FM-UE difference. Green filling highlights the essential parameter changes for a certain comparison. For the measurement dates columns, filled boxes indicate a measurement date was included in the analysis of the intervention effect, diagonally striped boxes signal a measurement date was also used for clustering-based patient selection. With L the longitudinal mixture model, C the cross-sectional model, τ the time constant used for data stimulation, τ_k the subgroup-specific time constants and σ_ϵ the residual error standard deviation.

power. For the cross-sectional analysis, the endpoint FM-UE of the intervention group and usual care group were compared with a non-parametric, two-tailed, Mann-Whitney U-test, as the distribution of the endpoint FM-UE was non-normal in the 412 ischemic stroke patients described before[9] (Kolmogorov-Smirnov test: $D(233) = 0.229$, $p = 3.58E\text{-}18$). The cut-off for the

p-value was 0.05. We assessed the contribution of the non-parametric test by repeating the analysis with a parametric t-test and found lower study power for the parametric test (data not shown).

For the longitudinal mixture model analysis without patient selection, the amended model from equations 3-6 with the mean model parameters $r_k, \tau_k, \mu_{\alpha,k}, \sigma_{\alpha,k}$ derived from Van der Vliet et al.[9] was fitted to the study dataset. The intervention group effects θ_2 were estimated either for the entire intervention group $\theta_{2,All}$ or for the three clusters separately ($\theta_{2,Poor}$, $\theta_{2,Moderate}$, and $\theta_{2,Good}$). The usual care group effect θ_1 was always calculated in the entire usual care group.

$$\theta_1, \theta_{2,All}, \theta_{2,Poor}, \theta_{2,Moderate}, \theta_{2,Good} \sim N(0,10^3) \quad (7)$$

The standard deviation and degrees of freedom were estimated as:

$$\sigma_\epsilon \sim 1/\Gamma(10^{-3}, 10^{-3}) \quad (8)$$
$$df \sim 1/\Gamma(10^{-3}, 10^{-3}) \quad (9)$$

For the longitudinal mixture model analysis with patient selection, patients were assigned at baseline to one of the clusters (poor, moderate or good), using the longitudinal mixture model from equation (1-4) with the model parameters from van der Vliet et al.[9]. Next, the amended model (equations 3-6) was fitted to estimate cluster-specific intervention effects in the selected populations. The primary parameter was the intervention effect specific to the cluster-of-interest. For example, if patients from the poor FM-UE recovery cluster were selected, $\Delta\theta_{Poor}$ would be the primary parameter.

For all longitudinal analyses, power was calculated as the percentage of simulated trials wherein the lower border of the 95%ETI of the intervention effect $\Delta\theta$ was larger than zero. Additional outcome parameters were: (1) the number of patients needed to obtain 90% power and (2) the positive predictive value and the miss rate for the predicted cluster in the patient selection designs, presented as the mean with 95%ETIs calculated over all simulations.

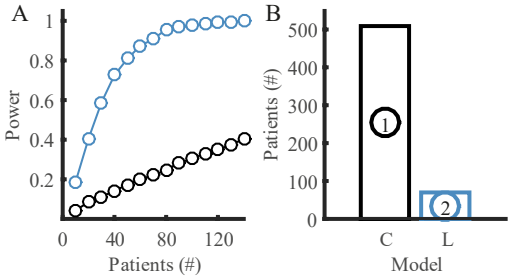


Figure 1. Comparing the longitudinal model to a cross-sectional test. A-B. Power plots (**A**) and 90% power bar graphs (**B**) for the cross-sectional model, shown in **black**, and the longitudinal model, shown in **blue**. The numbers in the plots refer to the study designs in Table 1.

Markov-chain Monte-Carlo sampling

We followed the Bayesian framework and performed all statistical analyses using Markov-chain Monte-Carlo sampling implemented in JAGS 4.3.0 (available from: https://sourceforge.net/projects/mcmc-jags/). Matlab 2015a (MathWorks, Natick, Massachusetts, United States) and Matjags (available from: http://psiexp.ss.uci.edu/research/programs_data/jags/) were used for data and sample processing. Settings for data generation were: 10^2 burn-in samples and 10^3 posterior distribution samples and one chain. Settings for estimating the intervention effect and patient clustering were: 10^3 burn-in samples and 10^3 posterior distribution samples with five chains. All scripts can be accessed at: https://github.com/rickvandervliet/Bayesian-Proportional-Recovery. In addition, we have created an online application to predict individual recovery profiles and FM-UE recovery cluster assignments in a user-friendly format at https://emcbiostatistics.shinyapps.io/LongitudinalMixtureModelFMUE/. These scripts can be used to (re-)estimate intervention effects in stroke rehabilitation trials and to perform power analyses.

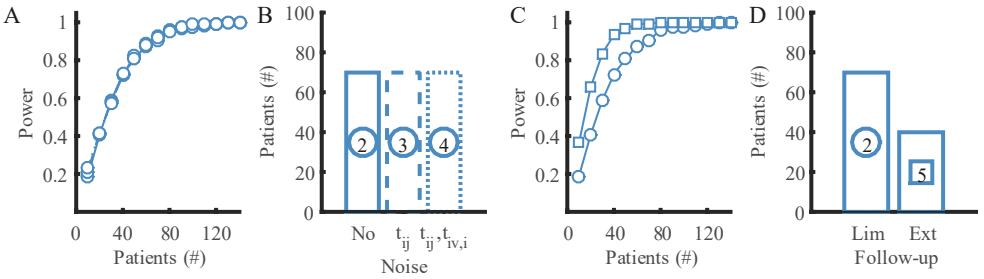


Figure 2. The effect of (1) between-patient variability in measurement dates and intervention onset and (2) follow-up on study power. A-B. Power plots (**A**) and 90% power bar graphs (**B**) for a study design without between-patient variability, illustrated with a solid line, a design with between-patient variability of measurement dates, illustrated with a striped line and a design with between-patient variability of measurement dates and intervention onset (see Table 1 design 4), illustrated with a dotted line. **C-D.** Power plots (**C**) and 90% power bar graphs (**D**) for a limited follow-up design, illustrated with a circular marker, and an extensive follow-up design, illustrated with a square marker.

Results

The longitudinal mixture model had more power than a cross-sectional non-parametric (Mann-Whitney U) test to detect an overall intervention effect (see results in Figure 1A-B for study designs 1-2 in Table 1). More specifically, the number of patients needed to obtain 90% power was reduced more than sevenfold from 510 patients for the cross-sectional non-parametric test to 70 patients for the longitudinal mixture model.

Both between-patient variability of measurement dates as well as between-patient variability of intervention onset relative to stroke did not influence study power for the longitudinal mixture model simulations (see results in Figure 2A-B for study designs 1, 3 and 4 in Table 1). In all three study designs, 70 patients were needed to obtain 90% power. Having repeated measurements at four time points (at 1, 4, 12 and 26 weeks poststroke) during the six month follow-up increased the study power compared to two repeated measurements (at three weeks and 26 weeks poststroke) and decreased the number of patients needed for 90% power with approximately 40% from 70 to 40 patients (see results in Figure 2C-D for study designs 1 and 5 in Table 1).

The power to detect an intervention effect in a specific FM-UE recovery cluster was relatively low in an unselected study population (see Figures 3A-B and Table 1 study design 6), due to the fact that the number of patients from each cluster is approximately a third of the total number of patients. We found the smallest required sample size for the poor FM-UE recovery cluster as it is not affected by ceiling. Patient selection based on the predicted FM-UE recovery cluster maximized the proportion of patients from a cluster of interest in a trial and therefore increased the power to detect a cluster-specific intervention effect (see Figures 3C-D and Table 1 study design 7 and Figure 3E for the positive predictive value of the included patients). However, patient selection early after stroke comes at the cost of falsely excluding patients when the predicted cluster is wrong, which was mainly an issue in the cluster of moderate recovery (see Figure 3F for the miss rate). Better performance for the cluster of moderate recovery can be obtained by delaying patient selection until week two (see Figures 3E-F and Table 1 study design 8). In contrast, having two measurements to predict the FM-UE recovery cluster (at week one and week two; study design 9) rather than one (at week two; study design 8) only slightly improved the miss rate (0.41 95%ETI=[0.38 0.44] versus 0.50 95%ETI=[0.47 0.53]) and did not affect the positive predictive value (0.76 95%ETI=[0.73 0.79] versus 0.77 95%ETI=[0.73 0.80]) and therefore the power (60 patients needed for 90% power).

Finally, we tested the sensitivity of the power results to violations of the model assumptions. Assuming fixed rather than subgroup-specific time constants for the study-specific exponential function only marginally affected study power. A fixed long (10 weeks) time constant slightly decreased study power whereas a fixed short (2 weeks) time constant slightly increased study power (see Figures 4A-B and Table 1 study designs 10-13). Second, increasing the residual

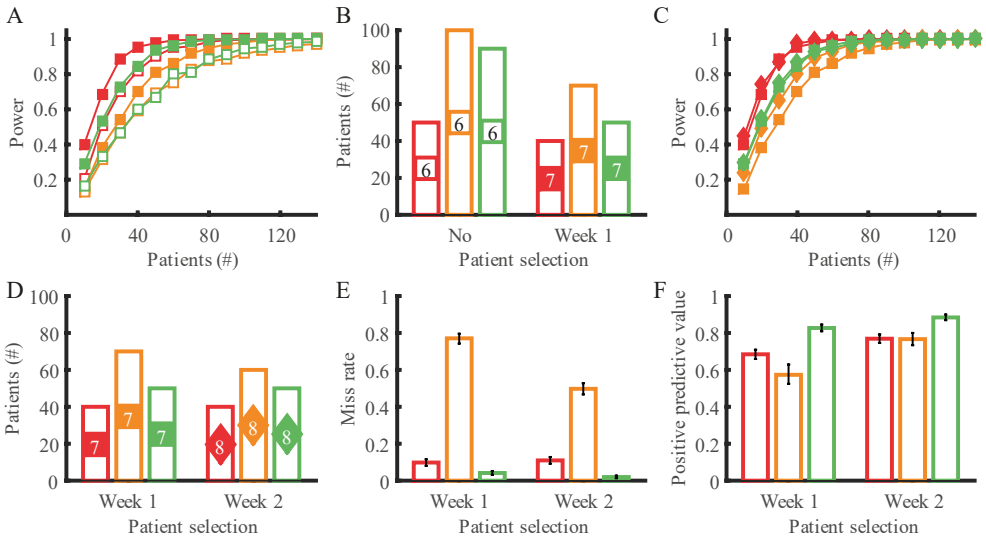


Figure 3. The effect of cluster-based patient selection on study power. A-D. Power plots (**A,C**) and 90% power bar graph (**B,D**) showing the cluster-specific analyses for the poor FM-UE recovery cluster in **red**, for the moderate FM-UE recovery cluster in **orange**, and for the good FM-UE recovery cluster in **green**. Square markers indicate an extensive follow-up with patient selection at week 1, diamond markers indicate an extensive follow-up with clustering at week 2. Open markers indicate non-selected study populations, filled markers selected study populations. **E-F.** Miss rate and positive predictive value for the predicted cluster at week one, or week two compared to the optimal cluster. Error bars represent 95% equal tailed intervals.

error of the model compared to the value found by Van der Vliet et al.,[9] which simulates a poorer model fit, decreased study power (see Figures 4E-F and Table 1 study design 14).

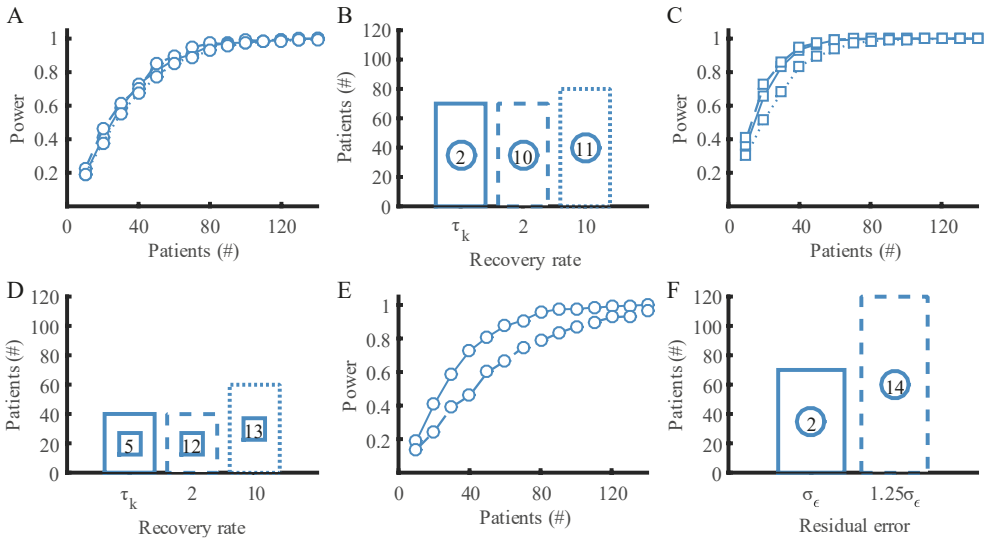


Figure 4. Testing model assumptions. A-D. Power plots (**A, C**) and 90% power bar graphs (**B, D**) for the assumptions on the rate of recovery. The solid line shows the results for data simulated with subgroup-specific time constants (as assumed), the striped line for data simulated with a fixed time constant of two weeks and the dotted line for data simulated with a fixed time constant of ten weeks. **E-F.** Power plot (**E**) and 90% power bar graph (**F**) for the assumptions on the standard deviation of the residual error. The solid line shows the results for data simulated with the model-derived residual error standard deviation, the striped line for data simulated with a 25% higher standard deviation. The circular markers indicate a limited follow-up, the square markers an extensive follow-up.

Discussion

Proper powering of trials is a key challenge considering current weak evidence for the effectiveness of interventions to improve motor recovery after stroke.[1–3] The goal of this paper was to compare the power to detect a difference on the Fugl-Meyer assessment of the upper extremity (FM-UE) between a longitudinal mixture model[9] and a cross-sectional non-parametric (Mann-Whitney U) test. We amended a longitudinal mixture model of FM-UE recovery to simulate FM-UE data of patients in the first six months poststroke and calculated the power to detect a 4.25 point FM-UE difference. We found that the number of patients needed to obtain 90% power drops approximately sevenfold from 510 patients for the cross-sectional test to 70 patients for the longitudinal mixture model. Our results will be useful for (1) designing future stroke rehabilitation trials, (2) re-analyzing already completed stroke rehabilitation trials.

The results of the cross-sectional power analysis presented in this paper are in line with earlier results.[4] Winters et al. investigated the sample size needed in a stroke rehabilitation trial to find an intervention effect 26 weeks after inclusion based on data from the EXPLICT study.[4] Using the power equations and standard deviation of the FM-UE 26 weeks after inclusion (18.47) described by Winters et al., we calculate approximately 800 patients would be required to detect a 4.25 difference on the FM-UE with a power of 0.9. Our cross-sectional analysis differs from the cross-sectional analysis by Winters et al. in that (1) more complex and realistic FM-UE data was simulated, (2) non-parametric statistics were used, and (3) comparisons were made at 26 weeks

poststroke rather than 26 weeks after inclusion. Therefore, the 510 patients found in the present study diverge from the estimates based on the work by Winters et al.[4]

The longitudinal mixture model has a much higher power to detect intervention effects than a cross-sectional non-parametric (Mann-Whitney U) test applied to the endpoint FM-UE at 26 weeks poststroke. More specifically, based on a study design with a limited number of repeated measurements (at one week and 26 weeks poststroke), without between-patient variability in timing of measurements or treatment start, we found a study sample of 70 patients to be sufficient for obtaining 90% power versus a study sample of 510 for the cross-sectional analysis. This difference can be explained by (1) the relatively small residual error on the longitudinal mixture model compared with the relatively high standard deviation of the endpoint FM-UE (23.7 in the original dataset of 412 stroke patients) and (2) incorporation of the baseline FM-UE measurement in the analysis.

Between-patient variability of measurement dates as well as between-patient variability in intervention start dates did not impact study power in the mixture model approach. Underlying this finding is the incorporation of the exact measurement dates in the longitudinal mixture model rather than measurement moments (t_0, t_1, etc.) for more classical analyses. Furthermore, as expected, we found study power to increase with the number of repeated measurements. This is because additional measurements improve individual patient estimates of baseline Fugl-Meyer $\alpha_{i|k}$ and subgroup assignment k, and therefore also enhance estimates of the intervention effects. For studies designed to identify separate treatment effects for each of the three recovery clusters (poor, moderate and good), our results indicate that patient selection based on the predicted (at one or two weeks after stroke) FM-UE recovery cluster increases power. Trials that specifically target patients in the cluster of moderate FM-UE recovery may benefit from delaying patient selection until two weeks as this sharply increases successful cluster allocation. Patient selection increases the proportion of patients from a certain cluster in the study sample, as a function of the positive predictive value and therefore increases the power to detect an intervention effect in a certain cluster.

In simulating patient data and estimating the intervention effects, we assumed that the rate of recovery of the intervention effect is identical to the subgroup-specific spontaneous recovery rate. However, it is also imaginable that the time course of an intervention effect is fixed for patients from all five subgroups. We therefore performed a sensitivity analysis by simulating patient data with fixed time constants for the intervention effect (two weeks or ten weeks), while erroneously estimating the intervention effect with subgroup-specific time constants. Results indicate that power slightly decreases with a fixed long time constant (ten weeks), and slightly increases with a fixed short time constant (two weeks). This makes sense, as the follow-up includes a measurement late (26 weeks) after stroke, which will capture treatment effects no matter the rate. Furthermore, we assumed that the model derived by Van der Vliet et al.[9] will fit equally well to new datasets, meaning the residual error will be identical. To test this second assumption, we simulated data with 25% more residual error and indeed found a decrease in study power, although compared to the cross-sectional test, the number of patients needed for 90% power was still four times smaller. Therefore, our results seem robust to model assumption challenges.

In this study, we estimated the intervention effect as an increase in the amount of recovery, although the longitudinal mixture model of FM-UE recovery could also detect other types of intervention effects. For example, an intervention could speed up the recovery rate or even promote a patient to a better FM-UE recovery cluster. If the focus of a study is on the

recovery rate, the longitudinal mixture model would need to be amended with a parameter that can modulate the time constant, either for the entire population or for the FM-UE recovery clusters separately. Shifts in FM-UE recovery cluster as a result of the intervention could be estimated directly from the proportions in the intervention and usual care groups. Calculating the statistical power of analyses focused on recovery rate and cluster assignment could be the topic of future studies. Moreover, we published all code on GitHub (see Methods) allowing other researchers to analyze the effect of potential variations in trial design (e.g., the number of follow-up measurements, start of intervention, expected intervention effect, etcetera) on study power.

A limitation of our study is the lack of a real-world stroke rehabilitation dataset analysis. Although the model we use for data simulation is based on actual cohort data, the simulated data is not directly derived from patients. Therefore, it would be interesting to verify our findings using existing stroke rehabilitation trial datasets with or without a statistically significant intervention effect. In trials with a statistically significant result, magnitude, and significance of the estimated intervention could be compared between the applied cross-sectional approach and the longitudinal model. In addition, the model could be improved by testing different hypotheses on the time course of intervention effects. Finally, the FM-UE recovery clusters which benefit the most from the intervention could be identified, which would be helpful for personalizing the intervention. In trials without a statistically significant result, our modeling approach may be able to uncover previously unrecognized clinical intervention effects, either in the population as a whole or in specific FM-UE recovery clusters.[13,15] Another limitation of our study is the generalizability of our results. The longitudinal mixture model was derived from first-ever ischemic stroke patients with a clinical (hemi)paresis, who were included based on CT or MRI imaging. Therefore, the model and power may not generalize to patients with a hemorrhagic stroke or recurrent stroke. Moving forward, application of the longitudinal mixture model to stroke recovery and stroke rehabilitation trials with broader inclusion criteria would help to generalize the results to different patient populations. Alternatively, the approach could be taken to re-estimate the model parameters on a new clinical dataset using the current model parameters as initial values, also ensuring generalization. International collaboration will help in designing increasingly better models to understand recovery after stroke as well as the way interventions impact on recovery.[10]

Results of the power calculations presented in this paper can be used to design future stroke rehabilitation trials. We have found that the longitudinal mixture model has more power to detect an intervention effect than a cross-sectional approach based on study design with a baseline (at one week) and endpoint (at 26 weeks) measurements. Therefore, relatively small (40-100 patients) studies can be powered for a clinically important difference on the Fugl-Meyer scale (>=4.25 FM-UE point) to test new innovative interventions, whereas larger studies can be powered to detect small differences (<4.25 points). In addition, we found that the exact timing of measurements and the start of treatment is not essential from a statistical point of view. Biologically, however, interventions may have effects in specific stages poststroke,[19,20] which would be a reason for having comparable intervention onsets for patients included in a study.[19–21] Furthermore, trials have to balance the effort put into collecting a large number of patients and the effort of intensive follow-up measurements. Study power can be improved both by increasing the number of patients and by increasing the number of measurements. Having more than two measurements per patient will allow for in-depth analysis of the time course of the intervention effect whereas having more patients will increase generalizability and support secondary analysis of, for example, complication rate and cluster shifts.

References

1. Veerbeek, J.M. et al. *PLoS One* **9**, e87987 (2014).
2. Pollock, A. et al. *Cochrane Database Syst. Rev.* **2014**, CD010820 (2014).
3. Langhorne, P., Bernhardt, J. & Kwakkel, G. *Lancet* **377**, 1693–1702 (2011).
4. Winters, C., Heymans, M.W., van Wegen, E.E.H. & Kwakkel, G. *Trials* **17**, 468 (2016).
5. Kwakkel, G., Kollen, B. & Lindeman, E. *Restor. Neurol. Neurosci.* **22**, 281–99 (2004).
6. Prabhakaran, S. et al. *Neurorehabil. Neural Repair* **22**, 64–71 (2008).
7. Winters, C., van Wegen, E.E.H., Daffertshofer, A. & Kwakkel, G. *Neurorehabil. Neural Repair* **29**, 614–622 (2015).
8. Krakauer, J. & Marshall, R. *Ann. Neurol.* **78**, 845–847 (2015).
9. Vliet, R. et al. *Ann. Neurol.* ana.25679 (2020).doi:10.1002/ana.25679
10. Bernhardt, J. et al. *Int. J. Stroke* **11**, 454–458 (2016).
11. Ward, N.S. *Nat. Rev. Neurol.* **13**, 244–255 (2017).
12. Boyd, L.A. et al. *Int. J. Stroke* **12**, 480–493 (2017).
13. Kwakkel, G., Wagenaar, R.C., Twisk, J.W., Lankhorst, G.J. & Koetsier, J.C. *Lancet* **354**, 191–196 (1999).
14. Page, S.J., Fulk, G.D. & Boyne, P. *Phys. Ther.* **92**, 791–8 (2012).
15. Kwakkel, G. et al. *Neurorehabil. Neural Repair* **30**, 804–816 (2016).
16. Nijland, R.H.M.M., van Wegen, E.E.H.H., Harmeling-van der Wel, B.C. & Kwakkel, G. *Stroke.* **41**, 745–50 (2010).
17. Kwakkel, G. (2013).at <http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4221>
18. Hoonhorst, M.H. et al. *Arch. Phys. Med. Rehabil.* **96**, 1845–1849 (2015).
19. Zeiler, S.R. & Krakauer, J.W. *Curr. Opin. Neurol.* **26**, 609–616 (2013).
20. Murphy, T.H. & Corbett, D. *Nat. Rev. Neurosci.* **10**, 861–872 (2009).
21. Biernaskie, J., Chernenko, G. & Corbett, D. *J. Neurosci.* **24**, 1245–1254 (2004).

Chapter 4. Electrophysiology, genetics and neuromodulation

4.1 TMS motor mapping: comparing the absolute reliability of digital reconstruction methods to the golden standard.

Zeb D. Jonker, Rick van der Vliet, Christopher M. Hauwert, Carolin Gaiser, Joke H.M. Tulen, Jos N. van der Geest, Opher Donchin, Gerard M. Ribbers, Maarten A. Frens and Ruud W. Selles

Abstract

Background: Changes in transcranial magnetic stimulation motor map parameters can be used to quantify plasticity in the human motor cortex. The golden standard uses a counting analysis of motor evoked potentials (MEPs) acquired with a predefined grid. Recently, digital reconstruction methods have been proposed, allowing MEPs to be acquired with a faster pseudorandom procedure. However, the reliability of these reconstruction methods has never been compared to the golden standard.

Objective: To compare the absolute reliability of the reconstruction methods with the golden standard.

Methods: In 21 healthy subjects, both grid and pseudorandom acquisition were performed twice on the first day and once on the second day. The standard error of measurement was calculated for the counting analysis and the digital reconstructions.

Results: The standard error of measurement was at least equal using digital reconstructions.

Conclusion: Pseudorandom acquisition and digital reconstruction can be used in intervention studies without sacrificing reliability.

Introduction

Transcranial magnetic stimulation (TMS) can be used to measure plasticity in the human primary motor cortex by comparing the location, size and excitability of cortical muscle representations before and after an intervention [1,2]. In the golden standard procedure, data is acquired by measuring electromyography (EMG) while applying multiple stimuli at predefined grid points on the scalp, which is then analyzed by counting the grid points at which more than half of the stimuli produced a motor evoked potential (MEP) [1–4]

Recently, digital analysis methods have been proposed for reconstructing the muscle representation from scattered stimuli, most notably surface fitting [5], cubic spline interpolation [6] and Voronoi tessellation [6]. Cavaleri et al. [7] showed that the surface fitting analysis produces similar results with data acquired in a grid procedure, which takes 15 to 60 minutes [4,5], as with data acquired in a pseudorandom walk procedure, which takes less than 5 minutes [5,7]. Therefore, these reconstruction methods could improve the ability to measure short-term plasticity [5].

To replace the counting analysis, digital reconstruction methods should show at least equal absolute reliability (e.g. standard error of measurement, SEM), as this is a marker of sensitivity to change in an individual or group [8]. However, this analysis has not yet been done. Therefore, the primary goal of this study was to compare the absolute reliability of the motor map parameters (volume, area, center of gravity) of the digital reconstruction methods to the golden standard. The results can be used as reference values for power calculations of future intervention studies.

Materials and methods

Twenty-one healthy right-handed subjects were recruited for this study (age: 28±9 years, 12 females). Participants were screened for contraindications using the TMS adult safety questionnaire [9]. The experiment was approved by the Medical Ethical Committee of the Erasmus MC Rotterdam and performed in accordance with the Declaration of Helsinki.

Setup

A Visor2 XT system (ANT Neuro, The Netherlands) was used, consisting of a MagPro X100 stimulator, a MC-B70 coil (Magventure, Denmark), a custom-built amplifier (TMSi, The Netherlands), a Polaris Spectra motion tracking system (NDI, Canada) and Visor 2 software (ANT Neuro, The Netherlands). Electromyography (EMG) signals were recorded from the left first dorsal interosseous (FDI) muscle with silver-silverchloride electrodes in a belly-tendon montage, sampled at 5 kHz and stored for offline analysis.

Experimental protocol

During the whole experiment, participants were seated with their left hand resting pronated on a table. Monophasic TMS pulses with a posterior-anterior current direction were applied over the right hemisphere, with the coil handle pointing 45 degrees from the midsagittal line throughout the protocol. The experimenter received visual feedback of the current coil position as well as previous coil positions color coded with the MEP-amplitudes. First, the head of the subject was co-registered to a stock MRI scan by defining the nasion, pre-auricular points and at least 100 data points spread out over the scalp. Second, the hotspot, the location with the largest MEPs, was estimated with pseudorandom acquisition using 80 pulses with a 2 second interval [5]. The stimulation intensity was set to 50% of maximum stimulator output (MSO) and increased with 5% MSO if there were no measurable MEPs after 15 pulses. Third, the resting motor threshold

(RMT), the lowest stimulator intensity that has ≥ 50% chance to produce a MEP at the hotspot, was determined with the Motor Threshold Assessment Tool (MTAT 2.0) [10]. EMG-responses with a peak-to-peak amplitude ≥ 0.05 mV, between 5 and 45 milliseconds after stimulation, were considered MEPs. Finally, the motor maps were acquired with a stimulation intensity of 110% RMT [4].

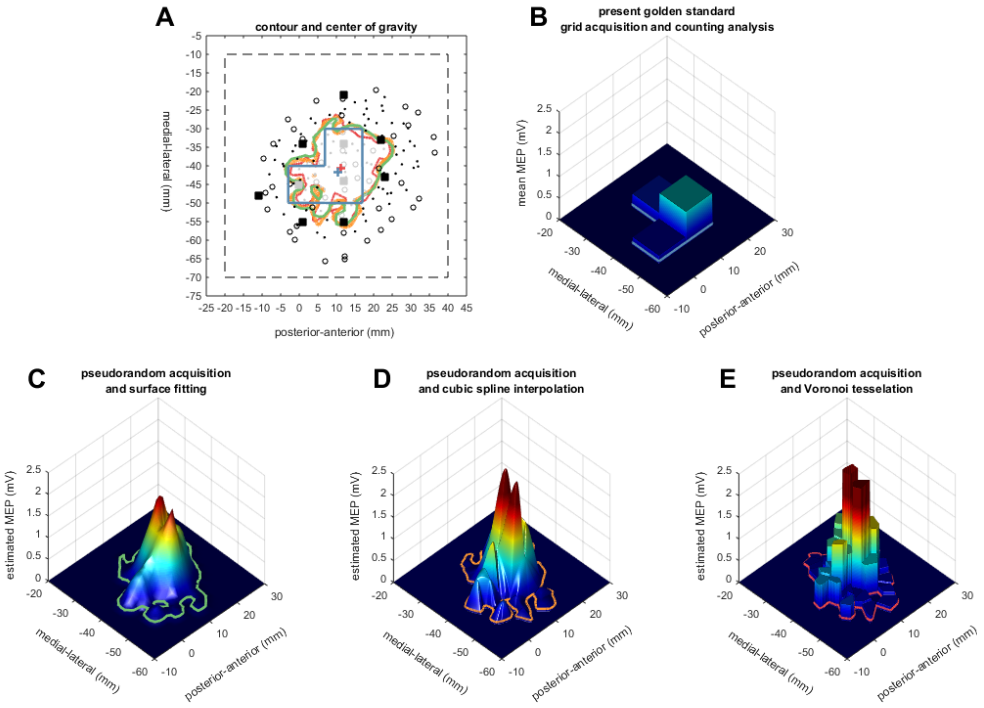


Figure 1. Methods of pseudorandom data acquisition and digital reconstruction with surface fitting (green), cubic spline interpolation (orange) and Voronoi tessellation (red) compared to the golden standard of grid acquisition and a counting analysis (blue). A. 2D representation of the motor maps. Black and grey markers depict negative and positive stimulation sites: squares depict the grid points of the grid acquisition and circles and points depict the first 50 and remaining 100 stimuli of the pseudorandom acquisition. The estimates of the cog and borders of the motor map are depicted by plus signs and solid lines in the corresponding colors of each method. The dashed square represents a 6-by-6cm predefined region which was used in previous studies with pseudorandom stimulation [5,7,11]. **B.** 3D representation of the counting analysis after grid acquisition. C-E: 3D representation of the three digital reconstruction methods, after the same pseudorandom acquisition. MEP = motor evoked potential.

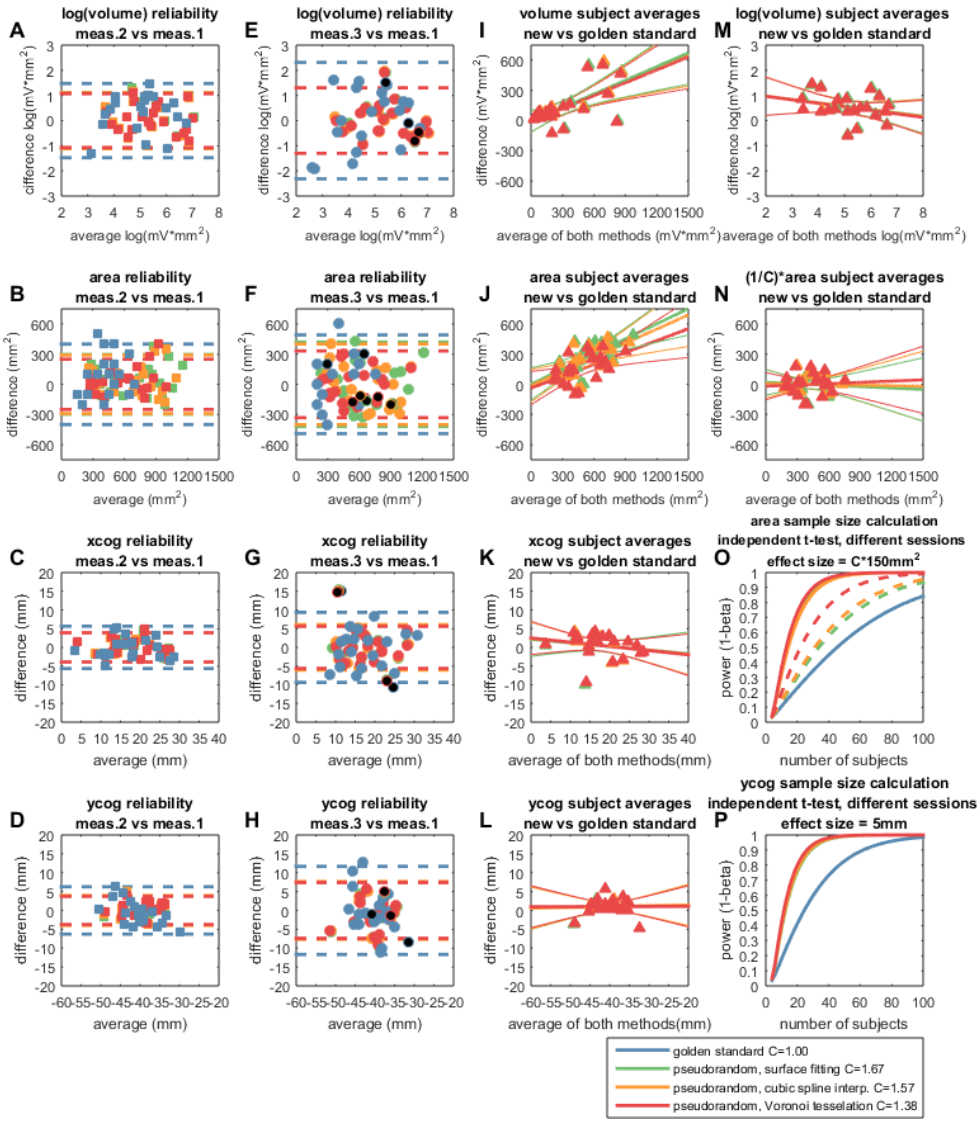


Figure 2. Results of pseudorandom data acquisition and digital reconstruction with surface fitting (green), cubic spline interpolation (orange) and Voronoi tessellation (red) compared to the golden standard of grid acquisition and a counting analysis (blue). A-H. Bland-Altman plots depicting the within subject differences of each method for measurements acquired in the same session (**A-D**) or in separate sessions (**E-H**). Black filled markers depict the two subjects that were removed because of a co-registration error. Dashed lines depict the smallest detectable change, which was smaller for the reconstruction methods, but still large compared to effect sizes found in clinical studies [1,2]. Importantly, the within subject differences did not increase with the averages. This indicates that the methods are equally reliable for small and large muscle representations. **I-N.** Bland-Altman plots depicting the between method differences of the subject averages. Solid lines depict linear regression lines and their confidence interval. The heteroscedasticity of the volume parameter was successfully removed with a log transformation. The estimations of the area parameter after digital reconstruction were systematically 67% (surface fitting), 57% (cubic spline

interpolation) and 38% (Voronoi tessellation) larger compared to the golden standard. **O-P.** Examples of power calculations (alpha = 0.05) using the absolute reliability of this experiment and the effect sizes from previous clinical studies [1,2], The effect size of area was adjusted for the bias between the methods. Dashed lines depict the power calculation without this adjustment. Meas. = measurement; C = scaling constant for motor map area; xcog = center of gravity in the posterior anterior direction; ycog = center of gravity in the medial-lateral direction.

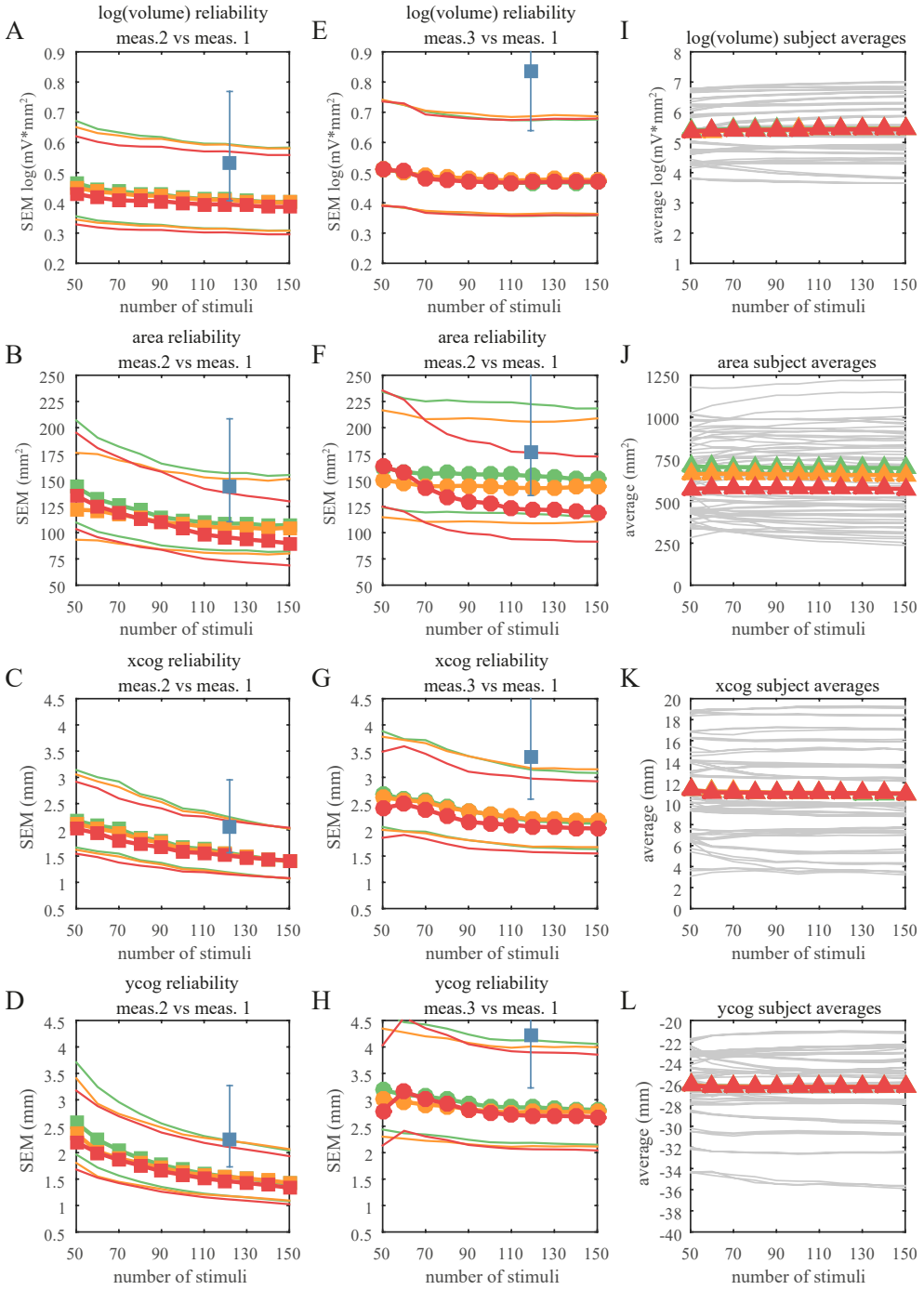
Grid acquisition

Grid acquisition was based on the well-established paradigm by Kleim, et al (2007) [4]. Ten pulses with an interval of 7 seconds were applied on the points of a 1-by-1cm spaced grid. A point was marked positive when at least half of the stimuli resulted in a MEP. Grid points were stimulated row by row, moving outward from the center, until a positive area was demarcated by negative points (Figure 1A). The pseudorandom acquisition was adapted from Van de Ruit et al. [5] and used 150 pulses. An improvement was made by first creating a subject-specific region of interest with 50 pulses, which prevented muscle representations exceeding the borders of a predefined region [7,11]. These 50 pulses were applied in 8 straight lines outward from the hotspot until 2 consecutive pulses (6.8±0.8mm apart) did not elicit a MEP, followed by a clockwise ellipsoid around these lines (Figure 1A). The experimenter applied the remaining 100 pulses pseudorandomly inside the ellipsoid.

The grid and pseudorandom acquisition were both performed three times in total (measurement 1-3), twice during the first session and once on the consecutive day. Each session started with determining the RMT, as is done in the follow up of intervention studies [2]. Then, the acquisition methods were performed alternatingly, with the two possible orders counterbalanced between subjects.

Data analysis

From each dataset, four parameters were calculated: area, volume and center of gravity in two dimensions. Data analysis was conducted offline with a custom-made MATLAB script (Mathworks, USA). First, stimuli were excluded if the root mean square of the background EMG, 100-5 milliseconds before stimulation, was more than 2 standard deviations above the average, or the coil position was outside the 99% probability interval. Second, a plane was fitted through the 3D coordinates (x,y,z) of the first measurement. The z-coordinates were transposed on this plane. The center of the coordinates (x,y,new-z) was used to translate the coordinates to the origin, which were then subsequently rotated around the x, y and z axis. [5]. The same plane, translation and rotation-matrix were used for the other two measurements. For each measurement, the error between the z and new-z coordinates was calculated. Third, after pseudorandom acquisition, grids were reconstructed with three methods: surface fitting, cubic spline interpolation and Voronoi tessellation. For surface fitting the gridfit algorithm was used to create a 1.2-by-1.2mm spaced grid [5,7]. For cubic spline interpolation and Voronoi tessellation, the griddata algorithm (method set to cubic or nearest) was used to create 0.1-by-0.1mm spaced grids [6]. Points in these reconstructed grids where the estimated EMG-amplitude was below 0.05mV, were set to 0. After standard grid acquisition, the counting of MEPs was repeated offline. For positive points the mean EMG-amplitude was calculated and negative points were set to 0. Finally, the motor map parameters were calculated. Volume was computed as the sum of the positive cell areas multiplied with their corresponding EMG-amplitudes and area as the sum of all positive cell areas. The cog was calculated for the posterior-anterior (xcog) and the medial-lateral direction (ycog) and added to the translation of the plane.



Supplementary Figure. Relation between the number of pseudorandom stimuli and the SEM and average of the motor map parameters for surface fitting (green), cubic spline interpolation (orange) or Voronoi tessellation (red). Thirty datasets were created by shuffling the last 100 stimuli of each

measurement. In each dataset volume, area, xcog and ycog were calculated for the first 50 to 150 stimuli in steps of 10. **A-H.** The average SEM of the 30 datasets. Colored squares and circles depict the SEM in the same session and in separate sessions a day apart. Thin lines depict the confidence interval. Even with the previously recommended 80 [5] and 110 [7], the reconstruction methods provide at least equal reliability compared to the golden standard (blue). However, the SEM in the same session appears to improve even after 110 pulses. **I-L.** The subject averages of the 30 datasets. Grey lines depict the subject averages and colored triangles the overall averages. Importantly, after approximately 90 stimuli the subject averages remain stable.

Statistical analysis

First, the subject averages of volume, area, xcog and ycog were calculated for the golden standard and the three digital reconstruction methods. These subject averages were used to inspect the between-method differences with Bland-Altman plots and to compute the overall average of each method. Second, for each method separately, the within-subject differences between measurement 2 and 1 (same session) and measurement 3 and 1 (separate sessions) were inspected with Bland-Altman plots as well. Finally, the standard error of measurement of each method was calculated from the standard deviation of these within-subject differences (SEM = $SD_{diff\_within} / \sqrt{2}$) as was the smallest detectable change (SDC = $SD_{diff\_within} * 1.96$) [12]. To illustrate the reliability, examples of sample size calculations are provided. Most intervention studies compare the plasticity in a treatment group to a control group. In this scenario, the primary outcome is a change in motor map parameters. Therefore, the SD_{diff_within} of this study is an estimate for the standard deviation of the groups. The parameter values are denoted as overall average ± between subject standard deviation and the SEM is denoted with a confidence interval.

Results and Discussion

During data analysis, 3.2±1.9% of the stimuli were excluded. One subject was removed from the within session analysis, because there were no positive sites during the second grid acquisition. Furthermore, two outliers were removed from the between sessions analysis because the xcog (3.3mm, first session) and the z-error (9.2mm, measurement 3) indicated a co-registration error. The average z-error was 1.1±0.4mm, 1.3±0.5mm and 2.3±0.9mm for measurement 1, 2 and 3.

The SEM of the reconstruction methods was equal or smaller than the golden standard (Table 1, Figure 2A-H). Therefore, the present golden standard using 122±44 stimuli in 17±7 minutes can be replaced by the much faster reconstruction methods using 150 stimuli in 5 minutes, without sacrificing reliability (Supplementary Figure).

Power calculations indicate the reconstruction methods can reduce the number of subjects needed in intervention studies (Figure 2O,P). It is important to note that the reconstruction methods increased the area estimates with 67% (surface fitting), 57% (cubic spline) and 38% (Voronoi tessellation) relative to the golden standard (Figure 2J,N). This bias was circumvented by normalizing the effect sizes to the overall mean of each method.

Regarding individual patients, all motor map parameters showed a considerable SDC and should be interpreted with caution on an individual level (Figure 2A-H).

References

1. Liepert, J. et al. *Neurosci. Lett.* 250, 5–8 (1998).
2. Sawaki, L. et al. *Neurorehabil. Neural Repair* 22, 505–513 (2008).
3. Wassermann, E.M., McShane, L.M., Hallett, M. & Cohen, L.G. *Electroencephalogr. Clin. Neurophysiol. Evoked Potentials* 85, 1–8 (1992).

4. Kleim, J.A., Kleim, E.D. & Cramer, S.C. *Nat. Protoc.* 2, 1675–84 (2007).
5. Van De Ruit, M., Perenboom, M.J.L. & Grey, M.J. *Brain Stimul.* 8, 231–239 (2015).
6. Julkunen, P. *J. Neurosci. Methods* 232, 125–133 (2014).
7. Cavaleri, R., Schabrun, S.M. & Chipchase, L.S. *Brain Stimul.* 1–5 (2018).doi:10.1016/j.brs.2018.07.043
8. Schambra, H.M. et al. *Front. Cell. Neurosci.* 9, 1–18 (2015).
9. Rossi, S., Hallett, M., Rossini, P.M. & Pascual-Leone, A. *Clin. Neurophysiol.* 122, 1686 (2011).
10. Awiszus, F. & Borckardt, J.
11. van de Ruit, M. & Grey, M.J. *Brain Topogr.* 29, 56–66 (2016).
12. Bartlett, J.W. & Frost, C. *Ultrasound Obstet. Gynecol.* 31, 466–475 (2008).

4.2 Cerebellar transcranial direct current stimulation interacts with BDNF Val66Met in motor learning

Rick van der Vliet, Zeb D. Jonker, Suzanne C. Louwen, Marco Heuvelman, Linda de Vreede, Gerard M. Ribbers, Chris I. De Zeeuw, Opher Donchin, Ruud W. Selles, Jos N. van der Geest and Maarten A. Frens

Abstract

Background. Cerebellar transcranial direct current stimulation has been reported to enhance motor associative learning and motor adaptation, holding promise for clinical application in patients with movement disorders. However, behavioral benefits from cerebellar tDCS have been inconsistent.

Objective. Identifying determinants of treatment success is necessary. BDNF Val66Met is a candidate determinant, because the polymorphism is associated with motor skill learning and BDNF is thought to mediate tDCS effects.

Methods. We undertook two cerebellar tDCS studies in subjects genotyped for BDNF Val66Met. Subjects performed an eyeblink conditioning task and received sham, anodal or cathodal tDCS (N=117, between-subjects design) or a vestibulo-ocular reflex adaptation task and received sham and anodal tDCS (N=51 subjects, within-subjects design). Performance was quantified as a learning parameter from 0 to 100%. We investigated (1) the distribution of the learning parameter with mixture modeling presented as the mean (M), standard deviation (S) and proportion (P) of the groups, and (2) the role of BDNF Val66Met and cerebellar tDCS using linear regression presented as the regression coefficients (B) and odds ratios (OR) with equally-tailed intervals (ETIs).

Results. For the eyeblink conditioning task, we found distinct groups of learners ($M_{Learner}$=67.2%; $S_{Learner}$=14.7%; $P_{Learner}$=61.6%) and non-learners ($M_{Non\text{-}learner}$=14.2%; $S_{Non\text{-}learner}$=8.0%; $P_{Non\text{-}learner}$=38.4%). Carriers of the BDNF Val66Met polymorphism were more likely to be learners (OR=2.7 [1.2 6.2]). Within the group of learners, anodal tDCS supported eyeblink conditioning in BDNF Val66Met non-carriers (B=11.9% 95%ETI=[0.8 23.0]%), but not in carriers (B=1.0% 95%ETI=[-10.2 12.1]%). For the vestibulo-ocular reflex adaptation task, we found no effect of BDNF Val66Met (B=-2.0% 95%ETI=[-8.7 4.7]%) or anodal tDCS in either carriers (B=3.4% 95%ETI=[-3.2 9.5]%) or non-carriers (B=0.6% 95%ETI=[-3.4 4.8]%). Finally, we performed additional saccade and visuomotor adaptation experiments (N=72) to investigate the general role of BDNF Val66Met in cerebellum-dependent learning and found no difference between carriers and non-carriers for both saccade (B=1.0% 95%ETI=[-8.6 10.6]%) and visuomotor adaptation (B=2.7% 95%ETI=[-2.5 7.9]%).

Conclusions. The specific role for BDNF Val66Met in eyeblink conditioning, but not vestibulo-ocular reflex adaptation, saccade adaptation or visuomotor adaptation could be related to dominance of the role of simple spike suppression of cerebellar Purkinje cells with a high baseline firing frequency in eyeblink conditioning. Susceptibility of non-carriers to anodal tDCS in eyeblink conditioning might be explained by a relatively larger effect of tDCS-induced subthreshold depolarization in this group, which might increase the spontaneous firing frequency up to the level of that of the carriers.

Introduction

Over the past decade, cerebellar transcranial direct current stimulation (tDCS) has been reported to enhance motor associative learning [1] and motor adaptation [2–10] (see [11] for a review of the technical details), holding promise for patients with movement disorders [12]. However, cerebellar tDCS effects are inconsistent across the literature, as recent studies failed to replicate initial behavioral benefits [13–15]. This could mean that the behavioral gains reported in earlier studies result from chance and/ or that determinants predicting successful tDCS are incompletely understood. Genetic differences between individuals might influence (1) the background performance level and therefore the potential to improve with tDCS [16] or (2) the susceptibility to tDCS. Therefore, to increase predictability of cerebellar tDCS effectiveness it is important to identify factors which modify treatment success [17], like genetic variants.

The common [18,19] BDNF Val66Met polymorphism, which decreases activity-dependent BDNF release [20], is a candidate determinant of cerebellar tDCS effectiveness, because (1) the polymorphism is associated with motor skill learning ability [21,22] and (2) BDNF is thought to mediate tDCS effects on synaptic plasticity and motor skill learning [22]. Since BDNF supports long-term potentiation [22,23] and formation of inhibitory synapses [24], Val66Met carriers have subtle behavioral alterations such as decreased memory [20], slowed motor skill learning [21,22] and more pronounced fear conditioning [25]. In addition, in mouse cortical brain slices, concurrent DCS and synaptic activation only leads to long-term potentiation when BDNF is not knocked out or blocked [22], suggesting that Val66Met carriers may benefit less from tDCS. However, whether BDNF Val66Met interacts with cerebellar tDCS in cerebellum-dependent motor learning has not yet been investigated.

Eyeblink conditioning and vestibulo-ocular reflex (VOR) adaptation are particularly well-characterized cerebellum-dependent learning tasks for which positive effects of cerebellar tDCS have been found. Eyeblinks are protective eyelid closures against damage to the cornea. They can be activated in response to predictive neutral cues such as auditory tones. This learned motor association is made in a relatively simple circuitry involving the interposed nucleus and lobule VI of the cerebellum [26–29] and extracerebellar areas in the hippocampus and amygdala [30–34]. Eyeblink conditioning is mediated by a sudden, carefully timed decrease in simple spike activity of cerebellar Purkinje cells that fire at a relatively high spontaneous firing frequency [28,35,36]. Zuchowski et al. found an increase in eyeblink conditioning with anodal tDCS and a decrease with cathodal tDCS [1], which is in line with the concept that eyeblink Purkinje cells should operate at a sufficiently high simple spike firing frequency during spontaneous activity, because anodal tDCS is supposed to increase the baseline firing frequency of neurons [37–40]. The VOR generates eye movements opposite in direction, but with identical speed as head rotation to stabilize objects of interest on the retina. Changes in the environment or the body can make this relation inappropriate and result in retinal slip [41]. Retinal slip will recruit adaptive mechanisms in the cerebellar flocculus and downstream vestibular nuclei to increase (gain-increase adaptation) or decrease eye (gain-decrease adaptation) movement velocity and regain foveal stabilization [42–48]. VOR gain-decrease adaptation, which will be studied in this paper, is mediated by decreased velocity sensitivity of neurons in vestibular nuclei, at least partially induced by plasticity mechanism involving floccular Purkinje cells [44–47]. Recently, anodal cerebellar DCS during VOR adaptation was found to enhance learning rate of a gain-decrease paradigm in mice [9]. Therefore, eyeblink conditioning and VOR adaptation are two cerebellum-dependent, but fundamentally different tasks, which entail different cellular mechanisms, and which concern

conceptually different paradigms in that conditioning implies learning new associations, whereas adaptation involves recalibrating and optimizing existing behavior.

The primary aim of this study was to investigate the interaction between BDNF Val66Met and cerebellar tDCS in eyeblink conditioning and VOR adaptation. To this end, we undertook two studies in genotyped subjects who received cerebellar tDCS and performed either an eyeblink conditioning task (N=117, between-subjects design) or a VOR adaptation task (N=51, within-subjects design). Based on motor skill learning studies [21,22], we expected better performance for non-carriers in both tasks and therefore a more pronounced effect of cerebellar tDCS in carriers. Based on fear conditioning studies [25], we expected better performance for carriers in the eyeblink conditioning task, which depends on the amygdala as well as the cerebellum, but not in the VOR adaptation task and therefore a more pronounced effect of cerebellar tDCS in non-carriers.

Materials and methods

Subjects

Healthy right-handed, defined as having a Edinburgh handedness inventory score [49] larger than zero, individuals participated in the eyeblink conditioning (genetic analysis failed in 3/120 subjects leaving 117 for analysis), VOR adaptation (genetic analysis failed in 4/55 subjects leaving 51 for analysis) (see Table 1). 9/51 subjects dropped out before the second VOR session but the available data of the first session was included in the analysis. The order for the visuomotor and saccade adaptation experiment was counterbalanced. The experiments were approved by the Erasmus MC medical ethics committee and performed in accordance with the Declaration of Helsinki.

Paradigm	Group	Gender (%Male)	Age (M±SD)	Ethnicity (%Western-European)	Edinburgh handedness scale
Eyeblink conditioning (N=117)	Sham (N=39)	41	21.5±2.8	85	
	Carriers (N=16)	31	21.3±2.4	81	
	Non-carriers (N=23)	48	21.7±3.2	87	
	Anodal (N=40)	40	20.6±2.5	85	
	Carriers (N=17)	35	20.8±2.9	82	
	Non-carriers (N=23)	43	20.4±2.2	87	
	Cathodal (N=38)	42	20.9±2.4	82	
	Carriers (N=14)	57	21.4±2.5	79	
	Non-carriers (N=24)	33	20.6±2.4	83	
VOR adaptation (N=51)	Carriers (N=18)	11	21.8±3.1	89	
	Non-carriers (N=33)	42	21.7±2.7	82	
Saccade adaptation and visuomotor adaptation (N=72)	Carriers (N=25)	40	21.6±2.0	87	82.2±15.6
	Non-Carriers (N=47)	38	21.1±2.5	91	79.0±17.9

Table 1. Subject characteristics for the eyeblink conditioning and VOR adaptation experiments. M=mean; S=standard deviation.

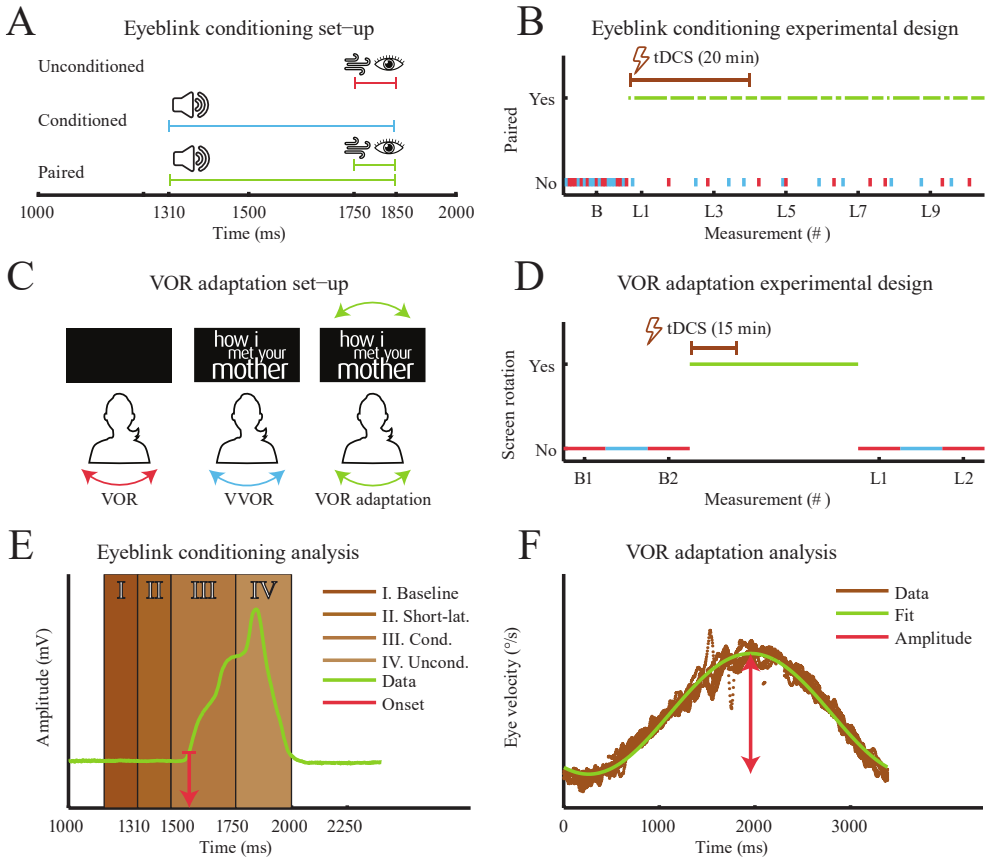


Figure 1. Experimental procedures and data-analysis for eyeblink conditioning and VOR adaptation. **A.** Eyeblink conditioning set-up. The experiment consisted of unconditioned stimulus trials (red line), conditioned stimulus trials (blue line) and paired stimulus trials (green line). During each trial, eyelid movements were recorded for 3 seconds. For the unconditioned stimulus trials, an air puff (3 bar at source, 100 ms duration) was delivered from 1750 until 1850 ms after recording onset. For the conditioned stimulus trials, a tone (650 Hz, 75 dB, 540 ms duration) was played from 1310 ms until 1850 ms after recording onset. In paired stimulus trials, subjects received both the tone and the air puff, which overlapped for 100 ms. **B.** Eyeblink conditioning experimental design. The experiment started with a baseline measurement (B) comprised of ten unconditioned stimulus trials (red lines) and ten conditioned stimulus trials (blue lines), followed by ten learning measurements (L1-L10) consisting of ten paired trials (green lines), one unconditioned stimulus trial (red lines) and one conditioned stimulus trial (blue lines). **C.** VOR adaptation set-up. The experiment consisted of VOR trials (red line), VOR adaptation trials (blue lines) and VVOR trials (green line). During VOR measurements, subjects were asked to keep their eyes fixated at the middle of the screen during rotation in total darkness for 40 seconds. Rotation of the chair was paused for 30 seconds before each VOR measurement. During VVOR trials, subjects were rotated for 1 minute while the movie was projected stationary on the middle of the screen. During VOR adaptation trials, the projection was rotated for 60 minutes with identical phase and amplitude as the chair. **D.** VOR adaptation experimental design. The experiment started with two baseline VOR trials (red lines, measurements B1 and B2), separated by a single VVOR trial (green line). Subsequently, subjects underwent a single VOR adaptation trial (blue line) and two VOR trials (red lines, measurements L1 and L2). **E.** Eyeblink conditioning analysis. Each trial (green line) was divided into (I) a baseline window of 150 ms before the start of the conditioned stimulus from t_{start}=1160 ms

until t_{end}=1310 ms; (II) a short-latency response window of 150 ms after the start of the conditioned stimulus from t_{start}=1310 ms until t_{end}=1460 ms; (III) a conditioned response window of 290 ms before the start of the unconditioned stimulus from t_{start}=1460 ms until t_{end}=1750 ms; and (IV) an unconditioned response window of 250 ms after the start of the unconditioned stimulus from t_{start}=1750 ms until t_{end}=2000 ms. Trials were classified as the window that contained the eyeblink onset. **F.** VOR adaptation analysis. Forty-second eye velocity signals were cut into eleven rotations (brown line) of 3.39 s, aligned in time and fitted with a sine wave (green line) of the same frequency to extract the amplitude (red arrow). tDCS = transcranial direct

Cerebellar tDCS

Cerebellar tDCS was delivered through two saline-soaked 5x5cm sponge electrodes (DC stimulator, NeuroConn GmbH, Ilmenau, Germany) placed on the right side of the scalp, 3 cm lateral to the inion (target electrode) and on the ipsilateral buccinator muscle (reference electrode). This electrode configuration is the standard for cerebellar tDCS in motor learning tasks [1–5,10] and is supported by electrophysiological [50] and modeling studies [51]. In the active conditions, we applied 2mA anodal or cathodal tDCS during 20 minutes for the eyeblink conditioning experiment (similar to: [14]) and 2mA anodal tDCS during 15 minutes for the VOR adaptation experiment (most commonly used duration: [52]). In the sham condition, 2mA anodal or cathodal tDCS was delivered for only 30 seconds, which is an effective method for blinding subjects [53]. In both the active and sham condition, current amplitude was increased and decreased in a ramp-like fashion over 30 seconds according to a well-established protocol [2]. Experimenters were blinded using a list of stimulation codes corresponding with sham or active stimulation. This list was semi-randomized, balancing the number of subjects in each condition.

Genetics

The BDNF Val66Met polymorphism (rs6265) was genotyped using TaqMan assays as described before [54]. Subjects with at least one Met allele were termed “carriers”, others “non-carriers”.

Eyeblink conditioning

Eyeblink conditioning was studied by presenting an auditory tone (conditioned stimulus) shortly before applying an air puff to the eye (unconditioned stimulus) [55,56], similar to Zuchowski et al. [1]. Over trials, subjects learn to predict the air puff from the tone and close the eyelid before the puff reaches the cornea. We chose a between-subject design for this task, even though a within-subject design could have removed between-subject variability, because the motor memory in eyeblink conditioning is retained for a long time [57]. Furthermore, we included anodal as well as cathodal tDCS because both have been found to modulate eyeblink conditioning [1].

We used a SheBot system (Neurasmus, Rotterdam, The Netherlands, [58]) controlled by a custom-built LabVIEW program (National Instruments Corporation, Austin, Texas, United States) to provide precisely timed (1) auditory tones via a headphone and (2) air-puffs via a nozzle placed 15 mm from the lateral corner of the eye. Eyelid closures were recorded with a small magnet on the upper eyelid and a sensor slightly below the eye [58]. During the experiment, subjects watched the movie “A Beautiful Mind” (Universal Pictures, 2005, Internet Movie DataBase #tt0268978) with audio but without subtitles.

The experiment consisted of unconditioned stimulus trials, conditioned stimulus trials and paired stimulus trials (see Figure 1A). The experiment started with a baseline measurement (B) comprised of ten unconditioned stimulus trials and ten conditioned stimulus trials, followed by ten learning measurements (L1-L10) consisting of ten paired trials, one unconditioned stimulus trial and one conditioned stimulus trial (Figure 1B). For each measurement, trial order

and trial interval (ranging from 20 to 35 seconds) were pseudo-randomized. Cerebellar tDCS or sham stimulation started with L1.

Eyeblink data was automatically processed using a custom MATLAB program (The MathWorks Inc., Natick, Massachusetts, United States) (see Figure 1E). First, trials were low-pass filtered with a zero-phase 6th order Butterworth filter using a 100 Hz cut-off frequency. Subsequently, trials were divided in four time windows: (1) a baseline window; (2) a short-latency response window; (3) a conditioned response window; and (4) an unconditioned response window. Peak time (ms) occurred at maximum eyelid closure in the conditioned or unconditioned response window. Peak amplitude (mV) was the difference between the eyelid signal at peak time and the mean eyelid signal in the baseline window. Eyeblink onset (ms) occurred at the last time point when the eyelid signal was smaller than 7.5% of the peak amplitude. The analysis was robust to small changes in the peak amplitude threshold. Trials were classified by the window that contained the eyeblink onset. Short-latency response and baseline responses were counted as invalid trials.

The learning parameter for this experiment was the percentage of conditioned responses in the last six learning blocks L5-L10. That is, the number of conditioned responses divided by the total number of conditioned and unconditioned responses in the 60 paired trials of the learning measurements L5-L10 multiplied by 100 (0=no conditioning; 100=complete conditioning). In addition, we investigated the short-latency responses as a percentage of the sum of conditioned responses, unconditioned responses and short-latency responses (short-latency response fraction).

VOR adaptation

VOR adaptation was studied by directly coupling head rotation to visual display rotation, which requires suppression of the reflex to minimize retinal slip [59–61], similar to an animal study performed by Das et al. [9]. In contrast to the eyeblink conditioning experiment, we chose a within-subject design as the motor memory is expected to last no more than three days [62–64]. Both stimulation sessions were separated by at least 7 days to ensure wash-out of the cerebellar tDCS [65,66] and VOR adaptation effects [62–64] of the first session. Furthermore, we did not include a cathodal condition to limit the number of experimental conditions for our subjects.

Subjects were seated in a rotational chair placed 224 cm in front of a wide translucent screen (235 cm x 170 cm). Head position was fixed relative to the chair with a bite-board (Dental Tecno Benelux, Ede, Netherlands). Chair rotation frequency was set to 0.295 Hz with an amplitude of 12° around the vertical axis, resulting in a peak angular velocity of 22.2 °/s (similar to [67]). Two-dimensional binocular eye movements were recorded using infrared video-oculography (EyeLink I, SR Research, Ontario, Canada, [68]). An episode of "How I Met Your Mother" with audio but without subtitles (Twentieth Century Fox Film Cooperation, 2005, Internet Movie DataBase #tt0460649) was back-projected (Infocus LP 335, Portland, Oregon, United States) onto the translucent screen (size 104 x 74 cm) via rotatable mirrors (model number 6900, Cambridge Technology, Cambridge, United Kingdom).

Trial types included VOR, visually-enhanced vestibulo-ocular reflex (VVOR) trials and VOR adaptation trials (see Figure 1C). The experiment started with two baseline VOR trials (measurements B1 and B2), separated by a single VVOR trial. Subsequently, subjects underwent a single VOR adaptation trial and two VOR trials (measurements L1 and L2) (Figure 1D).

Eye movement data was processed in MATLAB (The MathWorks Inc., Natick, Massachusetts, United States) (see Figure 1F). Eye velocity gains were calculated per subject, eye

and measurement (B1-2 and L1-2) according to the following procedure. First, saccades and eyeblinks were removed from the horizontal eye position using an internal Eyelink routine. Subsequently, the horizontal eye position was smoothened and differentiated with a Savitzky-Golay filter (third order polynomial, 10 Hz critical frequency) to obtain an eye movement velocity signal (°/s). Eleven rotations of 3.39 s from this 40-second velocity signal were aligned in time and fitted with a sine wave of the same frequency. Fitted velocity amplitudes (°/s) of left and right eye velocity signals were combined for each block by weighing with the number of recorded data points. Finally, all amplitudes were divided by the mean amplitude in B1 and B2 resulting in a normalized gain.

The learning parameter for this experiment was one minus the average amplitude of learning measurements L1 and L2 multiplied by 100 (0=no adaptation; 100=complete adaptation).

Saccade adaptation

Saccade adaptation was studied by relocating a target in an inward direction during a saccade to induce a post saccadic error [69–71]. Over trials, subjects learn to decrease the size of their saccades to compensate for these target jumps.

Subjects were seated in front of a monitor covered with a red filter (53 cm width, 1280x1024 pixel resolution) in a completely dark room. Steady head position was maintained using a chin rest at a fixed viewing distance of 82 cm. Eye movements were recorded binocularly at 250 Hz by means of video-oculography (EyeLink II, SR Research, Ontario, Canada).

Task design was similar to Avila et al. [6], but with smaller amplitude saccades (10° rather than 20°) to reduce the occurrence of two-step saccades. The trial types were unperturbed and perturbed trials (see Figure 6A). The experiment included baseline measurements of 50 unperturbed trials (measurements B1-50), followed by learning measurements of 150 perturbed trials (measurements L1-150) (see Figure 6B).

Saccade amplitudes were calculated using an internal Eyelink routine. All amplitudes were divided by 10° to calculate normalized gains and corrected for an offset by subtracting the median amplitude of the baseline measurements. The learning parameter was defined as the quotient of 1 minus the median of L150-200, and the perturbation size 0.3.

Visuomotor adaptation

Reaching movement adaptation to visual mismatches was studied with visuomotor adaptation, wherein visual feedback of hand location is rotated with respect to actual reaching movement [72–74]. Subjects adjust their movement based on this visual mismatch by changing the angle of their reaches.

Subjects were seated in front of a vertical monitor (48 cm width, 1280x1024 pixel resolution, distanced 60 cm from the subjects) while holding a robotic handle in their right hand (custom-made, see [75]) which recorded hand position and velocity. To remove direct visual feedback of hand position, subjects wore an apron that was attached to the table around their neck.

Task design was similar to Galea et al. [2]. The trial types were unperturbed trials and perturbed trials (see Figure 6C). The experiment design included baseline measurements of unperturbed trials (measurements B1-192) and learning measurements of perturbed trials (measurements L1-200) (see Figure 6D).

	Single group			Learner / non-learner						
	DIC	**M**	**S**	**DIC**	**P_{NL}**	**M_{NL}**	**S_{NL}**	**P_L**	**M_L**	**S_L**
Eyeblink conditioning	1121	46.8	28.8	1067	38.4	14.2	8.0	61.6	67.2	14.7
VOR adaptation	384	29.1	10.3	416	69.5	26.3	7.0	30.5	38.5	10.5
Saccade adaptation	635	55.9	19.6	695	40.1	45.2	15.4	59.9	60.7	14.1
Visuomotor adaptation	548	71.0	10.7	603	50.2	66.5	5.7	49.8	77.8	6.6

Table 2. Mixture model results. The learning parameters for eyeblink conditioning, VOR adaptation, saccade adaptation and visuomotor adaptation were modeled with (1) a single normal distribution and (2) a learner / non-learner model composed of a mixture of two normal distributions. We compared model fit with the DIC, with lower DICs indicating better model fits. Eyeblink conditioning was best captured with a learner/ non-learner model, whereas the adaptation paradigms were best described with a single group model. DIC = deviance information criterion, L = learner, M = mean, NL = non-learner, S = standard deviation.

Visuomotor adaptation data was processed using MATLAB (The MathWorks Inc., Natick, Massachusetts, United States). From each trial, we extracted movement start, defined as the time point when movement velocity exceeded 0.03 m/s, and movement end, defined as the moment when displacement from origin was equal to or larger than 9.5 cm. Aiming direction was calculated as the signed (+ or -) angle in degrees between the vector connecting origin and target and the vector connecting the positions of the manipulandum at movement start and movement end. The clockwise direction was defined as positive. Aiming directions more than 30° away from the median of an epoch of 8 trials across all subjects, were removed from further analysis.

The learning parameter for this experiment was the negative average of L9 through L88 divided by the perturbation size of 30° (similar to Galea et al. [2]).

Statistics

We used a two-step approach to data-analysis of the learning parameter.

First, we investigated whether the distribution of the learning parameter was best captured by either a single normal distribution (unimodal) or a mixture of two normal distributions (bimodal). The latter distribution could arise if one group of subjects is able to learn the task (learners) whereas the other group is not (non-learners). For this analysis, we used a Bayesian Gaussian mixture model fitting one or two normal distributions to the learning parameter (averaged across stimulation conditions for the VOR adaptation experiment), with a beta prior for the probability of being a learner or a non-learner. We set the lower limit on the prior probability of being a learner or non-learner to 0.15 and the upper limit to 0.85 to neglect clusters smaller than 15% of the total population. Quality of the two models was compared for each paradigm with the deviance information criterion (DIC) according to [76], which rewards high likelihood and penalizes model complexity.

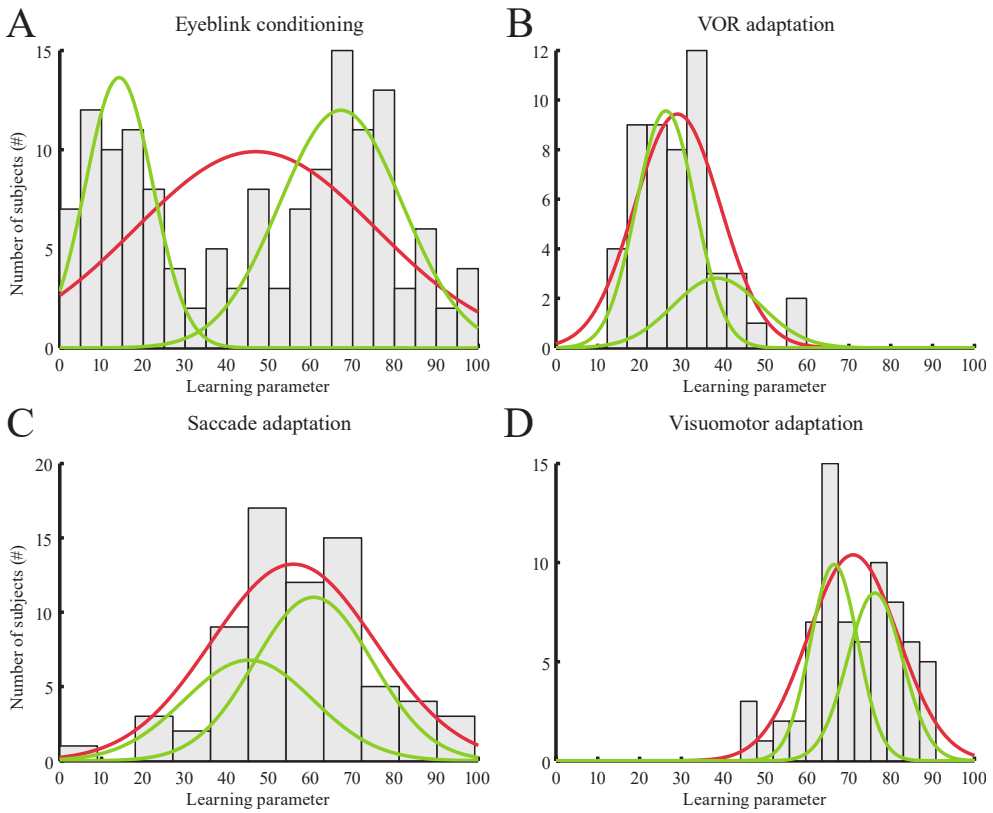


Figure 2. Learning parameter distributions. A-D. Histograms of the learning parameter for eyeblink conditioning (A), VOR adaptation (B), saccade adaptation (C) and visuomotor adaptation (D). The red Gaussians show the unimodal distributions. The green Gaussians the bimodal distribution.

Second, in case the learning parameter was best captured by a unimodal distribution, the learning parameters of all subjects were studied with a 'single group' Bayesian linear regression model (independent variables described below). However, if the learning parameter was best captured by a bimodal distribution, we performed an additional 'learner / non-learner' regression analysis. For this analysis, we labeled the subjects as "learners" and "non-learners" based on the group the subjects were assigned to most in the mixture model and calculated (1) a Bayesian logistic regression model for the probability of being a learner, and (2) a Bayesian linear regression model for the learning parameter of the "learners" only (independent variables described below). For eyeblink conditioning (between-subjects), the regression model contained the independent variables "carrier", "$anodal_{Carrier}$", "$anodal_{Non\text{-}carrier}$", "$cathodal_{Carrier}$" and "$cathodal_{Non\text{-}carrier}$". For VOR adaptation (within-subjects), the regression model contained the independent variables "carrier", "$anodal_{Carrier}$" and "$anodal_{Non\text{-}carrier}$".

The short-latency response fraction was analyzed by fitting beta distributions to the carriers and non-carriers and calculating the difference in group means. It was necessary to use beta distributions because short-latency response fraction was heavily skewed towards zero.

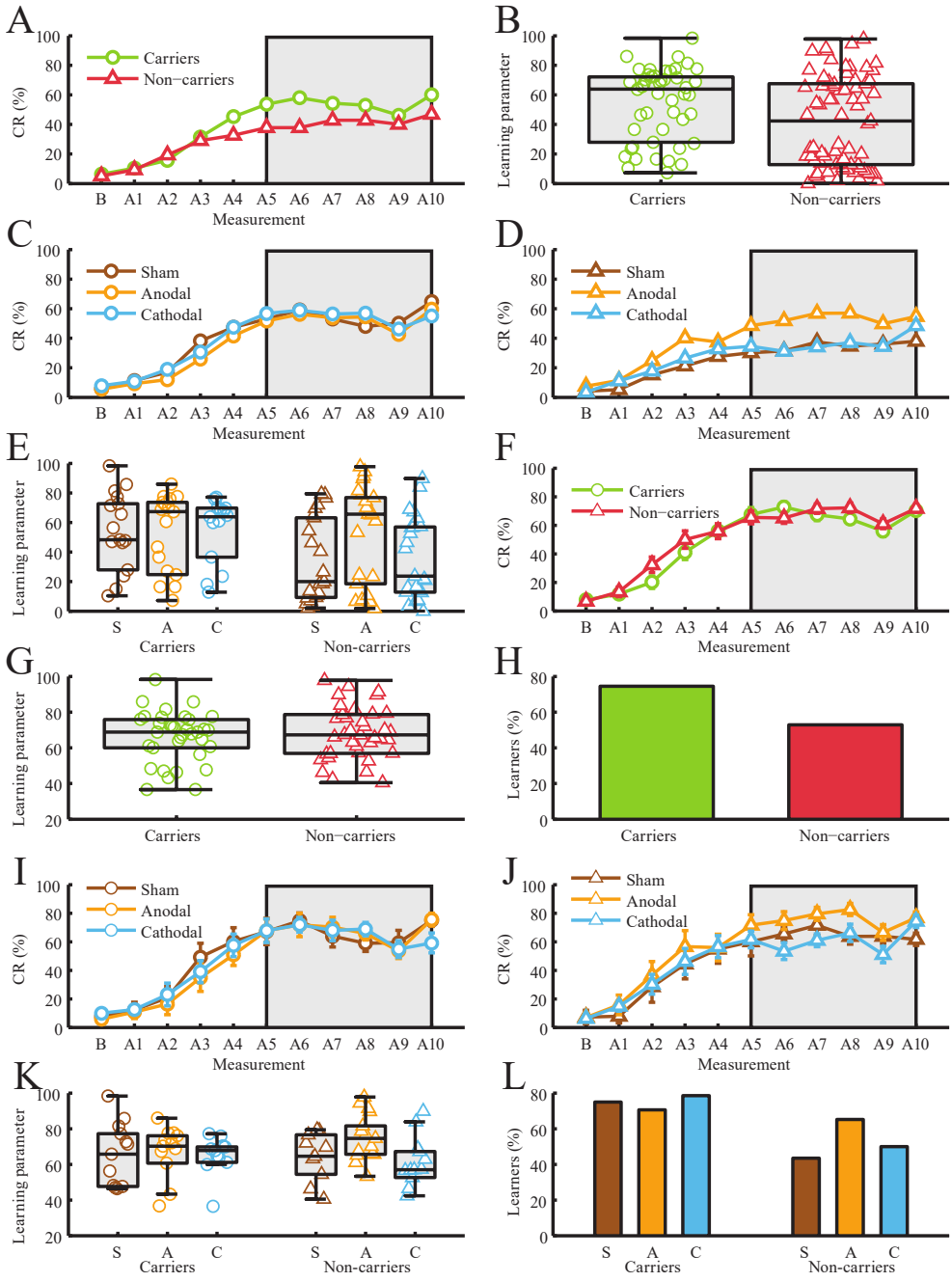


Figure 3. Role of BDNF Val66Met and cerebellar tDCS in eyeblink conditioning. A-E. Overall plots showing learners and non-learners combined. **A.** Overall learning curves for carriers (n=47) and non-carriers (n=70). **B.** Overall whisker plots of the learning parameter for carriers (n=47) and non-carriers (n=70). **C.** Overall learning curves for carriers receiving sham (n=16), anodal (n=17) or cathodal tDCS (n=14). **D.** Overall learning curves for non-carriers receiving sham (n=23), anodal (n=23) or cathodal tDCS (n=24). **E.** Overall

whisker plots of the learning parameter for carriers and non-carriers receiving sham, anodal or cathodal tDCS. **F-L.** Plots showing learners only. **F.** Learning curves for carriers (n=35) and non-carriers (n=37) who were classified as learners. **G.** Whisker plots of the learning parameter for carriers and non-carriers who were classified as learners. **H.** Bar graphs of the proportion of learners in the carrier and non-carrier group. **I.** Learning curves for carriers who were classified as learners and received sham (n=12), anodal (n=12) or cathodal tDCS (n=11). **J.** Learning curves for non-carriers who were classified as learners and received sham (n=10), anodal (n=15) or cathodal tDCS (n=12). **K.** Whisker plots of the learning parameter for carriers and non-carriers receiving sham, anodal or cathodal tDCS who were classified as learners. **L.** Bar graphs of the proportion of learners for carriers and non-carriers receiving sham, anodal or cathodal tDCS. Carriers are displayed in green, non-carriers in red. Sham tDCS is shown in brown, anodal tDCS in orange and cathodal tDCS in blue. Circles indicate carriers. Triangles indicate non-carriers. Error bars represent the standard error of the mean. S = sham; A = anodal; C = cathodal.

The additional saccade adaptation and visuomotor adaptation experiments were analyzed similarly to the eyeblink conditioning and VOR adaptation experiments with "carrier" as the independent variable.

Results for the linear and logistic regressions are reported as the median regression coefficient with 95% equally-tailed intervals (ETIs). Results for the direct comparison of beta distributions are presented as the median difference between carriers and non-carriers with 95%ETIs. An effect size is considered significant if the ETI does not overlap with zero. All analyses were performed using three chains with 50,000 samples each and 20,000 burn-in samples in Openbugs version 3.2.3 (Openbugs foundation). Missing values are automatically handled by the Bayesian analysis and do not contribute to the posterior estimates of the model parameters [77]. Group results are described with medians and interquartile ranges.

Sample size calculation

We powered the eyeblink conditioning and VOR adaptation studies to find a positive effect of anodal cerebellar tDCS in the smaller non-carrier group (estimated as 30% of the population [18,19]). Based on [21,22], BDNF Val66Met carriers were predicted to learn 50% less than non-carriers. All power analyses included a drop-out rate of 10%. For eyeblink conditioning, tDCS effect sizes were based on [1] ($B_{Anodal,carrier=}30\%$, $B_{Cathodal,carrier=}30\%$; population standard deviation of 20%). We estimated 35 subjects per group would give >90% power and included 40 subjects per group. For VOR adaptation, tDCS effect size was based on [2,61] ($B_{Anodal,carrier=}15\%$, within-subject standard deviation of 15%). We estimated a group size of 50 subjects would give >90% power and included 55 subjects. The additional saccade and reaching adaptation experiments were powered to find a $B_{Carrier}$ =10% given a population standard deviation of 10% and included 75 subjects.

Results

Eyeblink conditioning

We found that eyeblink conditioning was best captured with a bimodal distribution of the learning parameter (see Figure 2A and Table 2), which is line with a recent study [78]. The main statistical analysis was therefore based on the 'Learner / non-learner model'.

Results of the single group analysis are presented in Table 3 ('Single group model') and Figures 3A-E. We found (1) an increase in the learning parameter for carriers compared to non-carriers (see Figure 3A-B and Table 3), (2) no effect of cerebellar tDCS on the learning rate for carriers (see Figure 3C,E and Table 3) and (3) an increase in the learning parameter with anodal stimulation for non-carriers (see Figure 3D-E and Table 3).

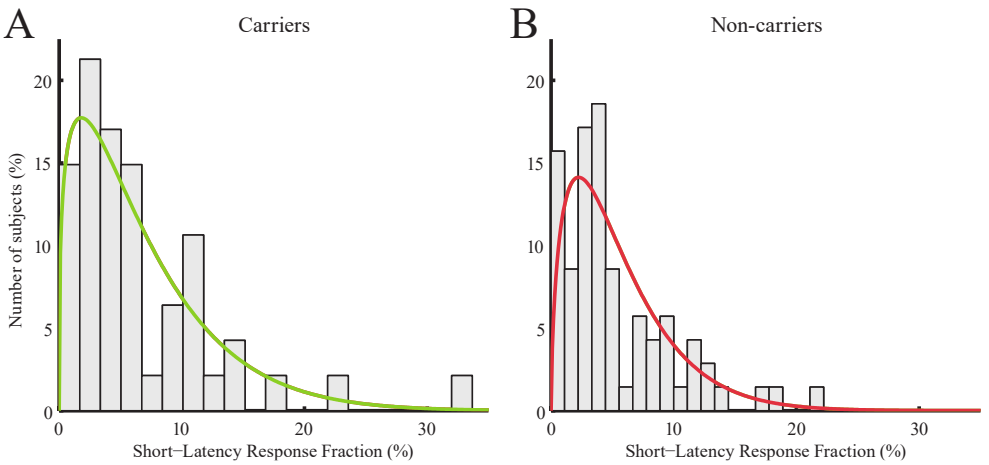


Figure 4. Role of BDNF Val66Met in the short-latency response fraction. A-B. Histograms for the short-latency response fraction in carriers (A) and non-carriers (B).

Paradigm	Model	Factor	OR	B
Eyeblink conditioning	Single group	Carrier	-	18.8 [2.3 35.3]
		$Anodal_{Carrier}$	-	-0.7 [-18.6 17.2]
		$Anodal_{Non\text{-}carrier}$	-	18.0 [2.6 33.3]
		$Cathodal_{Carrier}$	-	1.2 [-17.7 19.9]
		$Cathodal_{Non\text{-}carrier}$	-	2.3 [-13.0 17.5]
	Learner / non-learner	Carrier	4.2 [1.1 19.8]	2.9 [-8.5 14.5]
		$Anodal_{Carrier}$	0.8 [0.1 3.9]	1.0 [-10.2 12.1]
		$Anodal_{Non\text{-}carrier}$	2.5 [0.8 8.9]	11.9 [0.8 23.0]
		$Cathodal_{Carrier}$	1.2 [0.2 8.3]	-1.4 [-12.8 10.0]
		$Cathodal_{Non\text{-}carrier}$	1.3 [0.4 4.3]	-1.7 [-13.2 9.9]
VOR adaptation	Single group	Carrier	-	-2.0 [-8.7 4.7]
		$Anodal_{Carrier}$	-	3.4 [-3.2 9.5]
		$Anodal_{Non\text{-}carrier}$	-	0.6 [-3.4 4.8]
Saccade adaptation	Single group	Carrier	-	1.0 [-8.6 10.6]
Visuomotor adaptation	Single group	Carrier	-	2.7 [-2.5 7.9]

Table 3. Linear and logistic regression models. VOR adaptation, saccade adaptation and visuomotor adaptation were best modeled as a single group (see Table 2) and therefore further analyzed with a linear regression of the learning parameter of all subjects. Eyeblink conditioning was best captured with a learner / non-learner model and therefore analyzed both with a logistic regression for the probability of being a learner and a linear regression for the learning parameter of the learner group. B = correlation coefficients, OR = odds ratio.

Results of the learner / non-learner analysis are presented in Table 3 ('Learner / non-learner model') and Figures 3F-L. We found that whereas the learning parameter was similar for carriers and non-carriers (see Figure 3F-G and Table 3), the percentage of learners was higher for carriers than for non-carriers (see Figure 3H and Table 3). In the carrier group, neither anodal tDCS nor cathodal tDCS affected the learning parameter compared with sham (see Figures 3G and 3K and Table 3). Similarly, neither anodal tDCS nor cathodal tDCS affected the percentage of learners compared with sham (see Figure 3L). In the non-carrier group, anodal tDCS increased the learning parameter (see Figures 3J-K and Table 3) compared with sham, but not the percentage of learners compared with sham (see Figure 3L and Table 3). Cathodal tDCS did not affect the learning parameter nor the percentage of learners (see Figures 3J-L and Table 3).

here was no significant difference in the short latency response fraction between non carriers and carriers ($M_{Non\text{-}carrier}$ - $M_{Carrier}$ = -1.2% 95%ETI =[-3.3 – 0.6]%) (see Figure 4).

VOR adaptation

The learning parameter for VOR adaptation was best described by a unimodal distribution (see Figure 2B and Table 2). We therefore performed statistical analysis based on the 'single group' model (see Figure 5 and Table 3).

The learning parameter was similar for carriers and non-carriers (see Figures 5A and 5C and Table 3). In carriers, no effect of anodal tDCS was found compared with sham (see Figures 5B and 5E and Table 3). Similarly, in non-carriers, no effect of anodal tDCS was found compared with sham (see Figures 5C and 5F and Table 3).

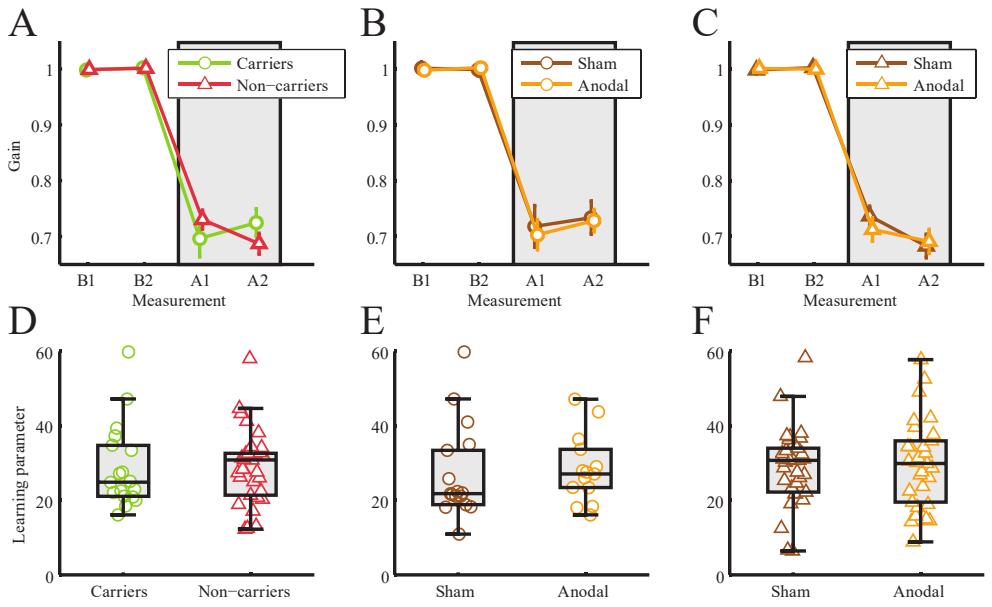


Figure 5. Role of BDNF Val66Met and cerebellar tDCS in VOR adaptation. A-C. Learning curves for (A) carriers and non-carriers, averaged over the two tDCS conditions, (B) carriers receiving sham and anodal tDCS and (C) non-carriers receiving sham and anodal tDCS. **D-F.** Learning parameters for (D) carriers and non-carriers, averaged over the two tDCS conditions, (E) carriers receiving sham and anodal tDCS and (F) non-carriers receiving sham and anodal tDCS.

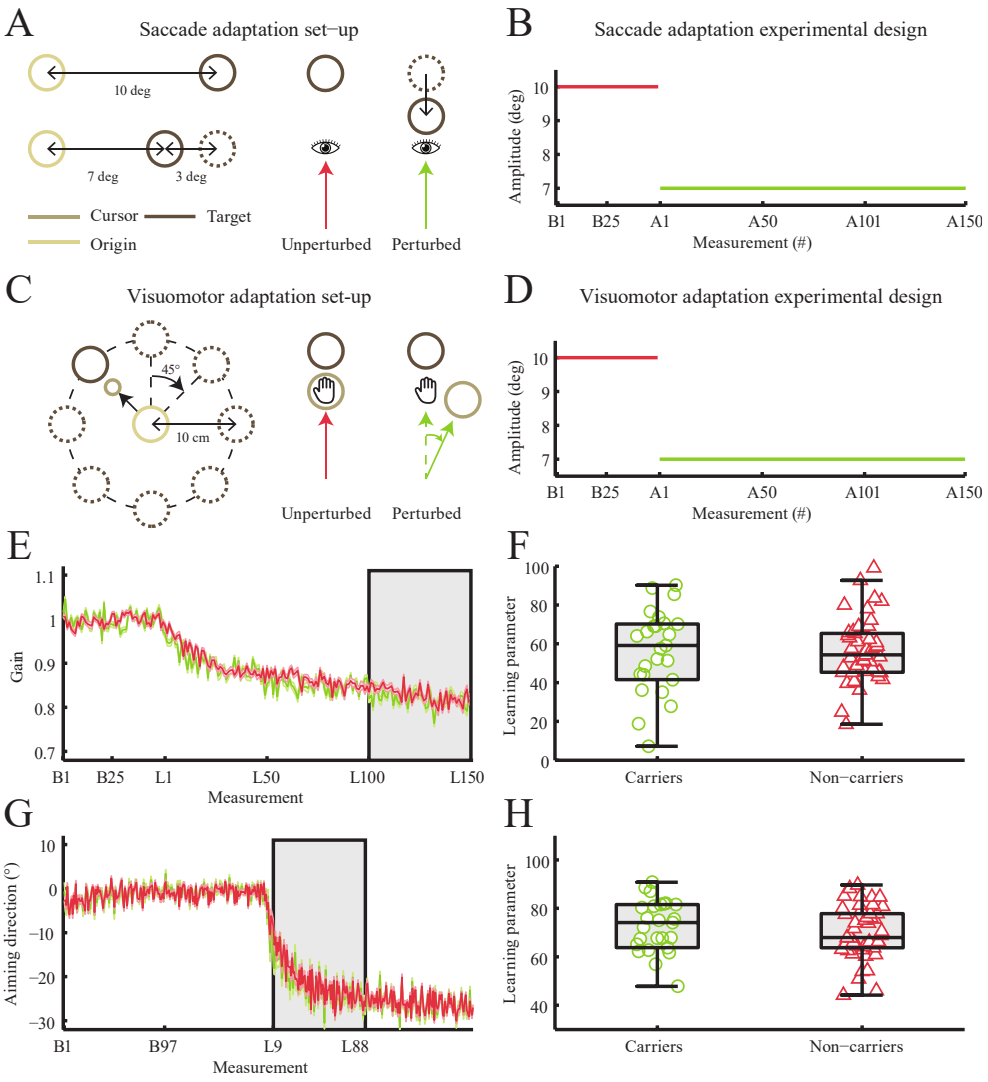


Figure 6. Role of BDNF Val66Met in saccade and visuomotor adaptation. A. Saccade adaptation set-up. Subjects were instructed to look at the origin (red circle 0.25 degrees of visual angle radius) displayed on a black background, 5 degrees of visual angle left of the center of the screen. After a uniformly distributed random delay between 400 and 1400 milliseconds, the origin disappeared and a target (red circle 0.25 degrees of visual angle radius) appeared 5 degrees of visual angle right of the center of the screen. Saccades were detected online using a velocity threshold of 60°/s and a boundary threshold of 1.2° to the right of the fixation position. If no saccade was detected after 500 ms, the screen was blanked for 500 ms and the trial was restarted showing the origin. The duration of one trial was 3000 ms. In unperturbed trials (red line), the target was shown at a fixed location 10° to the right of the origin from presentation start until trial end. In perturbed trials (blue line), the target was displaced 3 degrees of visual angle inward as soon as a saccade was detected i.e., during the saccade. **B.** Saccade adaptation experimental design. The experiment included baseline measurements of 50 unperturbed trials (red line, B1-50), followed by learning measurements of 150 perturbed trials (green line, L1-150). **C.** Visuomotor adaptation set-up. Subjects were instructed to make

straight rapid shooting movements from the origin towards the target. A trial started when the cursor (position of robotic handle; green circle 2 mm radius) was within 0.5 cm of the origin (red circle 2 mm radius) for 1 second, with the appearance of the target (red circle 2 mm radius) at one of 8 positions. A trial ended when the robotic handle passed an (invisible) circle with 10 cm radius around the origin or trial duration exceeded 2 seconds. At this point, the cursor was shown at its last position until the start of the next trial and the movement was dampened. Color cues were given to keep movement velocity in a tight range (blue when too slow >600 ms; yellow when too fast <400 ms; green when correct 400-600 ms). The cursor reappeared at its measured position when located 0.5 cm from the origin. In unperturbed trials (red lines), the cursor was shown at the location of the robotic handle while in perturbed trials (green lines) cursor position was rotated 30 degrees clockwise around the origin with respect to manipulandum position. **D.** Visuomotor adaptation experimental design. The experiment design included baseline measurements of unperturbed trials (red line, B1-192) and learning measurements of perturbed trials (green line, L1-200). **E-F.** Role of BDNF Val66Met in saccade adaptation. Learning curves (left column) and learning parameters (right column) for carriers of the BDNF Val66Met polymorphism (green) and non-carriers (red). **G-H.** Role of BDNF Val66Met in visuomotor adaptation. Learning curves (left column) and learning parameters (right column)

Saccade adaptation and visuomotor adaptation

To further investigate whether a role for BDNF Val66Met is absent in cerebellum-dependent motor adaptation, we performed additional saccade and visuomotor adaptation tasks in 75 subjects. Genetic analysis failed in 3/75 individuals leaving 72 for analysis.

The learning parameters of saccade and visuomotor adaptation were best described by a unimodal distribution (see Figures 2C-D and Table 2) and therefore analyzed with the 'single group' model only. For saccade adaptation, no difference was found for the learning parameter between carriers and non-carriers (see Figures 6 E-F and Table 3). Similarly, for visuomotor adaptation, no difference was found for the learning parameter between carriers and non-carriers (see Figures 6 G-H and Table 3).

Discussion

Role of BDNF Val66Met in cerebellum-dependent learning

The higher proportion of eyeblink conditioning learners in carriers compared to non-carriers could depend on modulation of cerebellar activity. Within the cerebellum, BDNF released from mossy fibers [24] may control the response of both granule cells and Purkinje cells to GABA [79] and thereby keep baseline simple spike firing frequency and the potential for conditioning within normal limits. Carriers of the BDNF Val66Met polymorphism on the other hand are expected to have an altered granule and Purkinje cell response to GABA, which may increase baseline simple spike firing frequency allowing for stronger disinhibition of cerebellar nuclei neurons and faster eyeblink conditioning [80]. Why then does the polymorphism not affect adaptation? Learning mechanisms for gain-decrease VOR adaptation, gain-decrease saccade adaptation and visuomotor adaptation are believed to depend more on synaptic plasticity in cerebellar and vestibular nuclei rather than the cerebellar cortex, and might be less directly related to baseline simple spike firing frequencies [44–48,81,82].

Alternatively, BDNF Val66Met might also influence other brain regions that are involved in eyeblink conditioning, like the amgydala [30–34] and the hippocampus [30]. We did not find a difference in short latency responses between carriers and non-carriers, which makes a direct effect of the amygdala unlikely [31]. However, it has been suggested that the amygdala can enhance eyeblink conditioning indirectly, by modulating the saliency of the conditioned stimulus [34]. In contrast, the hippocampus is believed to inhibit eyeblink conditioning. [30]. Indeed, lower BDNF concentrations in the mouse hippocampus have been associated with faster eyeblink

conditioning [83]. Furthermore, BDNF Val66Met carriers show stronger cued fear conditioning, with decreased activity in the hippocampus and increased activity in the amygdala [25]. This extracerebellar hypothesis is also compatible with the null effect of BDNF Val66Met in the adaptation tasks, which do not depend on the hippocampus or amygdala [41,84].
The relevance of BDNF Val66Met for eyeblink conditioning might extend to other cerebellum-dependent modalities of motor, emotional and cognitive associative learning [85] and pathologies of cerebellum-dependent associative learning such as schizophrenia [86,87].

Mechanisms of cerebellar tDCS
The interaction between cerebellar tDCS and BDNF Val66Met in eyeblink conditioning might point to a common effect on simple spike firing frequency. Anodal tDCS only increases eyeblink conditioning in non-carriers, who learn more slowly and have higher activity-dependent BDNF release. However, it seems unlikely that the effect of anodal tDCS in the cerebellum is mediated by BDNF release, as has been suggested for the motor cortex [22], because this would decrease rather than increase the eyeblink conditioning response. Rather, we expect anodal tDCS to directly modulate the baseline simple spike firing frequency of cerebellar neurons through subthreshold depolarization [36,37,39,40,88,89]. Carriers might be less sensitive to subthreshold depolarization, because baseline firing frequency is already increased (1) as a direct result of diminished BDNF release or (2) as a result of stronger excitation by the amygdala. In contrast, no effect of cerebellar tDCS on VOR adaptation was found for either carriers or non-carriers, which might be related to a minor role for simple spike firing in VOR adaptation compared to eyeblink conditioning [46,48]. Alternatively, the cerebellar flocculus, which is involved in VOR adaptation is located deeper in the cerebellum than Lobule VI, which is involved in eyeblink conditioning, and the local electric field strength [51] might therefore be insufficient for cerebellar tDCS to have an effect. Modeling-based approaches are necessary to further explore this open question [12].

The complex interaction between (1) cerebellar tDCS, (2) anatomical substrates and neurophysiological mechanisms of motor learning, and (3) genetic factors requires detailed animal studies combining electrophysiological and behavioral experiments to further develop cerebellar tDCS as a neuromodulatory technique.

Variable results of cerebellar tDCS
The interaction between BDNF Val66Met and anodal tDCS might explain some of the inconsistency in cerebellar tDCS literature. The null result for anodal tDCS found by Beyer et al. compared with increased eyeblink conditioning found by Zuchowski et al. [1] might have resulted from a higher proportion of carriers in the subject population of Beyer et al. [14]. However, decreased eyeblink conditioning with cathodal tDCS [1] could only be explained from our results by an uneven distribution of non-learners. In addition, since no interaction between cerebellar tDCS and BDNF Val66Met in VOR adaptation was found, as well as no direct effect of BDNF Val66Met on VOR adaptation, saccade adaptation and visuomotor adaptation, we do not think conflicting literature results in other tasks, such as visuomotor adaptation [2,4,10,13], can be explained by our results. Possibly, other individual determinants are important in these tasks.
Careful characterization of genetic and other individual factors will be necessary in future (pre)clinical studies of cerebellar tDCS to decrease response variability and identify non-learners who do not benefit from stimulation.

References

1. Zuchowski, M.L., Timmann, D. & Gerwig, M. *Brain Stimul.* 7, 525–31 (2014).
2. Galea, J.M., Vazquez, A., Pasricha, N., de Xivry, J.J. & Celnik, P. *Cereb Cortex* 21, 1761–1770 (2011).
3. Jayaram, G. et al. *J. Neurophysiol.* 107, 2950–7 (2012).
4. Block, H. & Celnik, P. *Cerebellum* 12, 781–93 (2013).
5. Herzfeld, D.J. et al. *Neuroimage* 98, 147–58 (2014).
6. Avila, E. et al. *Neural Plast.* 2015, 1–9 (2015).
7. Shah, B., Nguyen, T.T. & Madhavan, S. *Brain Stimul.* 6, 966–8 (2013).
8. Panouilleres, M.T.T.N., Miall, R.C. & Jenkinson, N. *J. Neurosci.* 35, 5471–5479 (2015).
9. Das, S. et al. *Brain Stimul.* 7, e3 (2014).
10. Hardwick, R.M. & Celnik, P.A. *Neurobiol. Aging* 35, 2217–21 (2014).
11. van Dun, K., Bodranghien, F.C.A.A., Mariën, P. & Manto, M.U. *Front. Hum. Neurosci.* 10, 199 (2016).
12. Ferrucci, R., Bocci, T., Cortese, F., Ruggiero, F. & Priori, A. *Cerebellum & ataxias* 3, 16 (2016).
13. Jalali, R., Miall, R.C. & Galea, J.M. *J. Neurophysiol.* jn.00896.2016 (2017).doi:10.1152/jn.00896.2016
14. Beyer, L., Batsikadze, G., Timmann, D. & Gerwig, M. *Front. Hum. Neurosci.* 11, 23 (2017).
15. Hulst, T. et al. *J. Neurophysiol.* 118, 732–748 (2017).
16. Furuya, S., Klaus, M., Nitsche, M.A., Paulus, W. & Altenmuller, E. *J. Neurosci.* 34, 13834–13839 (2014).
17. Buch, E.R. et al. *Clin. Neurophysiol.* 128, 589–603 (2017).
18. Cargill, M. et al. *Nat. Genet.* 22, 231–8 (1999).
19. Shimizu, E., Hashimoto, K. & Iyo, M. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 126B, 122–3 (2004).
20. Egan, M.F. et al. *Cell* 112, 257–269 (2003).
21. McHughen, S.A. et al. *Cereb. Cortex* 20, 1254–62 (2010).
22. Fritsch, B. et al. *Neuron* 66, 198–204 (2010).
23. Park, H. & Poo, M. *Nat. Rev. Neurosci.* 14, 7–23 (2013).
24. Chen, A.I. et al. *Sci. Rep.* 6, 20201 (2016).
25. Soliman, F. et al. *Science (80-.).* 327, 863–866 (2010).
26. Yeo, C.H., Hardiman, M.J. & Glickstein, M. *Exp. brain Res.* 60, 87–98 (1985).
27. Yeo, C.H., Hardiman, M.J. & Glickstein, M. *Exp. brain Res.* 60, 99–113 (1985).
28. Jirenhed, D.-A., Bengtsson, F. & Hesslow, G. *J. Neurosci.* 27, 2493–502 (2007).
29. McCormick, D.A. et al. *Bull. Psychon. Soc.* 18, 103–105 (1981).
30. Lee, T. & Kim, J.J. *J. Neurosci.* 24, (2004).
31. Boele, H.J., Koekkoek, S.K.E. & De Zeeuw, C.I. *Front. Cell. Neurosci.* 3, 19 (2010).
32. Siegel, J.J. et al. *eNeuro* 2, (2015).
33. Sakamoto, T. & Endo, S. *Eur. J. Neurosci.* 32, 1537–1551 (2010).
34. Farley, S.J., Radley, J.J. & Freeman, J.H. *J. Neurosci.* 36, 2190–2201 (2016).
35. De Zeeuw, C.I. & Ten Brinke, M.M. *Learn. Mem.* 389 (2016).
36. ten Brinke, M.M. et al. *Cell Rep.* 13, 1977–1988 (2015).
37. Chan, C.Y., Hounsgaard, J. & Nicholson, C. *J. Physiol.* 402, 751–71 (1988).
38. Bikson, M. et al. *J. Physiol.* 557, 175–190 (2004).
39. Chan, C.Y. & Nicholson, C. *J. Physiol.* 371, 89–114 (1986).
40. Reato, D., Rahman, A., Bikson, M. & Parra, L.C. *J. Neurosci.* 30, 15067–79 (2010).
41. Zee, R.J.L.D.S., Leigh, R.J. & Zee, D.S. (Oxford University Press: New York, 2006).doi:http://10.0.4.69/med/9780199969289.001.0001
42. Blazquez, P.M., Hirata, Y., Heiney, S.A., Green, A.M. & Highstein, S.M. *J. Neurosci.* 23, 9742–9751 (2003).
43. Hirata, Y. & Highstein, S.M. *J. Neurophysiol.* 85, 2267–88 (2001).
44. Lisberger, S.G., Pavelko, T.A., Bronte-Stewart, H.M. & Stone, L.S. *J. Neurophysiol.* 72, 954–73 (1994).
45. Lisberger, S.G. *J. Neurophysiol.* 72, 974–98 (1994).
46. Carcaud, J. et al. *eNeuro* 4, (2017).
47. Lisberger, S.G., Pavelko, T.A. & Broussard, D.M. *J. Neurophysiol.* 72, 928–53 (1994).
48. Voges, K., Wu, B., Post, L., Schonewille, M. & De Zeeuw, C.I. *J. Physiol.* 595, 5301–5326 (2017).
49. Oldfield, R.C. *Neuropsychologia* 9, 97–113 (1971).

50. Galea, J.M., Jayaram, G., Ajagbe, L. & Celnik, P. *J. Neurosci.* 29, 9115–22 (2009).
51. Rampersad, S.M. et al. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 441–52 (2014).
52. van Dun, K., Bodranghien, F., Manto, M. & Mariën, P. *The Cerebellum* (2016).doi:10.1007/s12311-016-0840-7
53. Gandiga, P.C., Hummel, F.C. & Cohen, L.G. *Clin Neurophysiol* 117, 845–850 (2006).
54. van der Vliet, R., Ribbers, G.M., Vandermeeren, Y., Frens, M.A. & Selles, R.W. *Brain Stimul.* 10, 882–892 (2017).
55. Cason, H. *J. Exp. Psychol.* 5, 153–196 (1922).
56. Smit, A.E. et al. *Genes. Brain. Behav.* 7, 770–7 (2008).
57. Christian, K.M. & Thompson, R.F. *Learn. Mem.* 10, 427–55 (2003).
58. Koekkoek, S.K.E., Den Ouden, W.L., Perry, G., Highstein, S.M. & De Zeeuw, C.I. *J. Neurophysiol.* 88, 2124–2133 (2002).
59. Tiliket, C., Shelhamer, M., Roberts, D. & Zee, D.S. *Exp. brain Res.* 100, 316–27 (1994).
60. Shelhamer, M., Tiliket, C., Roberts, D., Kramer, P.D. & Zee, D.S. *Exp. brain Res.* 100, 328–36 (1994).
61. Montfoort, I., Van Der Geest, J.N., Slijper, H.P., De Zeeuw, C.I. & Frens, M.A. *J. Neurotrauma* 25, 687–93 (2008).
62. Schubert, M.C., Migliaccio, A.A., Minor, L.B. & Clendaniel, R.A. *Exp. brain Res.* 187, 117–27 (2008).
63. Yakushin, S.B., Palla, A., Haslwanter, T., Bockisch, C.J. & Straumann, D. *Exp. brain Res.* 152, 137–42 (2003).
64. Yakushin, S.B., Bukharina, S.E., Raphan, T., Buttner-Ennever, J. & Cohen, B. *Ann. N. Y. Acad. Sci.* 1004, 78–93 (2003).
65. Monte-Silva, K. et al. *Brain Stimul.* 6, 424–32 (2013).
66. Lefebvre, S. et al. *Front. Hum. Neurosci.* 6, 343 (2012).
67. Watanabe, S., Hattori, K. & Koizuka, I. *Auris. Nasus. Larynx* 30 Suppl, S29-34 (2003).
68. van der Geest, J.N. & Frens, M.A. *J. Neurosci. Methods* 114, 185–95 (2002).
69. McLaughlin, S.C. *Percept. Psychophys.* 2, 359–362 (1967).
70. Coesmans, M. et al. *J. Psychiatry Neurosci.* 39, E3–E11 (2014).
71. Frens, M.A. & van Opstal, A.J. *Exp. brain Res.* 100, 293–306 (1994).
72. Krakauer, J.W., Pine, Z.M., Ghilardi, M.F. & Ghez, C. *J. Neurosci.* 20, 8916–24 (2000).
73. Tseng, Y.-W.W., Diedrichsen, J., Krakauer, J.W., Shadmehr, R. & Bastian, A.J. *J Neurophysiol* 98, 54–62 (2007).
74. Werner, S. et al. *PLoS One* 10, e0123321 (2015).
75. Rabe, K. et al. *J. Neurophysiol.* 101, (2009).
76. Celeux, G., Forbes, F., Robert, C.P. & Titterington, D.M. *Bayesian Anal.* 1, 651–673 (2006).
77. Kruschke, J.K.. (Academic Press: 2010).
78. Löwgren, K. et al. *PLoS One* 12, e0177849 (2017).
79. Cheng, Q. & Yeh, H.H. *J. Neurosci. Res.* 79, 616–627 (2005).
80. De Zeeuw, C.I. & Ten Brinke, M.M. *Cold Spring Harb. Perspect. Biol.* 7, a021683 (2015).
81. Robinson, F.R., Fuchs, A.F. & Noto, C.T. *Ann. N. Y. Acad. Sci.* 956, 155–63 (2002).
82. Donchin, O. et al. *J. Neurophysiol.* 107, 134–47 (2012).
83. Janke, K.L., Cominski, T.P., Kuzhikandathil, E. V, Servatius, R.J. & Pang, K.C.H. *Front. psychiatry* 6, 106 (2015).
84. Krakauer, J.W. et al. *J. Neurophysiol.* 91, (2004).
85. Timmann, D. et al. *Cortex* 46, 845–857 (2010).
86. Andreasen, N.C. & Pierson, R. *Biol. Psychiatry* 64, 81–88 (2008).
87. Yeganeh-Doost, P., Gruber, O., Falkai, P. & Schmitt, A. *Clinics (Sao Paulo).* 66 Suppl 1, 71–7 (2011).
88. Radman, T., Ramos, R.L., Brumberg, J.C. & Bikson, M. *Brain Stimul.* 2, 215–28, 228.e1–3 (2009).
89. Rahman, A. et al. *J. Physiol.* 591, 2563–78 (2013).

4.3 Cerebellar Cathodal Transcranial Direct Stimulation and performance on a verb generation task: a replication study

Kerstin Spielmann, Rick van der Vliet, W. Mieke E. van de Sandt-Koenderman, Maarten A. Frens, Gerard M. Ribbers, Ruud W. Selles, Stephanie van Vugt, Jos N. van der Geest and Peter Holland

Abstract

The role of the cerebellum in cognitive processing is increasingly recognized, but still poorly understood. A recent study in this field applied cerebellar Transcranial Direct Current Stimulation (c-tDCS) to the right cerebellum to investigate the role of prefrontal-cerebellar loops in language aspects of cognition. Results showed that the improvement in participants' verbal response times on a verb generation task was facilitated by cathodal c-tDCS, compared to anodal or sham c-tDCS. The primary aim of the present study is to replicate these findings and additionally to investigate possible longer term effects. A cross-over within-subject design was used, comparing cathodal and sham c-tDCS. The experiment consisted of two visits with an interval of one week. Our results show no direct contribution of cathodal c-tDCS over the cerebellum to language task performance. However, one week later, the group receiving cathodal c-tDCS in the first visit show less improvement and increased variability in their verbal response times during the second visit, compared to the group receiving sham c-tDCS in the first visit. These findings suggest a potential negative effect of c-tDCS and warrant further investigation into long term effects of c-tDCS before undertaking clinical studies with post-stroke patients with aphasia.

Introduction

Transcranial Direct Current Stimulation (tDCS) has become increasingly popular in neuroscience and neurorehabilitation. This user-friendly noninvasive form of brain stimulation can either increase or reduce neuronal excitability in a polarity-specific manner.[1,2] Positive or anodal stimulation is proposed to increase activity in the brain area under the electrode whereas negative or cathodal stimulation would do the opposite. tDCS has been used for fundamental research to understand the functional organization of the brain and additionally it has been investigated in a clinical setting. Examples of such clinical studies include attempts to treat patients with post-stroke aphasia or hemiplegia, Parkinson's disease, and depression.[3–6] However, despite a large body of tDCS literature reporting positive results, the reproducibility of these results is questioned.[7,8]

Recent studies have applied tDCS to understand the different functional domains of the cerebellum, a brain structure traditionally thought to be solely related to motor control but recently suggested to also be engaged in cognitive processes.[9] A role of the cerebellum in cognitive processing is supported by reports of cognitive deficits following injury to the cerebellum as well as anatomical and neuroimaging studies.[10,11] Topographically, cerebellar lobules VI and VII were found to have projections to cortical association areas involved in cognitive processes.[11] Neuroimaging studies have shown that regions of lobule VII are involved in prefrontal-cerebellar loops.[12–14] Specifically, language processing and executive functioning activated regions of lobule VII.[14] Taken together, these studies demonstrate the role of prefrontal-cerebellar loops in cognitive processing, specifically it has been suggested that the Purkinje cells in the right cerebellum have an inhibitory effect on the contralateral cortical prefrontal regions (i.e. cerebello-cortical inhibition).[9,11–14]

The efficacy of cerebellar tDCS (c-tDCS) in modulating cerebello-cortical inhibition has previously been confirmed by Galea et al.[15] They combined Transcranial Magnetic Stimulation (TMS) with c-tDCS and demonstrated that anodal c-tDCS to the right cerebellum increases the inhibitory effect to the primary motor cortex whilst cathodal c-tDCS to the right cerebellum reduces this effect. As Purkinje cells are the sole inhibitory output of the cerebellum, this observation suggests that anodal c-tDCS leads to increased activity of these neurons while cathodal c-tDCS lead to decreased activity. In addition, electrophysiological animal studies confirmed modulation of Purkinje cell activity with electrical stimulation.[16,17] However, in humans, whether these changes in Purkinje cells firing are direct or depend on other cerebellar neurons is currently unknown. Given the highly homogenous anatomy of the cerebellar cortex, it would seem likely that c-tDCS affects the prefrontal cortex similarly to the motor cortex. This means anodal c-tDCS would decrease prefrontal cortex activity whereas cathodal c-tDCS would increase prefrontal cortex activity. However, literature regarding the efficacy of c-tDCS is inconsistent, for example, a study by Doeltgen et al. reported[18] that anodal c-tDCS may reduce the inhibitory effect on the primary motor cortex. Also, a study focusing on language functioning[19] found that both anodal and cathodal c-tDCS enhanced the performance on a phonemic fluency task.

An interesting recent study that investigated right cerebellar involvement in cognitive processing employed c-tDCS to study prefrontal-cerebellar loops in arithmetic and language aspects of working memory and attention.[20] Pope and Miall[20] hypothesized that cathodal c-tDCS over the right cerebellum lobule VII would reduce the inhibitory tone exerted by the Purkinje cells over prefrontal regions, causing disinhibition of the contralateral prefrontal regions. Disinhibition of prefrontal regions in turn could improve performance, especially on cognitively

demanding tasks. Pope and Miall used arithmetic and language tasks with varying levels of cognitive demand and, reported that the improvement in participants' verbal response times was facilitated by cathodal c-tDCS over the right cerebellum, compared to anodal or sham c-tDCS over the same region. Additionally response times became less variable. As the improvement was greatest for the more cognitively demanding versions of the arithmetic and language task, the authors speculated that the cerebellum is capable of releasing cognitive resources by disinhibition of prefrontal regions, enhancing performance when tasks become cognitively demanding. Further support for this hypothesis was later found by demonstrating that stimulation of the prefrontal cortex with anodal tDCS achieves the same effect as cathodal c-tDCS, specifically for the task assessing arithmetic aspects.[21]

In the present study, we were specifically interested in the potential improvement in language task performance after c-tDCS, as reported by Pope and Miall.[20] Right cerebellar involvement in language processing has been highlighted in several studies.[22–24] Further, a Positron Emission Tomographic (PET) study[25,26] and a Functional Magnetic Resonance Imaging (fMRI) study[27] have demonstrated an involvement of left hemisphere areas and the right cerebellum during a verb generation task. The application of c-tDCS may contribute to our understanding of the prefrontal-cerebellar loops and language processing in healthy subjects, but could also be interesting for future clinical applications.[28] Recent clinical studies applying cerebral tDCS in post-stroke aphasia patients have already shown promising effects[29–31] and c-tDCS might possibly further contribute to the recovery of these patients. However, the results of cerebellar stimulation on language in healthy subjects awaits replication before translation to the clinical setting is justified.

The primary aim of the present study was to replicate the facilitatory effect immediately after cathodal c-tDCS on language task performance, as reported by Pope and Miall (i.e. their experiment 2).[20] The task setup and outcome measures are similar to their study. In contrast to their between-subject design, the present study performed a cross-over within-subject design, comparing cathodal and sham c-tDCS, in order to reduce the impact of individual variability in the response to tDCS.[32] The experiment consisted of two visits with an interval of one week; therefore, this design allowed us to investigate the long term effects of stimulation by measuring the same participants one week later.

Methods

Design

The present study replicates the task used in experiment 2 of the study of Pope and Miall.[20] Their study had a double-blind between-subject design comparing anodal c-tDCS, cathodal c-tDCS and sham c-tDCS (for further details see [20]). The present study has a double-blind cross-over within-subject design, comparing cathodal c-tDCS and sham c-tDCS (see Figure 1). The experiment consisted of two visits with an interval of one week. In each visit a different stimulation condition (cathodal or sham c-tDCS) was applied and this order was counterbalanced among participants. Similar to the study of Pope and Miall, response accuracy and verbal response times were collected before and after cathodal c-tDCS and sham c-tDCS on three language tasks: noun reading, verb reading and verb generation.

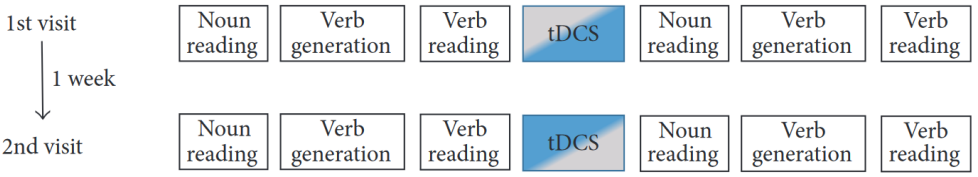


Figure 1. Study design. Participants complete 2 visits with a one-week interval, receiving cathodal (blue) or sham c-tDCS (grey) in a counterbalanced order.

Sample size calculation

Power calculations were based on the reported effects of the study of Pope and Miall,[20] specifically the interaction effect for verbal response times (Group x Block x Task, $F(20,570)=1.83$ corresponding to a Cohen's f of 0.18) and the interaction effect for a computed variable Learning (Session x Task x Group, $F(1,114)=4.50$ corresponding to a Cohen's f of 0.28). For a study design with 4 repeated measurements (cathodal compared to sham; before tDCS compared to after tDCS), a within-patient correlation of 0.75, an alpha of 0.05, a power of 0.80 and a Cohen's f effect size of 0.18, we need 23 subjects. For a study design with 4 repeated measurements (cathodal compared to sham; before tDCS compared to after tDCS), a within-patient correlation of 0.75, an alpha of 0.05, a power of 0.80 and a Cohen's f effect size of 0.28, we need 11 subjects. Based on these power calculations, our aim was to include 24 subjects (in order to have an even number of subjects for the counterbalancing procedure).

Participants

Twenty-four healthy and (near) native Dutch speakers (18 women, 6 men; age range 19-29 years, mean ± SD: 22 ± 2.36 years) with normal vision and normal speech (i.e. no stammer) were recruited from the Erasmus University Rotterdam for a small monetary reward. Exclusion criteria were left handedness and dyslexia. Right-handedness was based on an Edinburgh Handedness Inventory score ≥ 50,[33] and the absence of dyslexia was self-reported. All participants gave informed consent and the study has been approved by the Medical Ethics Committee of the Erasmus MC, University Medical Center Rotterdam.

Tasks and Stimuli

We used the three language tasks that were used in the study of Pope and Miall: a noun reading task, a verb generation task and a verb reading task. For a Dutch version of these tasks, we prepared Dutch word lists including 40 nouns and 40 matched verbs. First, all nouns of the verb generation task used by Pope and Miall[20] were translated. Some of the nouns could not be translated into Dutch and some verb productions were strongly related to the morphological form of the item due to an identical wordstem (e.g. *fiets- fietsen*, meaning 'bike- biking'). The list of nouns was therefore supplemented by the set of Dutch nouns of De Witte et al.,[34] resulting in a list of 124 concrete nouns related to manipulable tools and objects that were potential stimuli for the language experiment. The stimuli of the final word list were chosen on the basis of responses in a verb generation task from a pilot group (n = 22). Only noun-verb pairs generated by more than half of the pilot group were selected for the final word list. If two or more nouns elicited the same verb, these nouns were excluded. Also nouns eliciting non-action verbs (e.g. 'oven-bake') were excluded. The final word list, including 40 nouns and 40 matched verbs, was split up in two lists (list A and list B): one list was presented before c-tDCS and the other after c-tDCS. The order of list A and B was counterbalanced across participants. Specifically, during the first visit, half of

the group was presented with list A before c-tDCS and list B after c-tDCS. During the second visit, this same group was presented with list B before c-tDCS and list A after c-tDCS. The other half of the group received the opposite order, starting during the first visit with list B before c-tDCS, etc. The stimuli were presented on a computer screen (48 cm x 28 cm) placed 65 cm in front of the participants. The tasks were designed and presented using Matlab 2013a and Psychophysics Toolbox (v3.0.12).[35,36] Each task comprised 6 blocks of 10 trials (i.e. 10 words) each. In the first five blocks the same set of words was used but the order of the appearance of the words was randomized on a block by block basis. In the sixth block a new set of words was presented, again in a randomized order. Each task lasted approximately 5 minutes. Participants had a break of at least 10 seconds between each task.

A microphone (model: Trust-MC 1200) was used to register the verbal response times. Each stimulus was replaced by the next stimulus when the microphone recorded a response. After a response was recorded, a black screen was displayed for 2 s before the next stimulus was presented.

Transcranial Direct Current Stimulation

Cathodal and sham c-tDCS were delivered through a pair of saline-soaked sponge electrodes (25 cm^2 surface area) using a NeuroConn DC-stimulator. In the cathodal stimulation condition participants received active stimulation of 2 mA for a duration of 20 minutes. Stimulation was automatically activated with a fade in of 30 s and after 20 minutes the stimulation was automatically deactivated with a fade out of 30 s. In the sham condition, participants received pseudo stimulation with a fade in of 30 s and after 40 s the stimulation was automatically deactivated with a fade out of 30 s. The average impedance was 23.7 ± 8.0 kΩ (mean ± SD) among participants. The cathode was placed over the right cerebellar cortex, 1 cm under and 4 cm lateral to the inion, which is defined as the location of the cerebellar lobule VII. The anode was placed over the right shoulder, that is, the right deltoid muscle.[20]

Procedure

The experiment was performed inside a quiet cubicle. Participants performed the three tasks in the following order: noun reading, verb generation and verb reading. For the reading tasks, participants were instructed to read the presented noun or verb aloud as soon as it appeared on the computer screen. For the verb generation task, they were instructed to produce an appropriate verb as quickly as possible in response to the noun presented on the screen. It was explained that an appropriate verb could be a verb that described what the presented noun may do or what it may be used for. It was emphasized that only one verb was to be produced. At the beginning of each task, one example was given and three test items were presented, which were items other than those in the experiment. For all tasks, responses were checked for accuracy by the researcher. All verbs produced during the verb generation task were written down by the researcher.

After completion of the three tasks, 20 minutes of cathodal or sham c-tDCS was applied. The electrodes were placed by the researcher. Both the researcher and the participant were blinded for stimulation condition, which was achieved by using two 5-number codes that can be entered into the tDCS device. These 5-number codes are provided by the manufacturer of the tDCS device. One code is related to start the real tDCS stimulation condition and the other code is related to start sham tDCS. A researcher of our research team (JG), who was not involved in the assessment of the experiment, provided these two 5-number codes. During the 20 minutes

cathodal or sham c-tDCS, participants were instructed to look at a black computer screen. After the stimulation, participants performed the three tasks for the second time using parallel versions of word lists. In total, the experiment lasted approximately 90 minutes. After one week each participant took the experiment for the second time, in which the other stimulation condition was applied. Next to that, the word list previously presented after c-tDCS were now presented prior to c-tDCS.

Statistical analysis

Incorrect responses, missed responses, and outliers were removed before analysis. For the noun reading and the verb reading tasks, no incorrect responses were detected. For the verb generation task, non-words, multiple word responses and responses that were not

Variable	Effect	df	F	p	η2
Verbal response time	Condition	1, 23	4.81	0.038	0.173
	Task	1.16, 26.71	808.98	<0.001	0.9772
	Block	5, 115	121.63	<0.001	0.841
	Task x Block	4.22, 97.15	37.16	<0.001	0.618
	Session	1, 23	0.10	0.750	0.004
	Task x Session	1.38, 1.20	0.77	0.427	0.032
	Condition x Task x Block	4.33, 99,63	0.77	0.558	0.032
Response variability	Session	1, 23	6.49	0.018	0.220
	Task	1.19, 27.37	655.93	<0.001	0.966
	Block	5, 115	17,63	<0.001	0.434
	Task x Block	4.31, 99.12	8.65	<0.001	0.273
	Condition x Block	5, 115	0.62	0.689	0.026
	Condition x Task x Block	4.00, 91.96	1.42	0.233	0.058
Learning	Task	1.20, 27.52	21.76	<0.001	0.486
	Task x Session	1.22, 27.96	0.47	0.537	0.020
	Task x Condition	1.18, 27.11	1.48	0.240	0.060
	Session x Condition	1, 23	0.36	0.555	0.015
	Session x Task x Condition	1.27, 29.10	0.35	0.608	0.015
Learning variability	Session	1, 23	5.45	0.029	0.192
	Task	1.09, 25.00	6.66	0.014	0.225
	Condition	1, 23	0.63	0.435	0.027
	Task x Session	1.24, 28.44	7.09	0.009	0.236
	Task x Condition	1.17, 26.84	0.34	0.600	0.014
	Session x Condition	1, 23	0.70	0.411	0.030
	Session x Task x Condition	1.06, 24.34	0.44	0.524	0.019

Table 1. Results of the study: verbal response time, response variability, learning and learning variability.

representative for what the noun may do or what it may be used for (e.g. 'eyebrow – drawing'), were considered incorrect and were not included in the analysis. For each task, voice onset times were corrected manually from digital recordings if lip movement, swallowing and heavy breathing were prior to the verbal response, because this influenced the microphone recording. Outliers, responses exceeding more or less than 2 standard deviations from the mean of that task were removed.

Although we used test items, a novelty effect was found for the first trials (i.e. first word presented) of each block, shown by a larger reaction time. Because the mean for each block consisting of 10 trials was calculated, we decided to exclude the first trial in order to get a representative mean of the data. Further, in case of violations of sphericity, a Greenhouse-Geisser correction was applied and adjusted degrees of freedom are reported in the text.

In line with the study of Pope and Miall, the present study analyzed the data in terms of the mean and variability of verbal response times. Mean verbal response times for each block per task were analyzed with a repeated measures analysis of variance (ANOVA), using four factors. These factors are Condition (cathodal tDCS and sham), Session (pre-tDCS and post-tDCS), Task (noun reading, verb generation and verb reading) and Block (six blocks per task). The variability of verbal response times between the three tasks and six blocks per task was analyzed with pairwise comparisons; a Bonferroni correction was used. The level of significance was set at α = 0.05. For the response variability, an ANOVA was performed on the within block standard deviations of the verbal response times across Block, Task, Session and averaged by Condition.
Also in line with the study of Pope and Miall, the present study analyzed the data by computing the variables 'learning' and 'total learning variability'. The learning variable was computed by subtracting Block 5 from Block 1 and putting this as a variable in an ANOVA with Task x Session x Condition. For the total learning variability, the standard deviations of the learning variable (Block 5 – Block 1) across Task, Session and averaged by Condition, were entered into an ANOVA.

The present within-subject design allows us to investigate the long term effects of stimulation by measuring the same subjects a week later. We therefore also performed an ANOVA including the between-subject factor visit-order. This between-subject factor indicates whether a participant received cathodal c-tDCS or sham c-tDCS at the first visit.

Results

In general, results are reported in the same way as in the study of Pope and Miall.[20] Table 1 presents an overview of the statistical results for the 4 variables that were analyzed: mean verbal response times, verbal response variability, learning and total learning variability. Table 1 only includes the factors and interactions that were reported as (near) significant in the study of Pope and Miall, and will be explained further in the following paragraphs. Values are reported as mean ± standard error of the mean in the text unless otherwise specified.

Response accuracy

Participants made very few incorrect responses (1.9%) and very few missed responses (0.5%) were obtained. These incorrect and missed responses were excluded from further analysis.

Verbal response times

Figure 2 presents the results of the verbal response times for each task and across the 6 blocks, before and after tDCS. In general, the range of verbal response times of the present study (0.573 s – 1.082 s) was higher than the study of Pope and Miall.[20] A Condition x Task x Session x Block

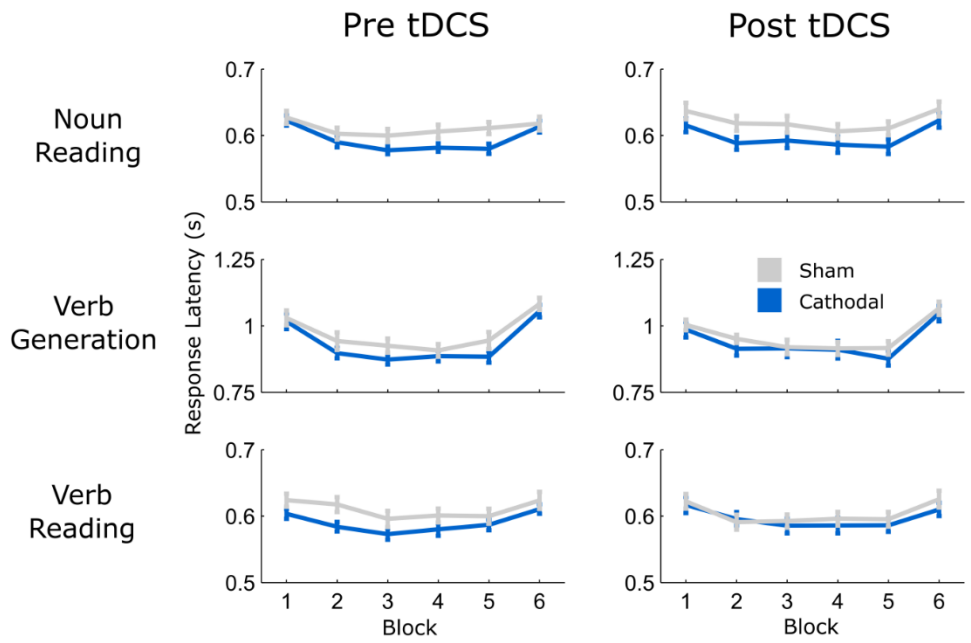


Figure 2. Results for the verbal response times (s), before and after tDCS, for each task and across the 6 blocks. Error bars present the Standard Error of the Mean (SEM).

ANOVA revealed a large main effect (see Table 1) of Condition, with larger verbal response times in the sham condition (0.730 ± 0.011 s) compared to the cathodal condition (0.709 ± 0.010 s). However, there was no main effect of Session and no interaction effect of Condition x Session, therefore indicating no overall effect of tDCS on verbal response times.

In line with the study of Pope and Miall, a large main effect of Task was found, with larger verbal response times on the verb generation task (0.953 ± 0.016 s) compared to the noun reading (0.606 ± 0.007 s) and verb reading task (0.600 ± 0.008 s). Also in line with Pope and Miall, a large main effect of Block was found. This can be described as a priming effect for block 1-5, meaning that the verbal response times are reduced across block 1-5 because the same words are repeated, and a novelty effect from block 5 to block 6, meaning an increase in verbal response time because new words are presented. The priming effect and the novelty effect were greater for the verb generation task, as shown by a large Task x Block interaction. Specifically, the verbal response times across block 1-5 were reduced more during verb generation than during noun reading and verb reading. The increase in verbal response times from block 5 to 6, was greater for verb generation than for noun reading and verb reading.

Response variability

For the response variability, a Condition x Task x Session x Block ANOVA revealed no main effect of Condition. A large main effect of Session was found, such that the response variability was greater after tDCS (0.096 ± 0.002 s) than before (0.091 ± 0.002 s). However, there was no Condition x Session interaction, indicating no overall effect of tDCS on verbal response variability. In line with the study of Pope and Miall, there was a large main effect of Task, such that verbal response times were more variable during verb generation (0.168 ± 0.004 s) than during noun

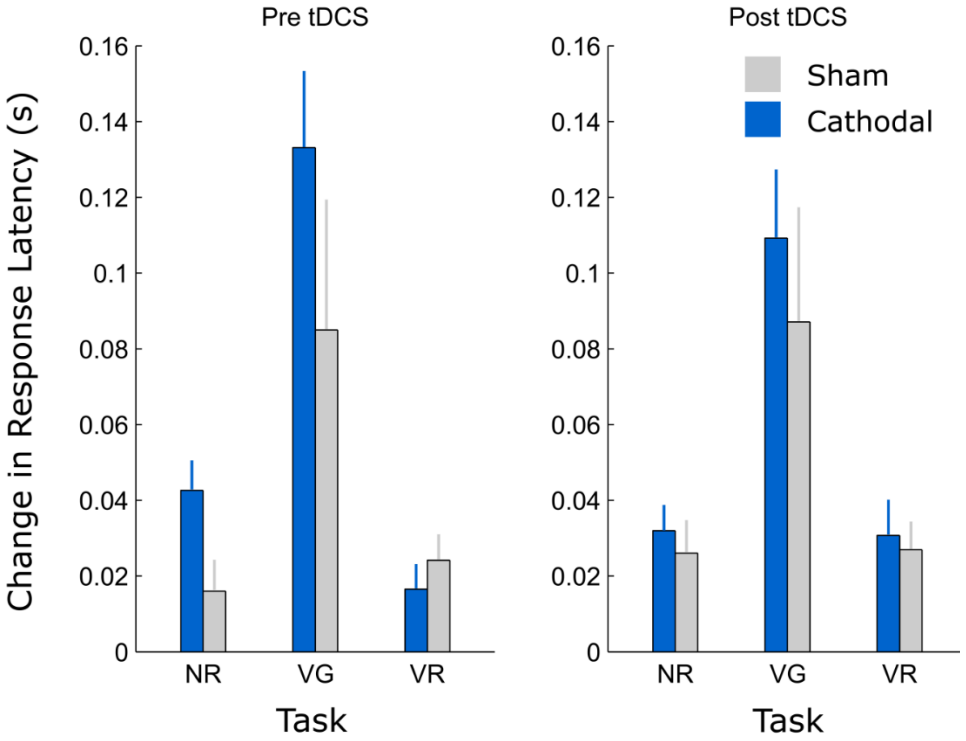


Figure 3. Results for the learning variable. Calculated by subtracting the verbal response times (s) in block 5 from the verbal response times (s) in block 1. This difference is presented for each task, before and after tDCS. Error bars present the Standard Error of the Mean (SEM).

reading (0.054 ± 0.002 s) and verb reading (0.059 ± 0.002 s). Also, in line with Pope and Miall, a large main effect of Block was found, where response variability decreased across the 5 blocks of repeated words (i.e. priming effect), then increased in block 6, when new word lists were shown (i.e. novelty effect). This pattern for the priming effect and the novelty effect was greater for the verb generation task, as shown by a large Task x Block interaction. Specifically, the response variability across block 1-5 was reduced more during verb generation compared to noun reading and verb reading. The increase in response variability from block 5 to 6 was greater for verb generation than for noun reading and verb reading.

Learning

The results for learning, as reflected in the difference in response times between block 1 and block 5, are presented in Figure 3. A Condition x Task x Session ANOVA revealed no significant main effect of Condition and no significant main effect of Session, indicating there was no effect of tDCS. In line with the study of Pope and Miall, there was a large main effect of Task, such that there was a larger improvement of verbal response times across block 1-5 for the verb generation task (0.104 ± 0.015 s), compared to noun reading (0.029 ± 0.005 s) and verb reading (0.025 ±

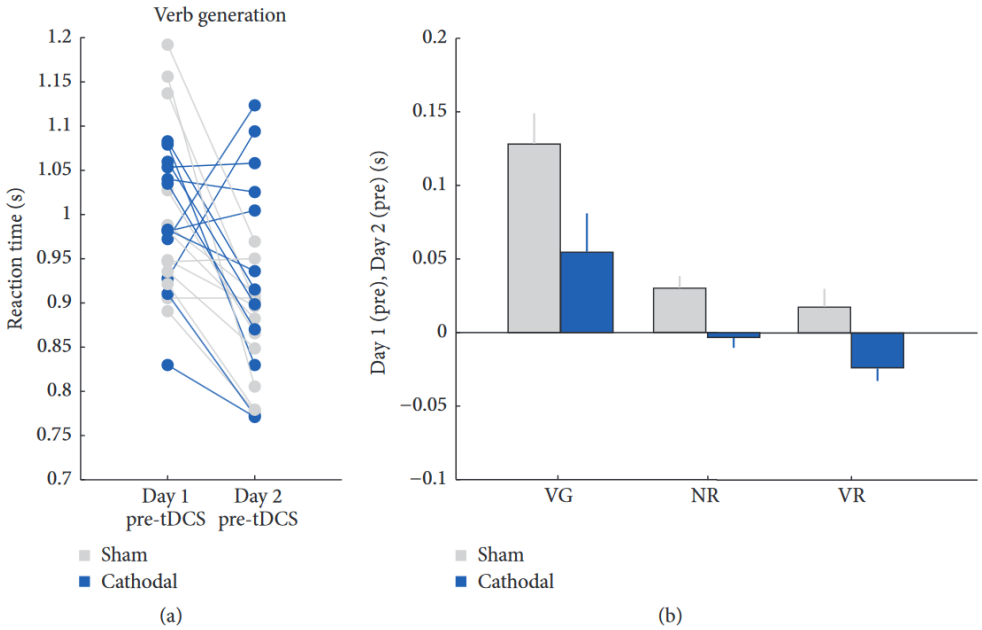


Figure 4. Results for the long term effects. A shows the individual verbal response times on the verb generation task, for visit 1 and visit 2. **B** shows the mean verbal response times for each task, subtracting performance in the second visit from the first visit. Error bars present the Standard Error of the Mean (SEM).

0.004 s). In contrast with the study of Pope and Miall, the present study did not demonstrate a Condition x Session x Task interaction.

Change in variability

For the total learning variability across block 1 to 5 (i.e. analyzing the standard deviations for the learning variable), a Condition x Task x Session ANOVA revealed no main effect of Condition. A large main effect of Session was found, such that the change in response variability was greater after tDCS (0.023 ± 0.004 s) than before (0.008 ± 0.005 s). However, there was no Condition x Session interaction, indicating no overall effect of tDCS on the change in variability. In line with the study of Pope and Miall, there was a large main effect of Task, such that the change in response variability between block 1 and 5 was greater for verb generation (0.035 ± 0.011 s), than for noun reading (0.005 ± 0.002 s) and verb reading (0.006 ± 0.002 s). A significant, large Task x Session interaction was found, such that the change in response variability between before and after tDCS was greater for the verb generation task, than for noun reading and verb reading. In contrast with the study of Pope and Miall, the present study did not demonstrate a Condition x Session x Task interaction.

Long term effects - Verbal response times

A Condition x Task x Session x Block ANOVA including block 1-5 and with visit-order as a between-subject factor (i.e. labeled as Visit) revealed a significant Condition x Visit interaction, $F(1,22)=8.362$, $p=0.008$, $\eta 2=0.275$, such that the mean verbal response times showed a greater reduction for the group receiving sham in the first visit (first visit: 0.727 ± 0.016 s; second visit: 0.681 ± 0.014 s), than for the group receiving cathodal stimulation in the first visit (first visit:

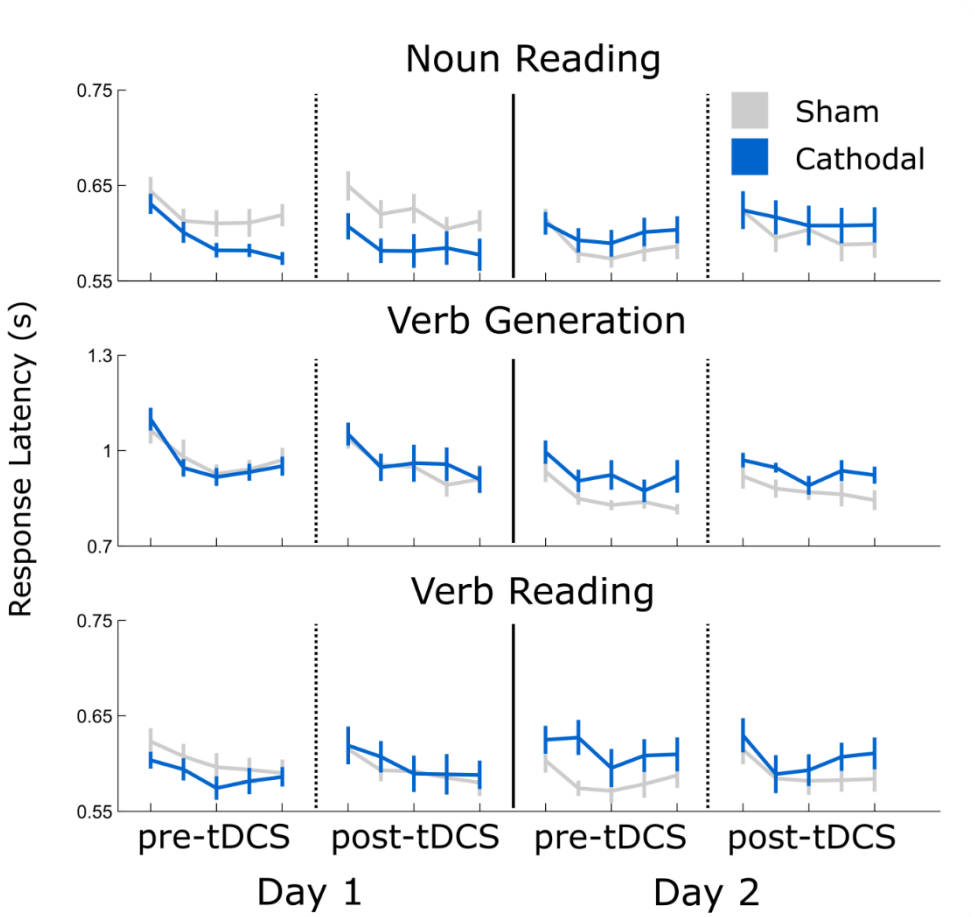


Figure 5. Verbal responses times (s) across block 1-5 and for each task, for the time points pre-tDCS visit 1, post-tDCS visit 1, pre-tDCS visit 2 and post-tDCS visit 2. Blue presents the group starting with the cathodal condition in the first visit and grey presents the group starting with the sham condition in the first visit. Error bars present the Standard Error of the Mean (SEM).

0.717 ± 0.014 s; second visit: 0.715 ± 0.016 s). This effect was greater for the verb generation task, as shown by a Condition x Task x visit interaction, $F(1.294,28.470)=25.266$, $p<0.001$, $\eta 2=0.535$. Figure 4 presents this interaction effect, showing the mean verbal response times for each task and stimulation condition, and comparing the first visit with the second visit. Specifically, the verbal response times for the verb generation task reduced more for the group receiving sham in the first visit (first visit: 0.963 ± 0.028 s; second visit: 0.864 ± 0.024 s), than for the group receiving cathodal first (first visit: 0.967 ± 0.024 s; second visit: 0.928 ± 0.028 s).

In line with the immediate c-tDCS results, the long term analysis shows a priming effect across block 1-5. Specifically, there was a Condition x Block x Visit interaction, $F(4,88)=3.026$, $p=0.022$, $\eta 2=0.121$, such that the verbal response times across block 1-5 reduced more for the group receiving sham the first time.

Long term effects - Response variability

For the response variability, the ANOVA analysis also revealed a large interaction of Condition x Visit, F(1,22)=14.274, p=0.001, $\eta 2$=0.394, such that the response variability reduced more for the group receiving sham the first time (first visit: 0.094 ± 0.004 s; second visit: 0.082 ± 0.003 s), than for the group receiving cathodal tDCS in the first visit (first visit: 0.096 ± 0.003 s; second visit: 0.089 ± 0.004 s). This effect was also more present for the verb generation task, as shown by a large, interaction effect of Stimulation x Task x Visit, F(1.558,34.280)=40.123, p<0.001, $\eta 2$=0.646. Specifically, the response variability for the verb generation task reduced more for the group receiving sham the first time (first visit: 0.171 ± 0.009 s; second visit: 0.132 ± 0.007 s), than for the group receiving cathodal the first time (first visit: 0.186 ± 0.007 s; second visit: 0.152 ± 0.009 s).

In line with the immediate c-tDCS results, the long term analysis for the response variability also shows a priming effect across block 1-5. Specifically, there was a significant interaction effect of Condition x Block x Visit. F(4,88)=2.596, p=0.042, $\eta 2$=0.106, such that the response variability across block 1-5 reduced more for the group receiving sham the first time. Finally, there was a significant interaction effect of Condition x Task x Block x Visit, F(3.728,82.018)=4.302, p=0.004, $\eta 2$=0.164, such that for the verb generation task, response variability across block 1-5 reduced more for the group receiving sham the first time.

Long term effects - Post-hoc tests: additional analysis of the long term effects

To further study the performance over time and the effect of visit-order, we have performed some additional analysis. Figure 5 presents the performance over time, for each task and across block 1-5, for the timepoints before tDCS visit 1 (pre-tDCS visit 1), after tDCS visit 1 (post-tDCS visit 1), before tDCS visit 2 (pre-tDCS visit 2) and after tDCS visit 2 (post-tDCS visit 2). Blue presents the group starting with the cathodal condition in the first visit and grey presents the group starting with the sham condition in the first visit.

We studied specifically the performance from time point post-tDCS visit 1 to the time point pre-tDCS visit 2 in order to analyze whether performance improved between visits (i.e offline learning). Also, the same set of words was under examination for these 2 time points. An ANOVA including these timepoints, with visit-order as the between-subject variable, revealed that the average performance across block 1-5 improves from post-tDCS visit 1 (0.719 ± 0.014 s) to the pre-tDCS visit 2 (0.693 ± 0.012 s), shown by a large effect, F(1,22)=9.716, p=0.005, η2=0.306. This effect could be interpreted as an effect of offline learning, so participants become better in a task after a time interval. Furthermore, the group receiving sham the first time improves more for these time points (0.721 ± 0.020 s in visit 1 compared to 0.674 ± 0.016 s in visit 2) than the group receiving cathodal tDCS the first time (0.717 ± 0.020 s in visit 1 compared to 0.712 ± 0.016 s in visit 2). This was shown by a large Stimulation x Visit interaction effect, F(1,22)=6,467, p=0.019, η2=0.227. However, these results include only the mean of all blocks, and so it is not possible to discern if any improvements in performance are a result of continued practice or if in fact performance has improved between visits (i.e. offline learning). Therefore, a further step in our analysis was to specifically analyze the time point post-tDCS block 5 of visit 1 and time point pre-tDCS block 1 of visit 2. An ANOVA including these timepoints, with visit-order as the between-subject variable, revealed that the performance on post-tDCS block 5 in visit 1 (0.696 ± 0.015 s) actually decreased in the pre-tDCS block 1 in visit 2 (0.730 ± 0.012 s). This was shown by a large effect of Visit, F(1,22)=9,190, p=0.006, η2=0.295. Therefore, these data show no evidence for offline learning.

Discussion

The aim of the present study was to replicate the results of Pope and Miall by demonstrating that cathodal stimulation of the right cerebellum improves task performance on a verb generation task.[20] The task setup and outcome measures were similar to their study. Based on their results, showing a facilitatory effect immediately after cathodal c-tDCS, we compared cathodal c-tDCS and sham stimulation. In contrast with the between-subject design study of Pope and Miall, the present study used a cross-over within-subject design, in order to reduce the impact of individual variability.[32] Participants had to complete two visits, with half of the group receiving cathodal c-tDCS the first time and half of the group receiving sham c-tDCS the first time. Our results did not show a facilitating effect of cathodal c-tDCS on verb generation, either in terms of verbal response times or variability. In line with Pope and Miall, the verbal response times were larger for the verb generation task, compared to noun reading and verb reading. This effect can be explained with the idea that the verb generation task requires lexical search processes and verbal response selection, while noun and verb reading requires only reading processes. Interestingly, the verbal response times on our tasks were longer than those reported by the original study. These longer reaction times could be due to linguistic factors of the words,[37] for example word length, i.e. words with more phonemes need more time to process.[38] Indeed, on average, the words in our word lists were longer (mean ± SD: 6.13 ± 2.188 phonemes) than the lists of Pope and Miall (mean ± SD: 4.77 ± 1.376 phonemes).[20] Further, in line with Pope and Miall, there was a reduction in response time across block 1-5 (i.e. priming effect) and an increase in block 6 (i.e. novelty effect).

The data of the present study do not confirm that cathodal c-tDCS over the right cerebellum lobule VII leads to disinhibition of the contralateral prefrontal regions and therefore to an improved performance on a cognitive demanding task (i.e. verb generation task). Previous studies have suggested that the Purkinje cells in the right cerebellum would have an inhibitory effect on the contralateral cortical prefrontal regions (i.e. cerebello-cortical inhibition).[9,11–14] For language processing, right cerebellar involvement has also been suggested.[22–24] Specifically, for the verb generation task, a PET scan study and an fMRI study showed that the contralateral cerebellar hemisphere was actively involved.[25–27] However, when investigating the efficacy of c-tDCS in modulating cerebello-cortical inhibition, motor-related studies demonstrate inconsistent findings. For example, one study demonstrates that anodal tDCS to the right cerebellum increases the inhibitory effect to the primary motor cortex whilst cathodal tDCS to the right cerebellum reduces this effect.[15] In contrast, another study in this field report that anodal c-tDCS may reduce the inhibitory effect to the primary motor cortex.[18]

Furthermore, the idea that the cerebellum constraints cortical activity which can be disinhibited by cathodal c-tDCS is also not consistently supported by cognition-related tDCS studies. For example, studies show contradictive results with regards to the application of tDCS to the right cerebellum and its effects on the performance on a verbal Working Memory (WM) task, i.e. forward and backward digit span task. One study shows that cathodal c-tDCS leads to reduced forward digit span and blocks the practice dependent increase in backward digit span,[39] while another study[40] shows that both anodal and cathodal tDCS impairs practice dependent improvement in reaction times in a WM task. Further, Turkeltaub et al.[19] found that both anodal and cathodal c-tDCS enhanced the performance on a phonemic fluency task, however, the anodal effect was found to be more robust. Taken together, it seems that c-tDCS studies are not yet consistent whether anodal or cathodal c-tDCS improves or disrupts task performance in healthy subjects. Future studies need to further explore the specific polarity effects of c-tDCS in order to understand its usage for cerebellar dependent cognitive processing.

Interestingly, we observe a long term effect of c-tDCS in our data. When analyzing the data further by taking into account visit-order, we found that the group receiving cathodal c-tDCS the first time demonstrated poorer performance in the second visit in comparison to those who received sham stimulation the first time. First of all, the group receiving cathodal c-tDCS in the first visit demonstrate less improvement from visit 1 to visit 2. Also, the group receiving cathodal c-tDCS in the first visit show less improvement during the second visit (i.e. performance across block 1-5) compared to the group receiving sham the first time. Regarding response variability, the same findings are found, thus the group receiving cathodal c-tDCS in the first visit show increased variability in verbal response times in the second visit and during the second visit (i.e. increased variability across block 1-5). In motor-related studies, this long term effect is often called a consolidation effect, meaning that after acquisition performance can become resistant to decay.[41] To our knowledge, studies investigating consolidation effects of c-tDCS on a language task are scarce, whereas there are several motor-related c-tDCS studies that investigate the effect of c-tDCS on a longer time scale. For example, one such study demonstrated that anodal c-tDCS would enhance general motor skill learning and sequence-specific learning, 35 minutes after tDCS stimulation.[42] Another study shows that anodal c-tDCS to the right cerebellum improves task performance on a temporal motor task in the follow-up tests (90 minutes and 24h after training).[43] Furthermore, a recent study provides evidence that cathodal c-tDCS impairs overnight retention of a force field reaching task.[44] Therefore, these motor-related studies show that, on a longer time scale, anodal c-tDCS may enhance performance, while cathodal c-tDCS may impair performance, which is in line with the long term results of the present study.

Studies focusing on the adaptation of movements and tDCS have demonstrated a dissociation between the acquisition phase and the consolidation phase.[45,46] Specifically, anodal tDCS to the right cerebellum leads to an increased acquisition of new internal models whereas anodal tDCS to the motor cortex leads to improved consolidation. Therefore, the cerebellum is believed to rapidly acquire new internal models that are also quickly forgotten whereas the motor cortex learns more slowly but retains better (i.e. consolidation). A similar transfer of learning from the cerebellar cortex to other structures has been proposed for other cerebellar dependent adaptation tasks such as eye-blink conditioning or adaptation of the vestibulo-ocular reflex.[47] In the present study, it is possible that these two partially separable effects are at work: short terms changes in firing rate of the cerebellum and additional effects on plasticity. First, cathodal c-tDCS may indeed reduce the firing rate of Purkinje cells and the inhibitory tone on the prefrontal cortex, and therefore improve performance in tasks relying on these cortical areas, as found in the study of Pope and Miall. However, it should be noted that there is no direct neurophysiological evidence for this effect of c-tDCS specifically on the prefrontal cortex. Secondly, cathodal c-tDCS may also reduce plasticity in the cerebellar cortex and therefore retard the rate of learning there, subsequently reducing the amount that can be transferred to other areas for consolidation, which may be in line with the results of the present study.

The present within-subject design with several time points allows us to evaluate different sub-concepts of consolidation. Consolidation can be described in terms of offline learning, i.e. improvements in performance between visits, and memory stabilization, i.e. reduced performance compared to the end of the previous visit but increased performance in comparison to the naïve state.[48] However, the degree to which either or both of these is possible is dependent on task structure and the particular skill under consideration. An important consideration in interpreting our results is separating the effect of repeated practice from true offline learning. The results of the present study show that the average performance across block 1-5 improves

from time point post-tDCS in the first visit to time point pre-tDCS in the second visit. Furthermore, the group receiving sham the first time improves more for these time points than the group receiving cathodal stimulation the first time. Therefore these results may show an effect of offline learning, however, if only the mean of all blocks is used as a measure of performance it is not possible to discern if any improvements are a result of continued practice or if in fact performance has improved between visits.[48] Further analysis demonstrates that performance in both groups (i.e. the group receiving cathodal stimulation the first time and the group receiving sham the first time) decreased between block 5 of the first visit and block 1 of the next, despite the fact that the same set of words was under examination. These data therefore show no evidence for offline learning but that may be due to the relatively long period of time between visits or because this particular task is not appropriate for such changes. In the future it will be interesting to test subjects again after a shorter interval to assay if offline learning is indeed possible with this task. It is important to note that offline learning has been investigated in an fMRI learning paradigm in which subjects had to learn a new lexicon and were tested 20 minutes later.[49] The degree of offline learning was positively correlated with the level of activation of the right cerebellum. Therefore, these data provide evidence for a role of the cerebellum in consolidation of a learning task that includes language/linguistic aspects. The differences between learning a new lexicon and learning associations within a known lexicon (as here), especially when concerning the cerebellum, are unknown and it is vital for proper delineation of tDCS effects that the specific task demands are well understood.

Limitations of the study

First of all, it should be noted that the design of the present study with 1 week between 2 visits could interfere with replication of the original immediate effect reported by Pope and Miall. This interference could be due to effects of retesting the same words or a ceiling effect. Furthermore, in the present study the subjects had one block of novel words at the end of the five blocks of repeated words which may have also acted as an interfering factor. As the majority of the results found in both the present study and the original Pope and Miall study can be found within blocks 1-5 it would be interesting to repeat the experiment with the omission of the novel words in block 6 to test if any interference is occurring. Finally, it should be noted that the majority of (c-)tDCS studies are described in the context of motor tasks and we therefore used these studies in order to interpret our results, however, the analogy between motor learning, consolidation and the type of results presented here may be stretched.

Conclusion and future recommendations

The present study shows that long term effects of c-tDCS need to be taken into account when investigating the effect of c-tDCS on language task performance. Most tDCS studies with a motor or non-motor learning task focus on direct results rather than long term learning effects (i.e. consolidation). Our findings warrant further investigation into long term effects of c-tDCS, to better capture its effect and how we can use this application to understand the complex role of the cerebellum on cognitive/language processing. Therefore, we first need to understand c-tDCS in healthy subjects, before undertaking clinical studies with post-stroke patients with aphasia. To further explore the long term effect of c-tDCS on a cognitive language task, we would suggest to combine the design of Pope and Miall with the design of the present study. This combined design would describe the effect of c-tDCS in 3 conditions - anodal c-tDCS, cathodal c-tDCS and sham (between-subject) - and participants need to come twice in each condition (within-subject). This

design allows us to evaluate the effect of anodal c-tDCS compared to the effect of cathodal c-tDCS, on a longer time scale. Furthermore, techniques such as EEG may be used to explore the effect of cerebellar tDCS and its polarity specific effects on ongoing or induced activity in areas of the cortex associated with language.

References

1. Nitsche, M.A. & Paulus, W. J Physiol 527 Pt 3, 633–639 (2000).
2. Nitsche, M.A. & Paulus, W. Neurology 57, 1899–1901 (2001).
3. Akhtar, H., Bukhari, F., Nazir, M., Anwar, M.N. & Shahzad, A. Neurosci. Bull. 32, 115–126 (2016).
4. Broeder, S. et al. Neurosci. Biobehav. Rev. 57, 105–117 (2015).
5. Kang, N., Summers, J.J. & Cauraugh, J.H. J. Neurol. Neurosurg. Psychiatry 87, 345–55 (2016).
6. Monti, A. et al. J. Neurol. Neurosurg. Psychiatry 84, 832–842 (2013).
7. Horvath, J.C., Forte, J.D. & Carter, O. Brain Stimul. 8, 535–50
8. Vannorsdall, T.D. et al. Cogn. Behav. Neurol. 29, 11–17 (2016).
9. Stoodley, C.J. The Cerebellum 11, 352–365 (2012).
10. Schmahmann, J. & Sherman, J.C. Brain 121, 561–579 (1998).
11. Stoodley, C.J., Valera, E.M. & Schmahmann, J.D. Neuroimage 59, 1560–1570 (2012).
12. Krienen, F.M. & Buckner, R.L. Cereb. Cortex 19, 2485–2497 (2009).
13. O'Reilly, J.X., Beckmann, C.F., Tomassini, V., Ramnani, N. & Johansen-Berg, H. Cereb. Cortex 20, 953–965 (2010).
14. Stoodley, C.J. & Schmahmann, J.D. Neuroimage 44, 489–501 (2009).
15. Galea, J.M., Jayaram, G., Ajagbe, L. & Celnik, P. J. Neurosci. 29, 9115–22 (2009).
16. Chan, C.Y., Hounsgaard, J. & Nicholson, C. J. Physiol. 402, 751–71 (1988).
17. Rahman, A., Toshev, P.K. & Bikson, M. Clin. Neurophysiol. 125, 435–8 (2014).
18. Doeltgen, S.H., Young, J. & Bradnam, L. V. The Cerebellum 15, 466–474 (2016).
19. Turkeltaub, P.E., Swears, M.K., D'Mello, A.M. & Stoodley, C.J. Restor. Neurol. Neurosci. 34, 491–505 (2016).
20. Pope, P.A. & Miall, R.C. Brain Stimul. 5, 84–94 (2012).
21. Pope, P.A., Brenton, J.W. & Miall, R.C. Cereb. Cortex 25, 4551–8 (2015).
22. De Smet, H.J., Paquier, P., Verhoeven, J. & Mariën, P. Brain Lang. 127, 334–342 (2013).
23. Mariën, P. et al. Cerebellum 13, 386–410 (2014).
24. Marien, P., Engelborghs, S., Fabbro, F. & De Deyn, P.P. Brain Lang. 79, 580–600 (2001).
25. Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M. & Raichle, M.E. Nature 331, 585–589 (1988).
26. Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M. & Raichle, M.E. J. Cogn. Neurosci. 1, 153–170 (1989).
27. Frings, M. et al. Neurosci. Lett. 409, (2006).
28. Pope, P.A. & Miall, R.C. Front. Psychiatry 5, 33 (2014).
29. Fridriksson, J., Richardson, J.D., Baker, J.M. & Rorden, C. Stroke 42, 819–821 (2011).
30. Marangolo, P. et al. Front. Hum. Neurosci. 7, 539 (2013).
31. Meinzer, M., Darkow, R., Lindenberg, R. & Flöel, A. Brain 139, 1152–1163 (2016).
32. Wiethoff, S., Hamada, M. & Rothwell, J.C. Brain Stimul. 7, 468–475 (2014).
33. Oldfield, R.C. Neuropsychologia 9, 97–113 (1971).
34. De Witte, E. et al. Brain Lang. 140, 35–48 (2015).
35. Brainard, D.H. Spat. Vis. 10, 433–436 (1997).
36. Pelli, D.G. Spat. Vis. 10, 437–442 (1997).
37. Rayner, K. & Duffy, S.A. Mem. Cognit. 14, 191–201 (1986).
38. Barton, J.J.S., Hanif, H.M., Eklinder Björnström, L. & Hills, C. Cogn. Neuropsychol. 31, 378–412 (2014).
39. Boehringer, A., Macher, K., Dukart, J., Villringer, A. & Pleger, B. Brain Stimul. null, 649–653 (2012).
40. Ferrucci, R. et al. J. Cogn. Neurosci. 20, 1687–1697 (2008).
41. Savic, B. & Meier, B. Front. Hum. Neurosci. 10, 26 (2016).
42. Ferrucci, R. et al. Cerebellum (2013).doi:10.1007/s12311-012-0436-9

43. Wessel, M.J. et al. Cereb. Cortex (2015).doi:10.1093/cercor/bhu335
44. Herzfeld, D.J. et al. Neuroimage 98, 147–58 (2014).
45. Galea, J.M., Vazquez, A., Pasricha, N., de Xivry, J.J. & Celnik, P. Cereb Cortex 21, 1761–1770 (2011).
46. Smith, M.A., Ghazizadeh, A. & Shadmehr, R. PLoS Biol. 4, e179 (2006).
47. Krakauer, J.W. & Shadmehr, R. Trends Neurosci. 29, 58–64 (2006).
48. Robertson, E.M., Pascual-Leone, A. & Miall, R.C. Nat. Rev. Neurosci. 5, 576–582 (2004).
49. Lesage, E., Nailer, E.L. & Miall, R.C. Brain Lang. 161, 33–44 (2016).

4.4 BDNF Val66Met but not transcranial direct current stimulation affects motor learning after stroke

Rick van der Vliet, Gerard M. Ribbers, Yves Vandermeeren, Maarten A. Frens and Ruud W. Selles

Abstract

Background: tDCS is a non-invasive neuromodulation technique that has been reported to improve motor skill learning after stroke. However, the contribution of tDCS to motor skill learning has only been investigated in a small number of studies. In addition, it is unclear if tDCS effects are mediated by activity-dependent BDNF release and dependent on timing of tDCS relative to training.

Objective: Investigate the role of activity-dependent BDNF release and timing of tDCS relative to training in motor skill learning.

Methods: Double-blind, between-subjects randomized controlled trial of circuit tracing task improvement (ΔMotor skill) in 80 chronic stroke patients who underwent tDCS and were genotyped for BDNF Val66Met. Patients received either short-lasting tDCS (20 minutes) during training (Short-lasting online group), long-lasting tDCS (10 minutes – 25 minutes break – 10 minutes) one day before training (Long-lasting offline group), short-lasting tDCS one day before training (Short-lasting offline group), or sham tDCS. ΔMotor skill was defined as the skill difference on the circuit tracing task between day one and day nine of the study.

Results: Having at least one BDNF Met allele was found to diminish ΔMotor skill ($\beta_{BDNF,Met}$=-0.217 95%HDI=[-0.431 -0.0116]), indicating activity-dependent BDNF release is important for motor skill learning after stroke. However, none of the tDCS protocols affected ΔMotor skill ($\beta_{Short\text{-}lasting,online}$=0.0908 95%HDI=[-0.227 0.403]; $\beta_{Long\text{-}lasting,offline}$=0.0242 95%HDI=[-0.292 0.349]; $\beta_{Short\text{-}lasting,offline}$=-0.108 95%HDI=[-0.433 0.210]).

Conclusion: BDNF Val66Met is a determinant of motor skill learning after stroke and could be important for prognostic models. tDCS does not modulate motor skill learning in our study and might be less effective than previously assumed.

Introduction

tDCS is a non-invasive neuromodulation technique that has been reported to improve upper limb rehabilitation after stroke in pilot studies [1–4], presumably by increasing the ability to learn a motor skill [5–14]. The favorable effects of tDCS on motor skill learning after stroke are thought to rely on a polarity-specific release of brain-derived neurotrophic factor (BDNF) [8], down-regulation of GABA [15–18] and restoration of the interhemispheric imbalance between the affected motor cortex and the unaffected motor cortex [19–22]. In this study, we investigate the influence of BDNF Val66Met and tDCS on motor skill learning in chronic stroke patients.

The role of BDNF as a link between tDCS and motor skill learning has been suggested by electrophysiological studies in mice and genetic analyses in healthy subjects. In mouse cortical

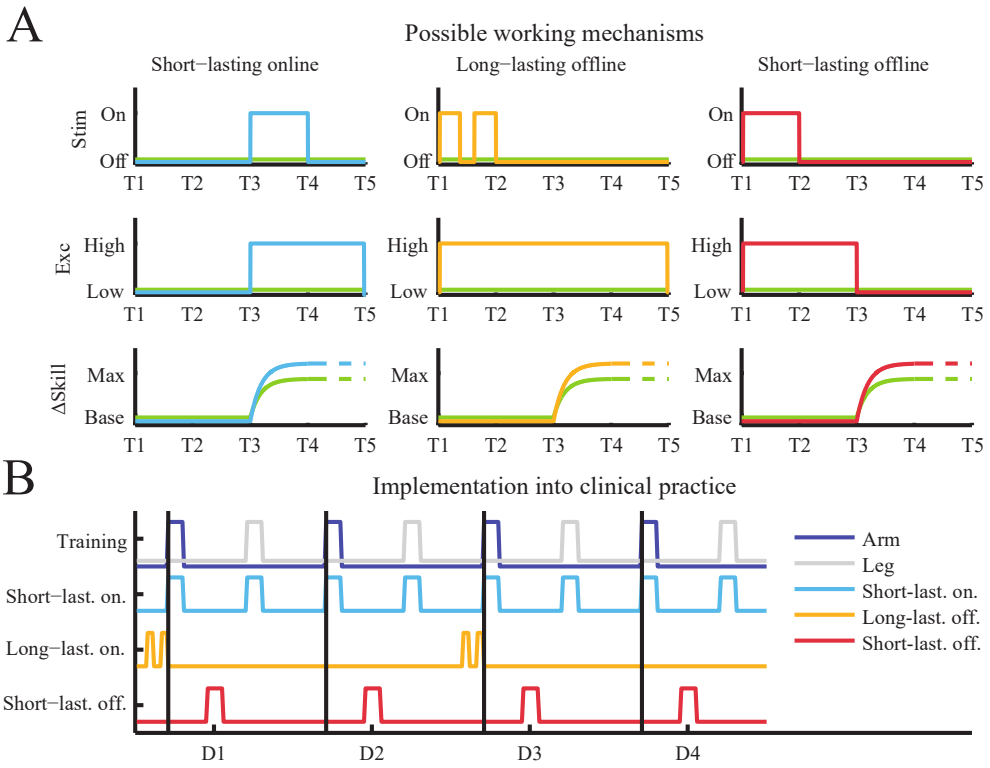


Figure 1. tDCS timing hypotheses. A. Schematic representation of the three possible ways to time tDCS relatively to the skill training. The top row shows timing of tDCS, the middle row the cortical excitability changes that may result from tDCS, and the bottom row the potential motor skill learning changes that may results from this. For tDCS to improve motor skill learning in stroke patients, tDCS itself should overlap with training in which case short-lasting online tDCS is appropriate (left column), tDCS aftereffects should overlap with training, in which case long-lasting offline tDCS might be optimal (middle column) or tDCS should precede training without tDCS itself or the aftereffects overlapping with training in which case shorting-lasting offline tDCS would suffice (right column). **B.** Potential implementation of the different tDCS protocols in motor rehabilitation. The top lines indicate when patients are doing arm or leg motor training. In case tDCS only improves motor skill learning when applied concurrently with training, short-lasting online tDCS should be applied during every single session (blue line). If, however, motor skill learning is facilitated during the aftereffects as well, patients could be stimulated every other day with long-lasting offline tDCS (orange line). Finally, if tDCS affects motor skill learning independently of the direct or aftereffects, short-lasting offline

slices, BDNF concentrations were shown to rise after direct current stimulation, increasing long-term potentiation of horizontal connections [8], which underlies motor skill learning [23,24]. Activity-dependent release of BDNF has been related to motor skill learning in healthy subjects by studying the role of the common (approximately 30% of the Caucasian population [25,26]) secretion-limiting [27] BDNF Val66Met polymorphism. Agreeing with the function of BDNF in motor cortex long-term potentiation, carriers of this polymorphism were found to more slowly acquire a new motor skill [8,28]. Since tDCS increases BDNF release in mouse brain slices and increased BDNF release is linked to faster motor skill learning in healthy subjects, tDCS may promote motor skill learning through BDNF release [8]. However, whether activity-dependent release of BDNF plays a role in motor skill learning after stroke as well and could therefore mediate tDCS effects in this patient group has yet to be established.

The contribution of tDCS to motor skill learning in chronic stroke patients has only been investigated in a small number of studies [9,11,14]. In addition, the importance of timing of tDCS and tDCS aftereffects relative to training is currently unclear [29]. Aftereffects of tDCS are periods of increased motor cortex excitability (usually measured with transcranial magnetic stimulation) following tDCS, which last up to 60 minutes for short-lasting protocols [30,31] and up to two days for long-lasting protocols [32,33]. Currently, it is unknown if tDCS itself should overlap with training [2,4,14,5–12] in which case short-lasting online (during training) protocols are appropriate, tDCS aftereffects should overlap with training, in which case long-lasting offline (before training) protocols might be optimal [1,3,13,34] or tDCS should just precede training without direct or aftereffects necessarily overlapping with training in which case short-lasting offline protocols would suffice (see Figure 1A). Resolving how timing relative to training influences the effect of tDCS on motor skill learning in stroke patients is therefore important because it determines the design of rehabilitation programs (see Figure 1B).

In this study, we evaluated motor skill learning in chronic stroke patients who were genotyped for BDNF Val66Met and received tDCS. Design of the motor skill learning task was identical to Lefebvre et al. [9,11], who found performance improvements with short-lasting tDCS in chronic stroke patients. We hypothesized that non-carriers (no Met alleles) would learn better than carriers (at least one Met allele). The timing hypotheses were addressed by comparing short-lasting tDCS applied during training (short-lasting online group), long-lasting tDCS applied one day before training (long-lasting offline group), and short-lasting tDCS applied one day before training (short-lasting offline group), to a sham tDCS protocol.

Materials and methods

Participants

Patients between the age of 18 and 80 who had suffered from stroke with hemiparesis at least 6 months prior to the study were eligible for participation. We excluded patients incapable of voluntary movement (Medical Research Council scale < 2) or unable to understand verbal instructions; with a history of head injury, cranial irradiation, epilepsy, substance abuse or psychiatric disorders; taking anticonvulsant or antiepileptic medication at the time of the study or carrying intracranial metal or a pacemaker.

We reviewed medical records of patients who were discharged from the Rijndam rehabilitation center between November 2008 and August 2015 to assess their eligibility. All patients who met the criteria were sent an invitation letter and called if they were willing to participate. After oral consent, visits to the rehabilitation center were planned at least 24 hours later. Patients were asked for written consent on the first day of the study. The study was

conducted in accordance with the 1964 Helsinki Declaration and approved by the medical ethics committee of the Erasmus MC university medical center.

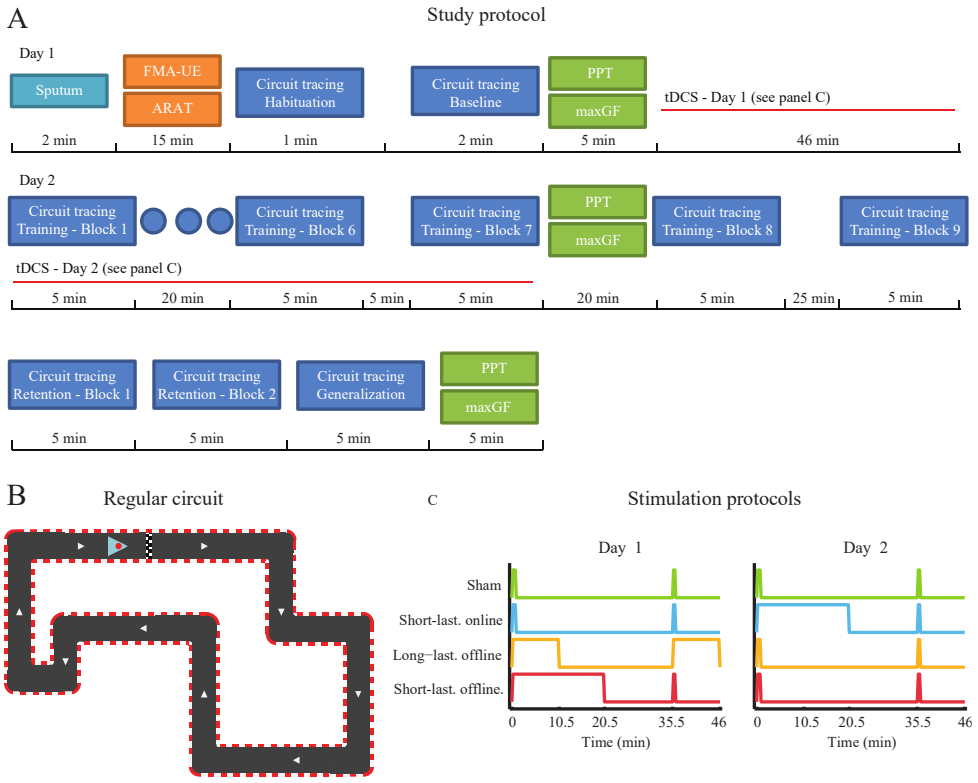


Figure 2. Study design. A. Study protocol. Chronic stroke patients visited the rehabilitation center on two consecutive days and again one week later (days one, two and nine of the study) to complete movement tasks and tDCS. Patients performed the circuit tracing task on day one to assess baseline performance, on day two to train and on day nine to determine ΔMotor skill. tDCS was added according to one of four tDCS protocols on day one during rest and on day two during training. **B.** Training circuit. Patients were seated behind a computer and asked to hold a computer mouse with their affected hand. The computer mouse controlled a cursor (blue arrow with a red dot in the center) on the screen which patients could freely move over a two-dimensional background. This background consisted of a black surface and a gray circuit sheathed with a red and white striped line. A black and white blocked line drawn perpendicularly over the circuit direction indicated the start and finish. White triangles, marking the center of the circuit, reminded the subject of the correct movement direction. **C.** tDCS protocols. We used three different tDCS protocols in our study: 30 seconds bihemispheric 1mA tDCS at t=0 and t=35.5 minutes (sham), 20 minutes bihemispheric 1mA tDCS at t=0 and 30 seconds bihemispheric 1mA tDCS at t=35.5 minutes (short-lasting tDCS) and 10 minutes bihemispheric 1mA tDCS at t=0 and t=35.5 minutes (long-lasting tDCS). All stimulation periods included a 15 second ramp-up and a 15 second ramp-down period to ensure comfort. The tDCS protocols were combined in different ways to create four tDCS groups: sham tDCS on day one and day two (sham group), sham tDCS on day one and short-lasting tDCS on day two (short-lasting online group), long-lasting tDCS on day one and sham tDCS on day two (long-lasting offline group), and short-lasting tDCS on day one and sham tDCS on day two (short-lasting offline group). ARAT = action research arm test; FMA-UE = Fugl-Meyer assessment of the upper extremity; maxGF = maximum grip force; PPT = Purdue Pegboard Test; tDCS = transcranial direct current stimulation.

Study design

Chronic stroke patients visited the rehabilitation center on two consecutive days and again one week later (days one, two and nine of the study). On the first day, we obtained a sputum sample (Oragene Discover OGR-500, DNA Genotek Inc., Ottawa, Ontario, Canada) for genetic analysis and quantified the patient's arm hand performance with the Fugl-Meyer assessment of the upper extremity [35] and the action research arm test [36]. All other procedures are described in more detail below. A schematic representation of the study is given in Figure 2A.

Motor skill learning

We measured motor skill learning with a circuit tracing task which has been used before to study the effects of tDCS [9,11]. Patients were seated in front of a PC and asked to hold a computer mouse with their affected hand. The computer mouse controlled a cursor on the screen which patients could freely move over a two-dimensional background (see figure 2B).

A trial started with the appearance of the circuit and cursor. The patient was instructed to move the cursor as fast and accurate as possible over the circuit in a clockwise direction for 30 seconds. The circuit could be traced more than once. After the trial, the screen turned black for 30 seconds, giving patients a short pause before the next trial. During some of these breaks, performance measures (cursor velocity, movement error and the skill score) of the previous trial were displayed on the screen.

The skill learning task included habituation, baseline, regular and generalization trials. In habituation trials, the circuit was a simple square. Baseline trials and regular trials introduced a more challenging circuit made out of a polygon with right angles. The generalization circuit was a comparable but different polygon of equal difficulty [9]. Performance scores were shown after regular and generalization trials but not habituation and baseline trials. Habituation blocks consisted of a single habituation trial, baseline blocks of two baseline trials, training and retention blocks of five regular trials and a generalization block of five generalization trials. The first day was comprised of a single habituation block (H) and a baseline block (B). The second day of nine training blocks (T1-T9), with T2-T6 preceded by brief pauses and T7-T9 by 5, 20 and 25 minute breaks. The ninth day consisted of two retention blocks (R1-R2) and a generalization block (G).

Motor skill change (ΔMotor skill) was the primary outcome of this study. This measure was based on Lefebvre et al. [9], but modified to improve accuracy. We first resampled movement on a 40Hz time grid using piecewise cubic interpolation and discarded the first and last second of every trial. Next, movement error was calculated as the shortest distance between the cursor and the circuit with larger distances resulting in higher error. Movement error was converted into precision by taking the negative natural logarithm. This measure assigns a higher score to lower movement error. Precision was averaged over all data points in a single block consisting of either two trials (baseline block) or five trials (training and generalization blocks) resulting in error measures per block. We determined movement velocity by calculating the vector projection of cursor displacement per time step on the closest segment of the circuit. This approach ensured that only velocities in the correct direction were rewarded with a higher velocity score. Movement velocity was averaged per block and patient.

Precision and velocity for all blocks were subsequently normalized into z-scores using the group mean and standard deviation of baseline precision and velocity. This normalization ensured improvements in error and velocity contribute equally to an improvement in skill score, with equal weighing factors for all patients.

Motor skill is the average of the two z-scores. This score gets higher with increased movement velocity and/or increased precision. Finally, we calculated ΔMotor skill by subtracting baseline motor skill for each patient.

Manual dexterity

We measured manual dexterity with the Purdue pegboard test (PPT, Lafayette Instrument Company, Lafayette, Indiana, USA). Patients had to place as many pegs as possible into vertically arranged holes in 30 seconds using the affected hand. This test was repeated three times on the first, second and ninth day of the study. The mean difference in number of pegs placed between the first and ninth day of the study (ΔPPT) was the outcome.

Maximum grip force

We determined maximum grip force with a digital hand dynamometer (Pattern medical, Warrenville, Illinois, USA). Patients had to squeeze down as hard as they could for a couple of seconds. This test was repeated three times on the first, second and ninth day of the study. The mean difference in maximum grip force between the first and ninth day of the study (ΔmaxGF) was the outcome.

tDCS

tDCS was applied using a Starstim device (Neuroelectrics, Barcelona, Spain) with 5 cm diameter sponge electrodes. Hand areas of both motor cortices were localized with transcranial magnetic stimulation (Neurosoft, Ivanovo, Russia). and thoroughly cleaned with Nuprep (Weaver and Company, Aurora, Colorado, United States). The positive sponge electrode was placed over the affected hemisphere, the negative sponge electrode over the unaffected hemisphere with an elastic headband (bihemispheric montage). We used bihemispheric tDCS because it (1) may have a larger effect on motor skill learning than unihemispheric tDCS [12] and (2) was also used by Lefebvre et al [9,11]. Stimulation intensity was set to 1mA, in agreement with Lefebvre et al [9,11] and many other tDCS studies [2,6–8,10,32,33].

We used three different tDCS protocols (see Figure 2C): 30 seconds bihemispheric 1mA tDCS at t=0 and t=35.5 minutes (sham), 20 minutes bihemispheric 1mA tDCS at t=0 and 30 seconds bihemispheric 1mA tDCS at t=35.5 minutes (short-lasting tDCS), and 10 minutes bihemispheric 1mA tDCS at t=0 and t=35.5 minutes (long-lasting tDCS). The long-lasting protocol was based on [33]. Duration of short-lasting tDCS was set to 20 minutes, which is a widely-used duration [2,6–8,10,14,37], to keep total tDCS duration of the non-sham protocols similar. Patients were blinded to the tDCS protocol with 30-second stimulation bursts that evoke a sensation similar to more prolonged tDCS sessions without affecting excitability [38]. Because tDCS started at t=0 and t=35.5 minutes for the long-lasting protocol, 30-second stimulation bursts were included at t=0 and t=35.5 minutes for the sham protocol, and at t=35.5 for the short-lasting protocols. All stimulation periods included a 15 second ramp-up and a 15 second ramp-down period to ensure comfort.

The three tDCS protocols were combined in different ways to create four tDCS groups: sham tDCS on day one and day two (sham group), sham tDCS on day one and short-lasting tDCS on day two (short-lasting online group), long-lasting tDCS on day one and sham tDCS on day two (long-lasting offline group), and short-lasting tDCS on day one and sham tDCS on day two (short-lasting offline group).

tDCS was controlled from a PC according to the patient's tDCS group and the day of the experiment. The program only showed day and group numbers without protocol names to blind the experimenter and patient from the protocol.

Randomization

Patients were randomized to one of the four tDCS groups on the first day of the study using a minimization approach. Minimization is a randomization technique which aims to balance predefined patient characteristics and group sizes by adapting the allocation probability of groups. This way, group characteristics and size can be better controlled [39]. We chose to minimize the difference between groups in the Fugl-Meyer assessment of the upper extremity (dichotomized: low < 50; high >= 50), stroke laterality (dominant/ non-dominant hand) and age (dichotomized: low < 60; high >=60). Minimization was implemented in MinimPy (freely available from: https://sourceforge.net/projects/minimpy/).

Sample size calculation

We powered our study to be 90% sure of finding a 95% highest density interval not containing zero for the short-lasting online tDCS group compared to sham given a ΔMotor skill of 0.4 and standard deviations of 0.35 based on Lefebvre et al. [9,11]. This resulted in a minimum 18 patients per group. Based on this number, we decided to include a total 80 patients in this study.

Genetics

BDNF Val66Met was genotyped with Taqman Allelic Discrimination using the Assay-On-Demand service of Life Technologies. Reactions were performed in a 384-wells format in a total volume of 2 µL containing 2 ng DNA, 1x Taqman assay, and 1x genotyping master mix. Polymerase chain reaction cycling consisted of initial denaturation for 15 minutes at 95° C, and 40 cycles with denaturation of 15 seconds at 96° C and annealing and extension for 60 seconds at 60.0° C. Signals were read with the Taqman 7900HT and analyzed using the sequence detection system 2.4 software. All materials and software were from Life Technologies (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Statistics

We used Bayesian linear regression for our data-analysis. This analysis is similar to regular linear regression, but is able to deal with missing data points and therefore does not require imputation. In addition, the confidence intervals for the regression coefficients can be interpreted directly as probability intervals, as opposed to regular linear regression, which makes the analysis more intuitive.

We applied Bayesian linear regression with ΔMotor skill on day nine (two training blocks and one generalization block) as the dependent variables and tDCS group (short-lasting online, long-lasting offline and short-lasting offline), $BDNF_{Met}$ (non-carriers or carriers), generalization (retention block or generalization block), age (<60 years or >=60 years), stroke dominance (dominant or non-dominant hand), gender and Fugl-Meyer assessment of the upper extremity (<50 points or >= 50 points) as the independent variables. The intercept for this model was calculated for each patient individually ("individual training effect") from a group prior ("group training effect") to account for differences in ΔMotor skill between patients unexplained by the independent variables. In addition, we calculated an unadjusted model including only the

tDCS groups and generalization. Because BDNF results were missing in eight patients, we modeled $BDNF_{Met}$ as a Bernoulli distribution with a beta prior.

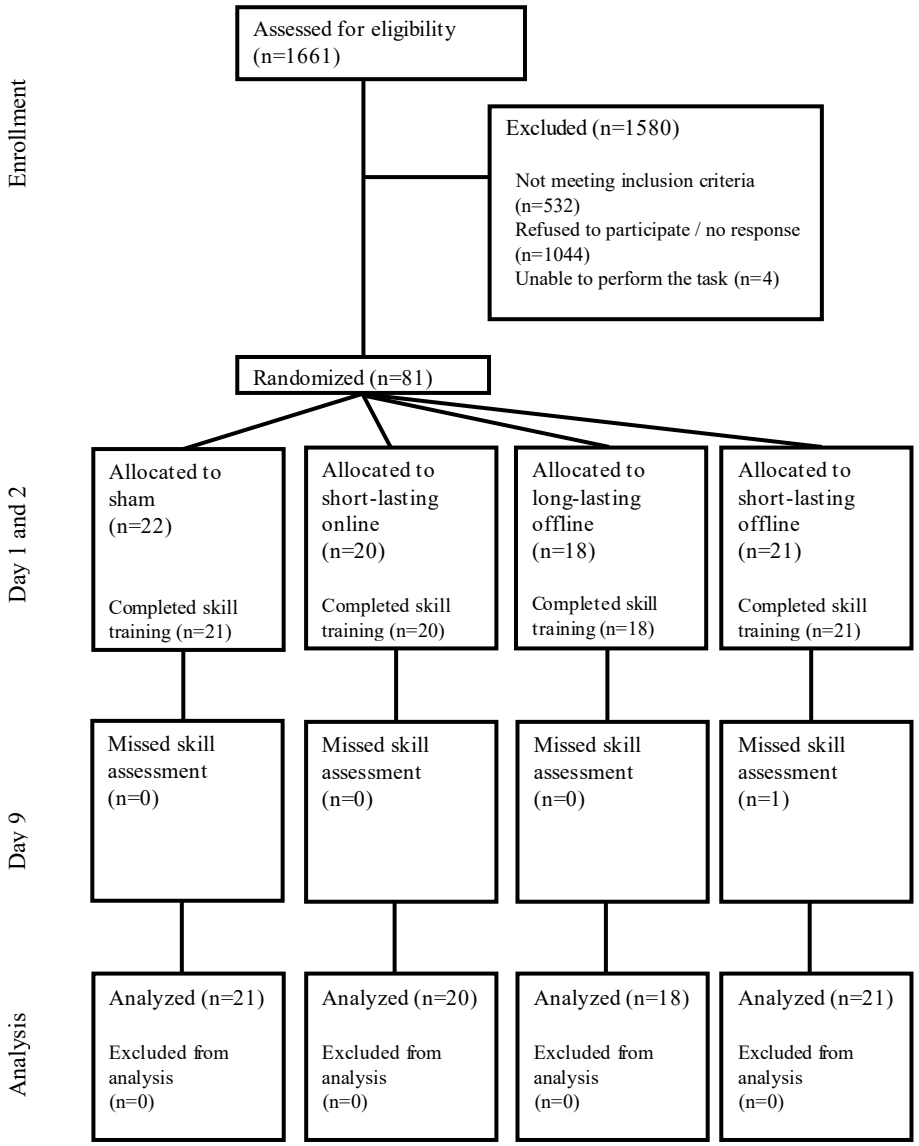


Figure 3. Patient inclusion flow diagram. Out of a total sample of 1661 screened chronic stroke patients, we invited 85 patients to participate. We excluded an additional four patients who were unable to control the computer mouse with their affected hand leaving 81 patients for randomization. One subject could not attend the last session but was kept in the analysis. A single subject quitted the study after the first day without notice and was removed from further analysis.

As secondary analyses, we investigated with a similar adjusted (independent variables: tDCS group, $BDNF_{Met}$, age, stroke dominance, gender and Fugl-Meyer assessment of the upper extremity) and unadjusted (independent variable: tDCS group) linear regression model the effects of tDCS on manual dexterity (ΔPPT) and maximum grip force (ΔmaxGF) of the affected hand. This model includes a single intercept for the entire patient group ("group training effect").

Finally, we calculated the Akaike information criterion to investigate whether our patient population could be separated in responders and non-responders. The Akaike information criterion indicates how likely a model is, corrected for the number of free parameters [40]. More likely models have a lower Akaike information criterion. We compared a null model and a responder / non-responder model using the mean score in R1 and R2 for patients in the sham and short-lasting online group. The null model, which assumes no effect for all patients, consisted of a single normal distribution. The responder / non-responder model, which assumes a population of responders and non-responders, was composed of two separate normal distributions with a shared standard deviation. Patients in the sham group were constrained to the non-responder group as they could not respond to tDCS.

	Sham (n=21)	**Short-lasting online (n=20)**	**Long-lasting offline (n=18)**	**Short-lasting offline (n=21)**
Demographic				
Age – yr (M±SD)	62±11	64±11	59±9	60±8
Male sex – (%)	67	75	67	40
Hand				
Right hand dominance –(%)	86	95	78	86
Paresis in right hand – (%)	52	45	39	57
Paresis in dominant hand – (%)	52	60	61	57
Fugl-Meyer assessment of the upper extremity	55±15	58±10	51±17	57±12
Action research arm test	50±14	52±10	45±19	50±13
Purdue pegboard test affected hand – pins	6.8±3	6.2±4	5.8±4	6.7±4
Maximum grip force affected hand – N	267±106	280±104	212±113	226±123
Genetics				
BDNF >= 1 Met allele (%)	28	58	29	45
Stroke				
Time post stroke – yrs.	4±5	3±3	2±2	2±2

Table 1. Characteristics of the study population with chronic stroke, N = 80.

Markov chain Monte Carlo simulation was performed with freely available software (Openbugs version 3.2.3, Openbugs Foundation). Results are presented as the mean of the posterior distribution with 95% highest density intervals (95%HDI). These intervals are the Bayesian equivalent to 95% confidence intervals. Any regression coefficient estimated to have a 95%HDI not containing a zero was considered statistically significant. Summaries of individual patient data are reported as group medians with interquartile ranges.

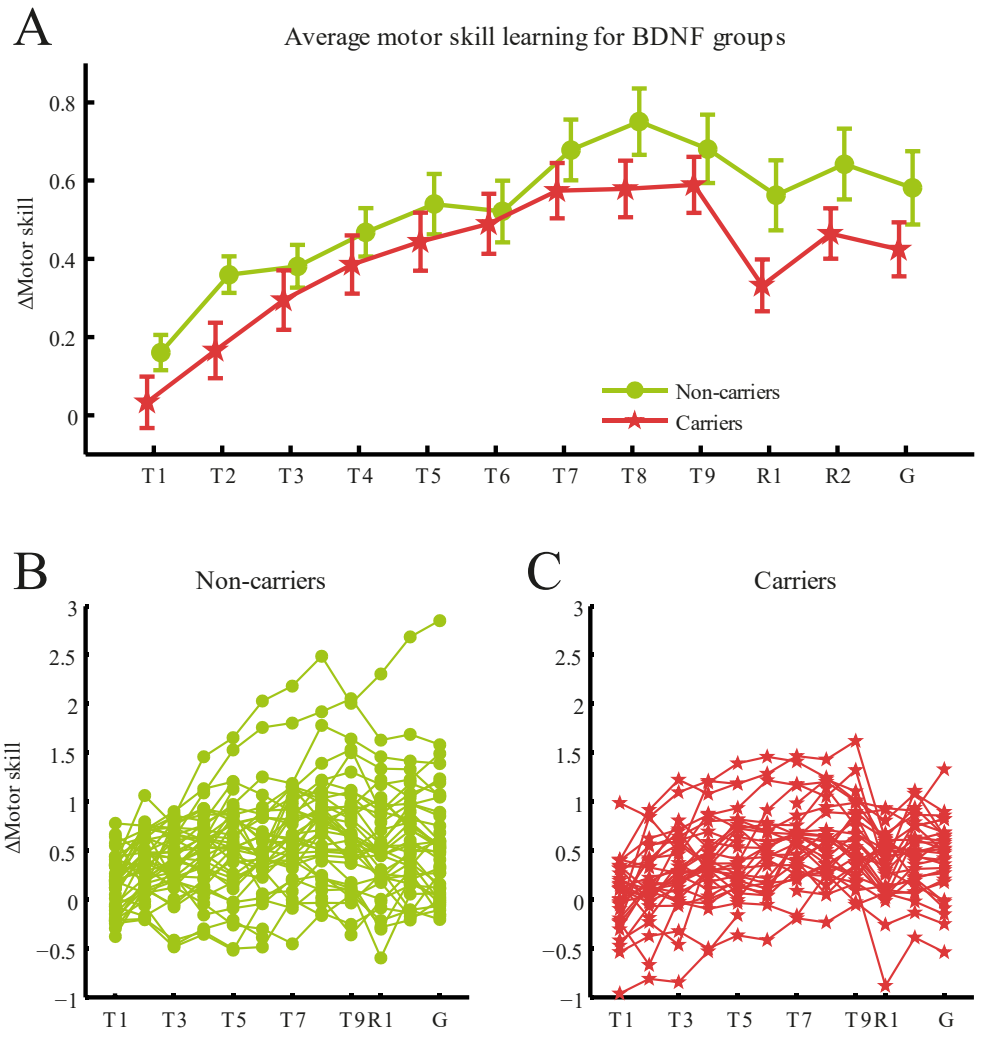


Figure 4. Effect of BDNF Val66Met on ΔMotor skill. A. Average motor skill learning for the BDNF Val66Met groups (M±SEM). Non-carriers have two Val alleles, carriers have at least one Met allele. ΔMotor skill is shown for day two and day nine of the study. Results on day nine were compared between BDNF Val66Met groups to investigate the role of the polymorphism in motor skill learning. BDNF Val66Met was found to negatively affect motor skill learning. **B-C.** Individual motor skill learning curves for non-carriers (**B**) and carriers (**C**).

Dependent variable	Independent variable	$\beta_{adjusted}$	$95\%HDI_{adjusted}$	$\beta_{unadjusted}$	$95\%HDI_{unadjusted}$
ΔMotor skill	Group training effect	0.467	0.0493 0.877	0.511	0.280 0.727
	Short-lasting online	0.0908	-0.227 0.403	0.0332	-0.284 0.350
	Long-lasting offline	0.0242	-0.292 0.349	0.0451	-0.280 0.372
	Short-lasting offline	-0.108	-0.433 0.210	-0.0706	-0.394 0.243
	Generalization	-0.00041	-0.0496 0.0514	-0.00035	-0.0521 0.0492
	$BDNF_{Met}$	-0.217	-0.431 -0.0116		
ΔPPT (#)	Group training effect	0.460	-0.364 1.272	0.511	0.280 0.727
	Short-lasting online	0.249	-0.379 0.88	0.0332	-0.284 0.351
	Long-lasting offline	-0.138	-0.786 0.489	0.0451	-0.2796 0.372
	Short-lasting offline	0.188	-0.449 0.826	0.270	-0.357 0.901
ΔMaxGF (N)	Group training effect	1.350	-6.180 8.9909	3.254	-0.670 7.187
	Short-lasting online	-2.044	-7.858 3.839	-1.166	-6.796 4.452
	Long-lasting offline	1.570	-4.664 7.440	2.058	-3.891 7.849
	Short-lasting offline	-4.170	-10.467 1.825	-4.411	-10.164 1.079

Table 2. Linear regression models for ΔMotor skill, ΔPPT and ΔMaxGF. The coefficients of an adjusted and an unadjusted model are presented as the mean with 95%HDIs. The adjusted model includes the dependent variables listed in the table as well as stroke dominance, gender, age and FMA-UE score, and an intercept ("group training effect"). Significant effects are printed in bold. BDNF = Brain-derived neurotrophic factor; HDI = highest density interval; MaxGF = maximum grip force; PPT = Purdue pegboard test.

Results

Patients

Patients were enrolled from February 2015 to May 2016 (see Figure 3). Out of a total screening sample of 1661 chronic stroke patients, we invited 85 patients to the rehabilitation center. We excluded an additional four patients who were unable to control the computer mouse with their affected hand, leaving 81 patients for randomization. One patient was unable to attend the last session but was included in the analysis. Another patient who quitted the study after the first day without notice was removed from further analysis. BDNF genotyping failed in eight out of 80 patients.

Our treatment groups were comparable in demographic, performance and stroke characteristics and representative of a mild to moderately affected chronic stroke group (see Table 1). Forty-two patients were genotyped as non-carriers (two Val alleles) whereas 30 patients were genotyped as carriers (at least one Met allele).

Motor skill learning

Overall, patients improved on the circuit tracing task with training (see Figures 4 and 5 and Table 2). 70 out of 80 patients had a positive ΔMotor skill during R1 and R2 (all patients: ΔMotor skill median=0.494 IQR=[0.194 - 0.772]). Additionally, we found a "group training effect" in both the unadjusted and adjusted model (see Table 2). Motor skill also generalized to an untrained circuit, as the median ΔMotor skill did not differ between the retention blocks (R1 and R2) and G1 (G1-

(R1+R2)/2 = -0.0278 with IQR=[-0.123 0.107]). In line with this result, both the adjusted and unadjusted model indicated no effect of the generalization circuit on ΔMotor skill (see Table 2).

Having at least one BDNF Met allele was found to negatively affect ΔMotor skill (see Figure 4 and Table 2). This difference was already visible on the second day during T1 (non-carriers median=0.180 IQR=[-0.0689 – 0.355]; carriers median=0.0052 IQR=[-0.234 0.316]) and became most apparent one week later during R1 (non-carriers median=0.506 IQR=[0.193 – 0.902]; carriers median=0.347 IQR=[0.0644 0.589]). No effect of tDCS on ΔMotor skill was found (see Figure 5 and Table 2).

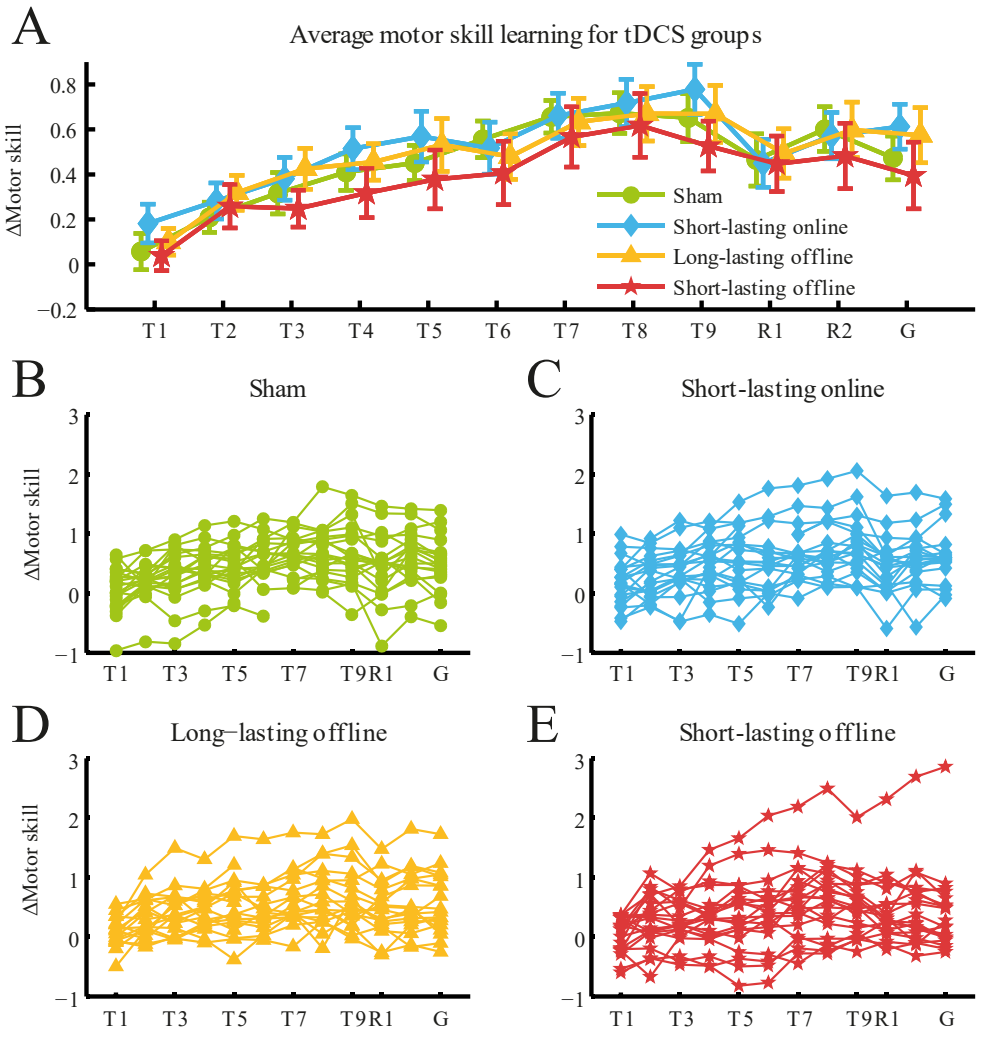


Figure 5. Effect of tDCS on ΔMotor skill. A. ΔMotor skill for the four tDCS groups (M±SEM). ΔMotor skill is shown for day two (training (T) block 1 to 9) and day nine (retention (R) block 1-2 and generalization (G)) of the study. Results on day nine were compared between tDCS groups to quantify tDCS motor skill learning effects. No effect of any tDCS group compared to sham on motor skill learning was found. **B-E.** Individual motor skill learning curves for chronic stroke patients in the sham (**B**), short-lasting online (**C**), long-lasting offline (**D**) or short-lasting offline group (**E**).

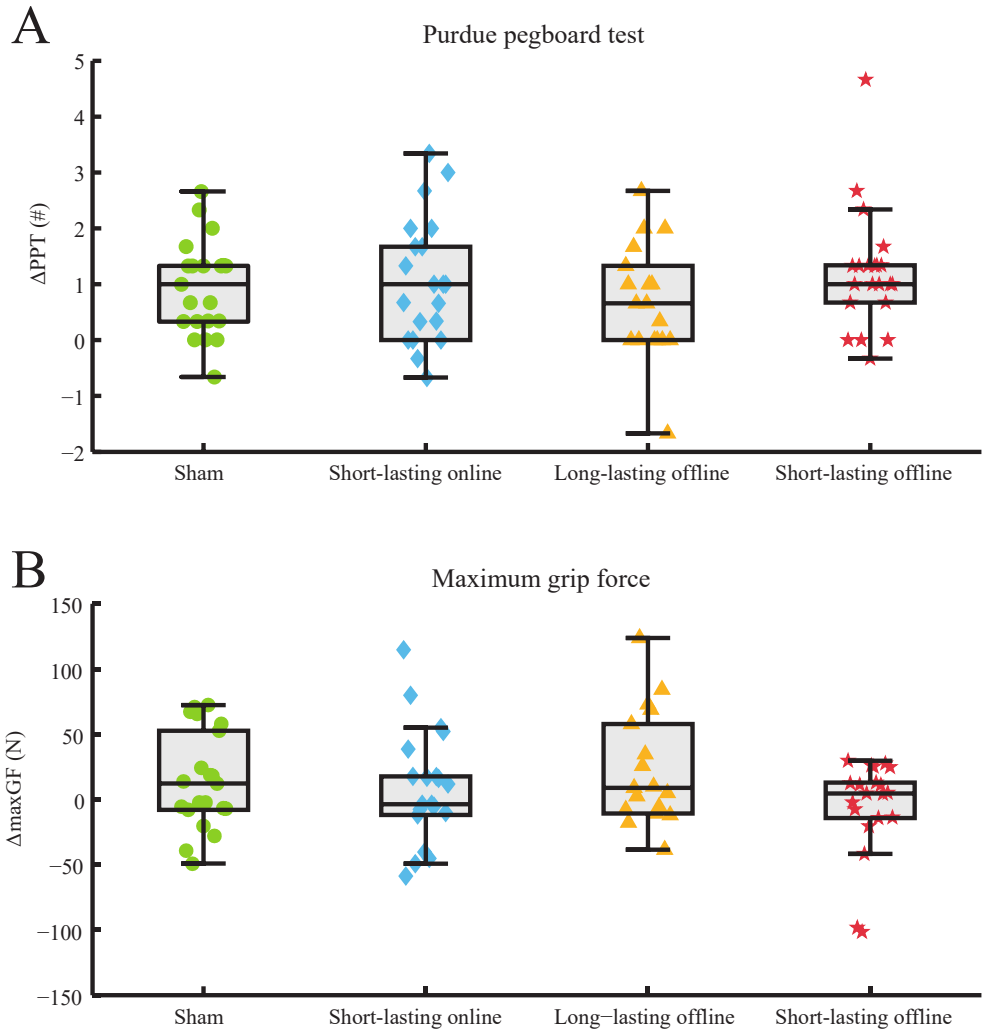


Figure 6. Effect of tDCS on manual dexterity and maximum grip force. A. Whisker plot of ΔPPT for the different tDCS groups. **B.** Whisker plot of ΔMaxGF for the different tDCS groups.

tDCS effects on manual dexterity and maximum grip force

Patients improved manual dexterity but not maximum grip force as a result of training (see Figure 6 and Table 2). 60 out of 80 patients had a ΔPPT larger than 0 (all patients: ΔPPT change median=1 IQR=[0.33 1.33] pegs). The "group training effect" from the unadjusted model similarly indicated more pegs were placed with the affected hand (see Table 2). In contrast, maximum grip force increased in 44 out of 80 patients for the affected hand (all patients: ΔmaxGF change=7.0 IQR=[-8.0 27.6] N) without a group training effect (see Table 2). No effect of tDCS on manual dexterity or maximum grip force was found (see Figure 6 and Table 2).

Evidence for a null effect

We calculated from the parameter estimates of the adjusted model how likely the effects of tDCS on ΔMotor skill reported in earlier studies [9,11] would be in our patient population. The probability of a ΔMotor skill of at least 0.4 (corresponding to the smallest effect size reported in Lefebvre et al. [9,11]) in our study was 3% (Short-lasting online group), 0.9% (Long-lasting offline group) and 0.1% (Short-lasting offline group). Second, the null model with M=0.511 had an Akaike information criterion=54.22, while the responder / non-responder model with $M_{NonResponder}$=0.485 and $M_{Responder}$=0.529 had an Akaike information criterion=56.20. The lower Akaike information criterion for the null model provides evidence against the presence of responders and non-responders.

Discussion

In this study, we investigated the role of BDNF Val66Met and tDCS in motor skill learning after stroke [9,11]. We found that non-carriers (no Met alleles) outperformed carriers (at least one Met allele) on day nine of the study. This result indicates activity-dependent release of BDNF is important for motor skill learning after stroke. Second, our results showed that none of the tDCS protocols affected motor skill learning, manual dexterity (ΔPPT) or maximum grip force (ΔmaxGF).

Role of BDNF in motor skill learning after stroke

Our finding that carriers of the BDNF Val66Met polymorphism more slowly acquire a motor skill is consistent with results in healthy subjects. BDNF modulates long term potentiation and synaptic plasticity [41,42], which is important for motor skill learning through strengthening of horizontal connections in the motor cortex [23,24]. Activity-dependent release of BDNF is decreased in BDNF Val66Met carriers [27] which explains why carriers have impaired motor skill learning [8,28] and reduced motor map expansion after motor training [43]. Our results indicate that interventions which can increase BDNF release might improve rehabilitation after stroke. However, even though tDCS has been suggested to promote BDNF release [8], we do not find an effect of tDCS on motor skill learning in our study (in contrast to [9,11,14], see below for a discussion) and therefore cannot conclude that BDNF is a likely mediator of tDCS.

Screening for BDNF Val66Met after stroke might improve models of stroke recovery [44] as motor skill learning is thought to play an important role in recovery after stroke [45]. Stroke patients who carry the BDNF Val66Met polymorphism will (re)learn motor skills more slowly than non-carriers and are therefore intuitively expected to recover less well. However, it has also been proposed that carriers do not have an overall worse clinical recovery but rather a different recovery pattern relying more on subcortical mechanisms [46]. Adding genetic screening to clinical prediction models of stroke such as the PREP model could be of great clinical interest [44] to increase predictive accuracy and help personalize rehabilitation programs.

Why did tDCS not affect motor skill in the current study?

Our tDCS results are at odds with previous studies showing a favorable contribution of short-lasting online tDCS to motor skill learning using (1) the exact same paradigm in a similar chronic stroke population [9,11] and (2) comparable paradigms in healthy subjects [5–8] and chronic stroke patients [14]. This might be related to a high percentage of non-responders in our study or training repetition. In addition, we should not preclude the possibility that tDCS plays a limited role in motor skill learning.

Absence of a favorable effect of tDCS on motor skill learning might be explained by a high proportion of non-responders in our sample population. First, one rationale behind tDCS in stroke patients is restoration of the interhemispheric imbalance in cortical activity [21], which has been found in chronic stroke patients [22]. This view was recently challenged as more severely affected patients might rely on their unaffected hemisphere for motor performance with their affected hand [19]. According to this idea, tDCS should be used to increase activity of the affected hemisphere in mildly affected patients, and of the unaffected in severely affected patients, which we did not do. However, our population consisted of well-recovered stroke patients capable of voluntary movement, for whom the applied tDCS protocol should be optimal. In addition, our patients were comparable to Lefebvre et al. [9,11]. Second, tDCS effects on motor skill learning might depend critically on the strength and orientation of the electric field in the motor cortex. These electric field parameters will differ between individuals because of anatomy of the skull, position and folding of the motor cortex and lesion characteristics which could explain why only some patients benefit from tDCS. Since we do not have imaging data for individual patients, this explanation cannot be further investigated. Statistically however, we found no evidence for separate "responder" and "non-responder" groups. We used a mixture modeling approach to compare a model of a single group with a model separating responders and non-responders in the sham and short-lasting online tDCS groups. The Akaike information criterion indicated the null model was a better description of the data than the responder / non-responder model, arguing against tDCS responders and non-responders in our sample.

A second hypothesis is that tDCS needs multiple days of training to develop its full effect on motor skill learning. However, all previous multiple day motor skill learning studies using motor sequence learning or a tracing task found a difference in motor performance already present at the start of the second training session [5–8,10,37]. In addition, our experimental design was identical to Lefebvre et al. [9,11] who did find favorable results after a single training session. It is thus unlikely that this hypothesis could explain the observed results.

Finally, it may be that tDCS has a limited contribution to motor skill learning. We calculated that the probability of a 0.4 ΔMotor skill improvement was smaller than 5% for all the tDCS groups, indicating large group differences found in earlier studies [9,11] are unlikely in our patient sample. Possibly, the effectiveness of tDCS in its contemporary form is lower than previously assumed. For successful neuromodulation in stroke patients, it might be necessary to characterize patients more carefully with for example transcranial magnetic stimulation [47], MRI diffuse tensor imaging [44,48], MRI spectroscopy [18,49] or EEG [50]. Research efforts in this direction will be important to develop tDCS into a reliable therapeutic intervention.

Limitations

Our experimental design did not include direct physiological measurements of cortical excitability. Furthermore, we did not collect imaging data of stroke lesions. We were therefore not able to correct for differences in lesion volume or location between patients. Finally, because we investigated the effects of tDCS on motor skill learning in the more stable chronic phase after stroke, in line with previous comparable studies [9,11,14], our results cannot be directly extrapolated to the subacute phase. Reorganizational processes in the first months after the lesion are not present in the chronic phase [19] and might be affected by tDCS in the subacute phase after stroke [1,2].

References

1. Khedr, E.M. et al. *Neurorehabil. Neural Repair* 27, 592–601 (2013).
2. Allman, C. et al. *Sci. Transl. Med.* 8, (2016).
3. Fusco, A. et al. *Restor. Neurol. Neurosci.* 32, 301–12 (2014).
4. Lindenberg, R., Renga, V., Zhu, L.L., Nair, D. & Schlaug, G. *Neurology* 75, 2176–84 (2010).
5. Waters-Metenier, S., Husain, M., Wiestler, T. & Diedrichsen, J. *J. Neurosci.* 34, 1037–50 (2014).
6. Prichard, G., Weiller, C., Fritsch, B. & Reis, J. *Brain Stimul.* 7, 532–40 (2014).
7. Reis, J. et al. *Proc Natl Acad Sci U S A* 106, 1590–1595 (2009).
8. Fritsch, B. et al. *Neuron* 66, 198–204 (2010).
9. Lefebvre, S. et al. *Front. Hum. Neurosci.* 6, 343 (2012).
10. Zimerman, M. et al. *Ann. Neurol.* 73, 10–5 (2013).
11. Lefebvre, S. et al. *Brain* 138, 149–63 (2015).
12. Vines, B.W., Cerruti, C. & Schlaug, G. *BMC Neurosci.* 9, 103 (2008).
13. Sriraman, A., Oishi, T. & Madhavan, S. *Brain Res.* 1581, 23–9 (2014).
14. Zimerman, M. et al. *Stroke* 43, 2185–2191 (2012).
15. Stagg, C.J. et al. *J. Neurosci.* 29, 5202–6 (2009).
16. Bachtiar, V. et al. *Curr. Biol.* 4, 1023–1027 (2015).
17. Nitsche, M.A. et al. *J Physiol* 553, 293–301 (2003).
18. Stagg, C.J., Bachtiar, V. & Johansen-Berg, H. *Curr Biol* 21, 480–484 (2011).
19. Di Pino, G. et al. *Nat. Rev. Neurol.* 10, 597–608 (2014).
20. Di Lazzaro, V. et al. *Brain Stimul.* 7, 841–848 (2014).
21. Ward, N.S. & Cohen, L.G. *Arch. Neurol.* 61, 1844–8 (2004).
22. Murase, N., Duque, J., Mazzocchio, R. & Cohen, L.G. *Ann. Neurol.* 55, 400–9 (2004).
23. Rioult-Pedotti, M.S., Friedman, D., Hess, G. & Donoghue, J.P. *Nat. Neurosci.* 1, 230–4 (1998).
24. Rioult-Pedotti, M.S., Friedman, D. & Donoghue, J.P. *Science (80-.).* 290, 533–536 (2000).
25. Cargill, M. et al. *Nat. Genet.* 22, 231–8 (1999).
26. Shimizu, E., Hashimoto, K. & Iyo, M. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 126B, 122–3 (2004).
27. Egan, M.F. et al. *Cell* 112, 257–269 (2003).
28. McHughen, S.A. et al. *Cereb. Cortex* 20, 1254–62 (2010).
29. Kang, N., Summers, J.J. & Cauraugh, J.H. *J. Neurol. Neurosurg. Psychiatry* 87, 345–55 (2016).
30. Nitsche, M.A. & Paulus, W. *J Physiol* 527 Pt 3, 633–639 (2000).
31. Nitsche, M.A. & Paulus, W. *Neurology* 57, 1899–1901 (2001).
32. Monte-Silva, K. et al. *Brain Stimul.* 6, 424–32 (2013).
33. Bastani, A. & Jaberzadeh, S. *Clin. Neurophysiol.* (2014).doi:10.1016/j.clinph.2014.01.010
34. Rocha, S. et al. *Disabil. Rehabil.* 38, 653–660 (2016).
35. Brunnstrom, S. *Phys. Ther.* 46, 357–75 (1966).
36. Lyle, R.C. *Int. J. Rehabil. Res.* 4, 483–92 (1981).
37. Reis, J. et al. *Cereb. Cortex* bht208- (2013).doi:10.1093/cercor/bht208
38. Gandiga, P.C., Hummel, F.C. & Cohen, L.G. *Clin Neurophysiol* 117, 845–850 (2006).
39. Scott, N.W., McPherson, G.C., Ramsay, C.R. & Campbell, M.K. *Control. Clin. Trials* 23, 662–74 (2002).
40. Akaike, H. *Int. Encycl. Stat. Sci.* 25–25 (2011).doi:10.1007/978-3-642-04898-2_110
41. Turrigiano, G.G. & Nelson, S.B. *Curr. Opin. Neurobiol.* 10, 358–364 (2000).
42. Park, H. & Poo, M. *Nat. Rev. Neurosci.* 14, 7–23 (2013).
43. Kleim, J.A. et al. *Nat. Neurosci.* 9, 735–7 (2006).
44. Stinear, C.M., Barber, P.A., Petoe, M., Anwar, S. & Byblow, W.D. *Brain* 135, 2527–35 (2012).
45. Krakauer, J.W. *Curr. Opin. Neurol.* 19, 84–90 (2006).
46. Di Pino, G. et al. *Neurorehabil. Neural Repair* 30, 3–8 (2016).
47. Wiethoff, S., Hamada, M. & Rothwell, J.C. *Brain Stimul.* 7, 468–475 (2014).
48. Stinear, C.M. et al. *Brain* 130, 170–180 (2006).
49. Kim, S., Stephenson, M.C., Morris, P.G. & Jackson, S.R. *Neuroimage* 99, 237–43 (2014).
50. Polanía, R., Nitsche, M.A. & Paulus, W. *Hum. Brain Mapp.* 32, 1236–49 (2011).

4.5 Long-lasting offline tDCS for upper extremity motor rehabilitation in the subacute phase after stroke: double-blind, randomized clinical trial

Rick van der Vliet, Zeb D. Jonker, Maarten A. Frens, Ruud W. Selles and Gerard M. Ribbers

Abstract

Background: Stroke is a common global health-care problem that is serious and disabling. A promising new tool in motor rehabilitation is transcranial direct current stimulation (tDCS), a safe, non-invasive technique that delivers low-intensity current to the scalp through a pair of electrodes. The most common approach to tDCS in clinical stroke rehabilitation trials has been to stimulate daily for 15-30 minutes either during (online) or just before (offline) rehabilitation training. However, a more efficient approach might be to utilize offline stimulation protocols that evoke longer-lasting increases in motor cortex excitability and therefore have the potential to support motor learning and rehabilitation beyond the stimulation sessions.

Objective: To investigate the effects of long-lasting tDCS on upper extremity motor recovery in first-ever, ischemic, subacute stroke patients.

Methods: In this parallel, placebo-controlled intervention trial with two arms, patients were assigned in a 1:1 ratio to either sham tDCS (placebo group) or long-lasting offline tDCS (intervention group). The regular upper extremity rehabilitation program (on Mondays, Wednesdays, and Fridays) was interspersed with tDCS (on Tuesday and Thursdays) for four weeks, adding up to a total eight stimulation sessions. The primary outcome was the Fugl-Meyer of the upper extremity (measured at baseline, 5, 8, and 12 weeks post-stroke, and 26 weeks for a subset of patients), which was analyzed with a longitudinal mixture model of FM-UE recovery to sensitively estimate the treatment effect over usual care. Secondary outcomes included (1) functional activity (action research arm test), (2) walking ability (10-meter walk test), (3) dependence in activities of daily living (Barthel index), (4) mood disorders (hospital anxiety and depression scale) at 12 weeks post-stroke.

Results: No effect of long-lasting tDCS on the upper extremity motor impairment was found. In addition, no differences were found in (1) action research arm test, (2) 10-meter walk test, (3) Barthel index, (4) hospital anxiety and depression scale. Adverse events were uncommon and comparable between treatment arms.

Conclusion: We found no evidence for the superiority of long-lasting offline tDCS over sham tDCS on upper limb recovery in the subacute phase after stroke. Based on this result we recommend future studies to (1) focus on online tDCS rather than offline tDCS, and (2) enroll larger patient populations for stroke severity subgroup analyses.

Introduction

Stroke is a common global health-care problem[1] that is serious and disabling.[2] However, as most patients with stroke survive the initial injury,[3] the largest effect on patients and families is usually through long-term impairment, limitation of activities (disability), and reduced participation (handicap).[4,5] Motor impairment after stroke, defined as a loss or limitation of function in muscle control or movement or a limitation in mobility,[6] typically affects the control of movement of the face, arm, and leg of one side of the body in about 80% of patients. [7,8] Therefore, much of the focus of stroke rehabilitation is on the recovery of movement and associated functions with high-intensity, repetitive task-specific practice.[7,8]

A promising new tool in motor rehabilitation is transcranial direct current stimulation (tDCS), which is a safe[9] and non-invasive technique that delivers low-intensity current to the scalp through a pair of electrodes.[10,11] Depending on the polarity of the electrodes and the spatial orientation of the underlying neurons,[12,13] direct current was found to alter the excitability of the motor cortex, as measured with transcranial magnetic stimulation, for approximately an hour.[14–16] Since then, tDCS has been reported to improve motor skill learning in healthy subjects[17–24] and chronic stroke patients,[25,26] and upper limb rehabilitation in subacute and chronic stroke patients with moderately severe cortical damage.[27–31] Presumably, tDCS effects result from releasing brain-derived neurotrophic factor[20], down-regulating GABA[32–35] and restoring the interhemispheric imbalance between the affected motor cortex and the unaffected motor cortex.[36–39] However, more research is still needed to establish the clinical relevance of tDCS for upper limb rehabilitation as well as to establish the optimal protocol for stimulation.

The most common approach to tDCS in clinical stroke rehabilitation trials has been to stimulate daily for 15-30 minutes either during (online) or just before (offline) rehabilitation training.[27–31] However, a more efficient approach is offline stimulation protocols, which evoke longer-lasting increases in motor cortex excitability[40,41] and therefore have the potential to support motor learning and rehabilitation beyond the stimulation sessions. Central to long-lasting stimulation protocols is the introduction of a short break in between two brief stimulation sessions, which induce after-effects for up to two days (termed late LTP-like plasticity),[40,41] rather than the 30-60 minutes found for conventional continuous stimulation. Long-lasting stimulation could, therefore, facilitate distributed practice with high task variability, thus complying with fundamental insights on robust motor learning [6,42] and minimize patient discomfort as well as the demand for stimulation equipment.

In this study, we investigate the effects of a long-lasting tDCS protocol on upper extremity motor recovery in subacute stroke patients. The primary outcome is upper extremity motor impairment, as measured with the Fugl-Meyer assessment of the upper extremity (FM-UE) in the subacute phase after stroke. Secondary outcomes include (1) upper limb capacity (action research arm test), (2) walking ability (10-meter walk test), (3) dependence in activities of daily living (Barthel index), (4) mood disorders (hospital anxiety and depression scale) and (5) adverse events.

Materials and methods

Patients

The trial was conducted at the Rijndam Rehabilitation Centre (Rotterdam, The Netherlands), which offers inpatient and outpatient stroke rehabilitation. The recruitment period ran from January 2015 until July 2019. Patient inclusion criteria were: acute hemiparesis with a first-ever, non-hemorrhagic infarction documented by a neurologist, no more than four weeks post-stroke

at start of the study, between the ages of 18 and 79. Exclusion criteria were: absence of voluntary movement of the affected upper extremity, head injury or the presence of intracranial metal or intracranial lesions, history of cranial irradiation, history of epilepsy, presence of a pacemaker, taking anticonvulsant or neuroleptic medication, substance abuse, and the inability to understand instructions.

Randomization and blinding

In this parallel, placebo-controlled intervention trial with two arms, patients were assigned in a 1:1 ratio to either sham tDCS (placebo group) or long-lasting offline tDCS (intervention group) using a minimization approach. Minimization is a randomization technique designed for relatively small studies which aims to balance predefined patient characteristics and group sizes by adapting the allocation probability of groups. This way, group characteristics and size can be better controlled.[43] We chose to minimize the difference between groups in the baseline FM-UE (dichotomized: low<33 points; high>=33 points), stroke laterality (dominant/non-dominant hand) and age (dichotomized: low<60 years; high>=60 years). Minimization was implemented in MinimPy (freely available from: https://sourceforge.net/projects/minimpy/).

Treatment

All patients received standard upper extremity rehabilitation treatment according to the Dutch rehabilitation guidelines[44,45] twice a day on Mondays, Wednesdays, and Fridays. In addition, participants were administered a total eight sessions of tDCS on Tuesday and Thursdays spread over four weeks.

tDCS was applied using a Starstim device (Neuroelectrics, Barcelona, Spain) with 5 cm diameter sponge electrodes. The scalp areas overlying the motor cortices were localized with the International 10/20 Electroencephalogram System and the identified area was thoroughly cleaned with Nuprep (Weaver and Company, Aurora, Colorado, United States). The positive sponge electrode was placed over the affected hemisphere and the negative sponge electrode was placed over the unaffected hemisphere with an elastic headband (bihemispheric montage). We used bihemispheric tDCS because it (1) may have a larger effect on motor skill learning than unihemispheric tDCS[22,46] and (2) and has been previously-used in tDCS trials with stroke patients.[25,26,30] Stimulation intensity was set to 1mA, which is also in agreement with many tDCS studies.[18–21,28,40,41] The long-lasting protocol was based on work by Monte-Silva et al.[40] and consisted of two stimulation blocks of 10 minutes separated by a 25-minute break. Both periods included a 15 s ramp-up and a 15 s ramp-down period to ensure comfort. In the sham protocol, only the ramp up and ramp down period was provided with stimulation bursts that evoke a sensation similar to more prolonged tDCS sessions, thus ensuring blinding without affecting excitability.[47] The tDCS was controlled from a PC according to the patient's tDCS group. The program only showed day and group numbers without protocol names to blind the experimenter and patient from the protocol.

Outcomes

Measurements were performed at baseline (within four weeks post-stroke), and 5, 8, and 12 weeks post-stroke and included the Fugl-Meyer assessment of the upper extremity[48] (FM-UE; baseline, 5, 8, and 12 weeks post-stroke), the Action Research Arm Test[49] (ARAT; 12 weeks post-stroke), 10-meter walk test (12 weeks post-stroke), Barthel index[50] (baseline and 12 weeks post-stroke) and the hospital anxiety and depression scale (baseline and 12 weeks post-stroke). The

FM-UE difference between inclusion and 12 weeks post-stroke (ΔFM-UE) was a secondary outcome. A subset of patients was also measured at 26 weeks post-stroke because they participated in the PROFITS study. The 26 weeks FM-UE measurements of these patients were included in our analysis. Patients were assessed for adverse events (new-onset convulsion, central pain, mood disorder or headache) at 12 weeks post-stroke.

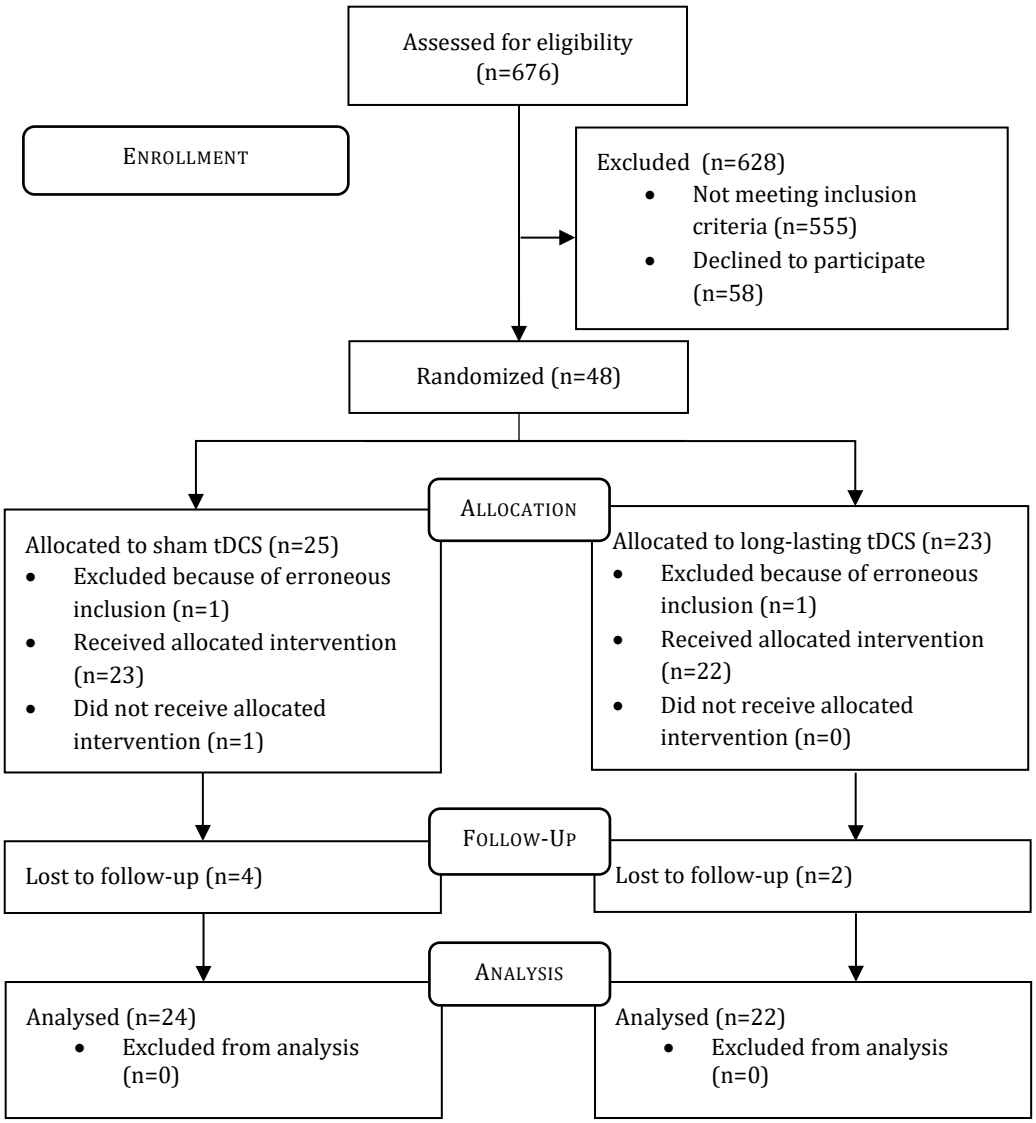


Figure 1. Flow chart.

Statistical methods

Statistical analysis of the primary outcomes (FM-UE) was based on a recently developed longitudinal mixture model of spontaneous neurological recovery after stroke. Details of this approach can be found in van der Vliet et al. In short, FM-UE at different time points post-stroke is modeled as five distinct subgroups which recover over time with a specific rate up until a fixed proportion of their potential recovery (66-baseline FM-UE) plus a study effect for the usual care (sham tDCS) and the intervention care (long-lasting tDCS) groups. This approach allows estimating treatment effects for all patients combined or for the poor (subgroup one), moderate (subgroups two and three) and good (subgroups four and five) FM-UE recovery clusters separately. For this study, we chose to estimate overall study effects rather than cluster-specific effects. Output of the model, therefore, includes the overall study effects for the sham tDCS group and the long-lasting tDCS group as well as the invention effect, which is obtained by subtracting the study effect of the sham tDCS group from the study effect of the intervention group. As shown precisely,[51] advantages of this approach include the increased study power compared to cross-sectional approaches, the incorporation of the exact timing of measurements after stroke, the handling of missing values and robustness to outlier data.

An intention-to-treat approach was used for the analysis of this randomized-controlled trials. Once randomized, each patient was analyzed in the group they were assigned to, independent of potential drop-out or compliance to the protocol. The only reason for post-randomization exclusion was erroneous inclusion.

The baseline demographics, clinical characteristics, and secondary outcomes (ΔFM-UE, ARAT, 10-meter walk test, Barthel index, and the HADS) were analyzed with independent sample t-tests for continuous, normally-distributed outcomes, with Mann-Whitney U tests for continuous non-normal data and with Fisher exact tests for proportions. Normality was checked for with the Kolmogorov-Smirnov test. The threshold for statistical significance was set to 0.05 for all analyses.

Sample size

We based our power calculation on a clinically-important difference for the FM-UE, which has been estimated at 5.25 points,[52] and four repeated measurements (3, 5, 8 and 12 weeks after stroke), with a residual error of 3.9.[51,53] We calculated 42 patients should be sufficient to obtain 85% power. Accounting for a drop-out rate of 10%, we aimed to include 48 patients in total.

Results

Between January 2015 until July 2019, 676 patients were screened and 48 patients were enrolled. Twenty-five individuals were randomized to sham tDCS and 23 patients to long-lasting tDCS (Figure 1). Two patients were excluded from the study after inclusion because they were erroneously enrolled. One individual suffered from a hemorrhagic stroke rather than an ischemic stroke and the other individual was diagnosed with a functional neurological disorder rather than an ischemic stroke. Out of the remaining 46 patients, six patients were lost to follow-up at 12 weeks. Therefore, FM-UE measurements of 40 patients were available at 12 weeks post-stroke.

Baseline demographics and clinical characteristics were similar between the two treatment arms (see Table 1). Our typical patient was close to 60 years of age, male and right-handed. Stroke risk factors were often found, with hyperlipidemia being the most common. Average stroke severity was mild to moderate, as can be deduced from the Bamford scale, the NIHSS and the baseline FM-UE.[54] Only a minority of patients was treated with either thrombolysis

Variable	Sham tDCS	Long-lasting tDCS	*p*
Patients no.	24	22	
Age M (SD)	57.8 (10.6)	58.8 (12.5)	0.755
Male no. (%)	18 (75.0)	20 (90.9)	0.247
Right handed no. (%)	19 (79.2)	20 (90.9)	0.418
Vascular risk factors			
Diabetes no. (%)	3 (12.5)	6 (27.3)	0.276
Hypertension no. (%)	17 (70.8)	18 (81.8)	0.497
Dyslipidemia no. (%)	22 (91.7)	18 (81.8)	0.405
Current smoker no. (%)	11 (45.8)	9 (40.9)	0.774
Heart attack no. (%)	0 (0.0)	3 (13.6)	0.101
Atrial fibrillation no. (%)	7 (29.2)	5 (22.7)	0.742
Stroke characteristics			
Lacunar infarcts no. (%)	6 (25.0)	12 (54.5)	0.116
Partial anterior circulation infarcts no. (%)	16 (66.7)	9 (40.9)	
Total anterior circulation infarcts no. (%)	2 (8.3)	1 (4.5)	
Right side affected no. (%)	13 (54.2)	11 (50.0)	1.000
Dominant side affected no. (%)	12 (50.0)	13 (59.1)	0.568
Baseline stroke severity			
FM-UE M (SD)	26.4 (19.6)	31.2 (20.9)	0.421
NIHSS M (SD)	6.5 (4.7)	6.2 (2.5)	0.790
Barthel index M (SD)	14.2 (4.8)	15.9 (4.3)	0.213
MoCA M (SD)	21.8 (4.7)	23.0 (4.1)	0.389
HADS M (SD)	5.4 (4.8)	7.0 (4.3)	0.597
Thrombolysis no. (%)	3 (12.5)	6 (27.3)	0.276
Thrombectomy no. (%)	1 (4.2)	1 (4.5)	0.736
Intervention details			
Days from stroke to treatment M (SD)	21.6 (8.0)	21.3 (9.4)	0.918
Number of treatments M (SD)	6.8 (1.9)	7.3 (1.5)	0.194

Table 1. Baseline demographics and clinical characteristics. Abbreviations: FM-UE, Fugl-Meyer assessment of the upper extremity; HADS, hospital anxiety and depression scale; MoCA, Montreal cognitive assessment; NIHSS: National institute of healthy stroke scale.

or thrombectomy. Most patients (30/46) underwent all stimulation sessions and 40 patients missed no more than two treatments. Out of the six patients missing at least three tDCS sessions, one patient withdrew from the intervention because of head injury before the first session, one patient backed out of the study because of headache, and the other four were discharged early from the rehabilitation centre.

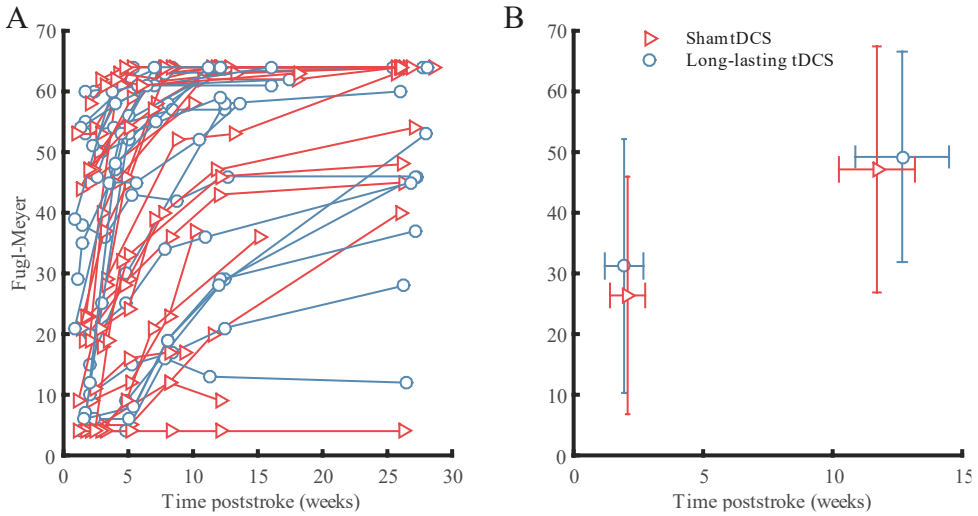


Figure 2. Fugl-Meyer upper extremity measurements.

No effect of long-lasting tDCS on the upper extremity motor impairment was found (see Table 2). Both the sham tDCS group (3.4 95%ETI=[1.3 5.3]) and the long-lasting tDCS group (1.9 95%ETI=[0.5 3.5]) performed slightly better than the reference group for spontaneous recovery, but the long-lasting tDCS group did not outperform the sham tDCS group (-1.5 95%ETI=[-3.8 1.2]). In addition, no differences were found in any of the secondary outcomes, which measured (1) the difference in FM-UE between inclusion and 12 weeks post-stroke, (2) functional activity (Action Research Arm Test), (3) walking ability (10-meter walk test), (4) dependence in activities of daily living (Barthel Index), and (5) mood disorders (Hospital Anxiety and Depression Scale) (see Table 3). Finally, adverse events occurred rarely, and the rates did not differ between both treatment arms (see Table 4).

Variable	Mean (95%ETI)
Sham tDCS	3.4 (1.3 to 5.3)
Long-lasting tDCS	1.9 (0.5 to 3.5)
Intervention effect	-1.5 (-3.8 to 1.2)
Residual error standard deviation	3.0 (2.3 to 3.9)
Degrees of freedom	2.3 (1.4 to 3.9)

Table 2. Primary outcome.

Discussion

In this study, we investigated the effects of long-lasting offline tDCS on upper extremity motor recovery in 48 subacute, ischemic, stroke patients. No difference was found between the sham tDCS and long-lasting offline tDCS groups in motor impairment (FM-UE) or in any of the secondary outcomes (functional activity, walking ability (10-meter walk test), dependence in activities of daily living (Barthel index), and mood disorders). Adverse events were uncommon and unrelated to the treatment modality.

Scale	Sham tDCS	Long-lasting tDCS	95%CI	*p*
Endpoint Fugl-Meyer	47.1 (20.3)	49.2 (17.3)	0.0 (-7.0 to 11.0)	0.966
ΔFugl-Meyer	22.0 (15.3)	21.9 (14.8)	-0.1 (-9.8 to 9.7)	0.991
ARAT M (SD)	37.3 (22.2)	35.2 (24.6)	0.0 (-11.0 to 9.0)	0.988
10-meter walk M (SD)	13.0 (7.8)	9.9 (4.1)	-1.0 (-4.9 to 0.9)	0.343
Barthel M (SD)	19.8 (0.5)	19.5 (1.9)	0.0 (0.0 to 0.0)	0.559
HADS M (SD)	8.1 (6.7)	5.1 (5.6)	-2.9 (-7.3 to 1.4)	0.181

Table 3. Secondary outcomes. Abbreviations: ARAT: action research arm test; HADS, hospital anxiety and depression scale.

The non-superiority of long-lasting offline tDCS compared to sham tDCS agrees with a previous study that found no benefit of a single session of long-lasting offline tDCS to motor learning in chronic stroke patients.[55] Together, these studies, therefore, do not support application of long-lasting offline tDCS for promoting motor learning or rehabilitation outcomes after stroke. However, no conclusions on the relevance of online tDCS protocols, which apply stimulation during training, for rehabilitation after stroke can be drawn from our current results. It is possible that online tDCS has a stronger effect on rehabilitation when stimulation is concurrent with training-related neural activity,[20] although increased motor cortex excitability after tDCS has been specifically found in people at rest.[14–16] Indeed, improvement of motor impairment with online tDCS has been found in populations of subacute[27] and chronic stroke patients,[28–31] even though these positive results have also not been consistently replicated[56,57] and the overall evidence is still relatively low.[58,59]

Limitations of this study include the lack of physiological markers of motor cortex excitability, which could have been obtained through transcranial magnetic stimulation. These measurements of motor cortex excitability might have helped in attributing the null effect to either the absence of neuromodulation or the lack of translation from neuromodulation to clinical effect. [14–16] Second, our sample size was based on the detection of an overall intervention effect across recovery subgroups and was therefore insufficient for estimating separate intervention effect in the poor (subgroup 1), moderate (subgroups 2 and 3) or good (subgroups 4 and 5) FM-UE recovery groups. Third, we did not have MRI brain scans of our patients, which restricts the exact description of stroke lesion volumes and location. Finally, generalizability of our results is restricted to first-ever, ischemic, subacute stroke patients.

In conclusion, we found no evidence for the superiority of long-lasting offline tDCS over sham tDCS on motor recovery in the subacute phase after stroke. Based on this result we recommend future studies to (1) focus on online tDCS rather than offline tDCS, and (2) enroll larger patient populations for stroke severity cluster analyses.

Adverse event	Sham tDCS	Long-lasting tDCS	*p*
Convulsion no. (%)	0 (0.0)	0 (0.0)	1.000
Central pain no. (%)	1 (4.2)	0 (0.0)	1.000
Mood disorder no. (%)	2 (8.3)	1 (4.5)	1.000
Headache no. (%)	1 (4.2)	0 (0.0)	1.000

Table 4. Adverse events.

References

1. Feigin, V.L. et al. *Lancet (London, England)* 383, 245–54 (2014).
2. Hankey, G.J. *Lancet (London, England)* 389, 641–654 (2017).
3. Davenport, R.J., Dennis, M.S., Wellwood, I. & Warlow, C.P. *Stroke* 27, 415–420 (1996).
4. Luengo-Fernandez, R. et al. *Stroke* 44, 2854–2861 (2013).
5. Poon, M.T.C., Fonville, A.F. & Al-Shahi Salman, R. *J. Neurol. Neurosurg. Psychiatry* 85, 660–667 (2014).
6. Krakauer, J.W. *Curr. Opin. Neurol.* 19, 84–90 (2006).
7. Langhorne, P., Coupar, F. & Pollock, A. *Lancet Neurol.* 8, 741–54 (2009).
8. Langhorne, P., Bernhardt, J. & Kwakkel, G. *Lancet* 377, 1693–1702 (2011).
9. Poreisz, C., Boros, K., Antal, A. & Paulus, W. *Brain Res. Bull.* 72, 208–14 (2007).
10. Nitsche, M.A. et al. *Brain Stimul.* 1, 206–23 (2008).
11. Nitsche, M.A. & Paulus, W. *Restor. Neurol. Neurosci.* 29, 463–92 (2011).
12. Rahman, A. et al. *J. Physiol.* 591, 2563–78 (2013).
13. Radman, T., Ramos, R.L., Brumberg, J.C. & Bikson, M. *Brain Stimul.* 2, 215–28, 228.e1–3 (2009).
14. Nitsche, M.A. & Paulus, W. *J Physiol* 527 Pt 3, 633–639 (2000).
15. Nitsche, M.A. & Paulus, W. *Neurology* 57, 1899–1901 (2001).
16. Wiethoff, S., Hamada, M. & Rothwell, J.C. *Brain Stimul.* 7, 468–475 (2014).
17. Waters-Metenier, S., Husain, M., Wiestler, T. & Diedrichsen, J. *J. Neurosci.* 34, 1037–50 (2014).
18. Prichard, G., Weiller, C., Fritsch, B. & Reis, J. *Brain Stimul.* 7, 532–40 (2014).
19. Reis, J. et al. *Proc Natl Acad Sci U S A* 106, 1590–1595 (2009).
20. Fritsch, B. et al. *Neuron* 66, 198–204 (2010).
21. Zimerman, M. et al. *Ann. Neurol.* 73, 10–5 (2013).
22. Vines, B.W., Cerruti, C. & Schlaug, G. *BMC Neurosci.* 9, 103 (2008).
23. Sriraman, A., Oishi, T. & Madhavan, S. *Brain Res.* 1581, 23–9 (2014).
24. Zimerman, M. et al. *Stroke* 43, 2185–2191 (2012).
25. Lefebvre, S. et al. *Front. Hum. Neurosci.* 6, 343 (2012).
26. Lefebvre, S. et al. *Brain* 138, 149–63 (2015).
27. Khedr, E.M. et al. *Neurorehabil. Neural Repair* 27, 592–601 (2013).
28. Allman, C. et al. *Sci. Transl. Med.* 8, (2016).
29. Fusco, A. et al. *Restor. Neurol. Neurosci.* 32, 301–12 (2014).
30. Lindenberg, R., Renga, V., Zhu, L.L., Nair, D. & Schlaug, G. *Neurology* 75, 2176–84 (2010).
31. Nair, D.G., Renga, V., Lindenberg, R., Zhu, L. & Schlaug, G. *Restor Neurol Neurosci* 29, 411–420 (2011).
32. Stagg, C.J. et al. *J. Neurosci.* 29, 5202–6 (2009).
33. Bachtiar, V. et al. *Curr. Biol.* 4, 1023–1027 (2015).
34. Nitsche, M.A. et al. *J Physiol* 553, 293–301 (2003).
35. Stagg, C.J., Bachtiar, V. & Johansen-Berg, H. *Curr Biol* 21, 480–484 (2011).
36. Di Pino, G. et al. *Nat. Rev. Neurol.* 10, 597–608 (2014).
37. Di Lazzaro, V. et al. *Brain Stimul.* 7, 841–848 (2014).
38. Ward, N.S. & Cohen, L.G. *Arch. Neurol.* 61, 1844–8 (2004).
39. Murase, N., Duque, J., Mazzocchio, R. & Cohen, L.G. *Ann. Neurol.* 55, 400–9 (2004).
40. Monte-Silva, K. et al. *Brain Stimul.* 6, 424–32 (2013).
41. Bastani, A. & Jaberzadeh, S. *Clin. Neurophysiol.* (2014).doi:10.1016/j.clinph.2014.01.010
42. Huang, V.S. & Krakauer, J.W. *J. Neuroeng. Rehabil.* 6, 5 (2009).
43. Scott, N.W., McPherson, G.C., Ramsay, C.R. & Campbell, M.K. *Control. Clin. Trials* 23, 662–74 (2002).
44. Duncan, P.W. et al. *Stroke* 36, e100-43 (2005).
45. Quinn, T. et al. *J. Rehabil. Med.* 41, 99–111 (2009).
46. Waters, S., Wiestler, T. & Diedrichsen, J. *J. Neurosci.* 37, 7500–7512 (2017).
47. Gandiga, P.C., Hummel, F.C. & Cohen, L.G. *Clin Neurophysiol* 117, 845–850 (2006).
48. Gladstone, D.J., Danells, C.J. & Black, S.E. *Neurorehabil. Neural Repair* 16, 232–240 (2002).
49. HSIEH, C.-L., HSUEH, I.-P., CHIANG, F.-M. & LIN, P.-H. *Age Ageing* 27, 107–113 (1998).
50. Collin, C., Wade, D.T., Davies, S. & Horne, V. *Int. Disabil. Stud.* 10, 61–63 (1988).
51. van der Vliet, R. et al. *Subm.* (2020)

52. Page, S.J., Fulk, G.D. & Boyne, P. *Phys. Ther.* 92, 791–8 (2012).
53. van der Vliet, R. et al. *Ann. Neurol.* 87 (3), 383-393 (2020)
54. Hoonhorst, M.H. et al. *Arch. Phys. Med. Rehabil.* 96, 1845–1849 (2015).
55. van der Vliet, R., Ribbers, G.M., Vandermeeren, Y., Frens, M.A. & Selles, R.W. *Brain Stimul.* 10, 882–892 (2017).
56. Dehem, S. et al. *Int. J. Rehabil. Res.* 41, 138–145 (2018).
57. Hesse, S. et al. *Neurorehabil. Neural Repair* 25, 838–46 (2011).
58. Elsner, B., Kwakkel, G., Kugler, J. & Mehrholz, J. *J. Neuroeng. Rehabil.* 14, 95 (2017).
59. Elsner, B., Kugler, J., Pohl, M. & Mehrholz, J. *Cochrane Database Syst. Rev.* (John Wiley & Sons, Ltd: Chichester, UK, 2016).doi:10.1002/14651858.CD009645.pub3

Chapter 5. General discussion

Optimal control models of movement

The optimal control model of movement has been successful in providing a unified explanation of motor control and motor learning.[1] In this framework, the motor system sets a motor goal (possibly in the prefrontal cortex) and judges its value based on expected costs and rewards in the basal ganglia.[2] Selected movements are executed in a feedback control loop involving the motor cortex and the muscles which runs on an estimate of the system's states.[2] Both the feedback controller and the state estimator are optimal in a mathematical sense. The feedback controller because it calculates optimal feedback parameters for minimizing motor costs and maximizing performance, given prescribed weighting of these two criteria.[3] The state estimator because it optimally combines sensory predictions from a forward model (cerebellum) with sensory feedback from the periphery (parietal cortex), similar to a Kalman filter. [4,5] In the optimal control model of movement, motor adaptation is defined as calibrating the forward model, which is optimal in the same sense as the state estimator.[6]

In Chapter 2 of this thesis, we showed that is both possible and useful to model (components of) the optimal control model of movement on an individual level. Possible, because recent statistical and computational advances have provided us with Bayesian tools which can be used to fit complex models to movement data, as we did for a state-space model of movement adaptation. Useful, because we could uncover optimal relations between movement variability and adaptation rate and between movement variability and EEG activity. Therefore, our studies build upon earlier findings on optimal movement behavior in humans[1,2,7,8] and on the role of theta activity during adaptation,[9–11] and extend the explanatory power of the optimal control model of movement to between-subject variations in motor learning ability.

How could we further harvest the capabilities of our modeling approach for motor learning and recovery after stroke? First, we could use the very precise estimates of learning parameters for the evaluation of new interventions. This might help increase study power relative to more crude averages of learning processes, such as used in Chapter 4 of this thesis, and allow for more natural perturbations in the experimental design. Second, the model could aid in mapping brain lesions to specific motor learning deficits. Currently, it is unknown whether motor learning itself is affected in stroke patients relative to healthy subjects. Finally, understanding of individual differences in motor learning ability between healthy individuals could be deepened by relating the learning parameters to variations in genetic make-up (for example the BDNF Val66Met polymorphism studied in Chapter 4) or cortical anatomy.

Indeed, we are currently implementing the visuomotor adaptation experiment on a large scale in the Generation R cohort study. This requires minimization of the number of trials to limit the duration of the experiment as much as possible while maintaining reliable parameter estimation. We have already seen in a comparably complex study of eye movements during reading and social responsiveness in Generation R, that implementation is indeed possible (see the additional publications 12-13 at the end of this thesis for further reference). We expect this project to provide us with a rich and unique dataset that will help us further investigate genetic and anatomical determinants of motor learning ability.

Proportional recovery models of stroke

Longitudinal studies have repeatedly demonstrated the time-dependency of neurological recovery after stroke, including upper[12,13] and lower limb motor impairment,[14,15] visuo-spatial neglect,[16] and speech.[17] This suggests that recovery follows a predictable pattern, which is often described as spontaneous biological recovery.[18,19] Understanding the mechanisms and individual dynamics that drive stroke recovery is vital for developing better prognostic models and more effective, personalized therapeutic interventions.[20–23] The proportional recovery rule has been instrumental in modeling spontaneous upper extremity recovery by linking baseline motor impairment[24] to the observed motor recovery.[25] More specifically, the proportional recovery rule states that in three to six months (1) the majority of patients (recoverers) gain a fixed proportion, estimated between 0.55 and 0.85,[13] of their potential recovery, calculated as the difference between baseline FM-UE and the scale's maximum score of 66, while (2) the minority of patients (non-recoverers) show only very moderate improvement which cannot be linked to potential recovery.[12,13,25] Mechanistically, the key underlying difference between recoverers and non-recoverers is currently understood as the intactness of the corticospinal tract early after stroke.[26–29]

In Chapter 3 of thesis, we extended the concepts of proportional recovery to a longitudinal mixture model which captures recovery on the FM-UE in the subacute phase after stroke with an exponential recovery function and five subgroups, which we organized in clusters of poor, moderate and good FM-UE recovery. This model will be useful for (1) refining the search for causal and prognostic biomarkers of spontaneous motor recovery[8]; (2) informing patients early about their clinical prognosis for motor recovery; (3) estimating clinical intervention effects with greater sensitivity; and (4) selecting patients for clinical studies to increase study homogeneity[9,10]. Addressing point (3), we amended the longitudinal mixture model with an extra term that covers participation in an intervention trial and found a much higher sensitivity to detect intervention effects compared to a classical cross-sectional model in a series of simulations. In addition (point (4)), we showed that selecting patients from FM-UE recovery clusters increases the power to detect intervention effects in those clusters specifically. We tested the robustness of this new approach in multiple ways. First, we violated the basic assumption on the time course of the intervention effect and identified a negligible effect on study power. Second, we simulated data with 25% more residual error and found that the number of patients needed for 90% power was still four times smaller than the number needed for the cross-sectional test. Third, the real-world implementation of the model presented in Chapter 4.5 shows that the confidence intervals on the intervention effect agree with the expected values.

How could we use our model to gain a better understanding of recovery after stroke? Currently, it is still unknown whether any therapeutic intervention impacts recovery of impairment. Using our models, it is possible to find differences in both the amount and timing of recovery as well as in shifts between FM-UE recovery subgroups, either for the entire study population or for specific subgroups or clusters. Therefore, either reanalysis of already completed studies or analysis of upcoming studies with our model of FM-UE recovery could indicate which therapies are effective for specific patients. In addition, the model of FM-UE recovery could be used to identify determinants of individual differences in spontaneous recovery after stroke, which could help to (1) improve prognostic accuracy and (2) open new avenues towards drug development. These open questions rely on large cohort studies for their answer.

Neuromodulation

At the outset of this thesis, a promising new neuromodulation tool in motor rehabilitation was transcranial direct current stimulation (tDCS), a safe[30] and non-invasive technique that delivers low-intensity current to the scalp through a pair of electrodes.[31,32] Depending on the polarity of the electrodes and the spatial orientation of the underlying neurons,[33,34] direct current had been found to alter the excitability of the motor cortex, as measured with transcranial magnetic stimulation, for approximately an hour.[35–37] In addition, tDCS had been reported to improve motor skill learning in healthy subjects[38–45] and chronic stroke patients,[46,47] and upper limb rehabilitation in subacute and chronic stroke patients with moderately severe cortical damage.[48–52] tDCS was presumed to improve motor learning by releasing brain-derived neurotrophic factor[41] and down-regulating GABA[53–56] and support motor rehabilitation after stroke through restoration of the interhemispheric imbalance between the affected motor cortex and the unaffected motor cortex.[57–60]

In Chapter 4 of this thesis, we found little supportive evidence for a positive effect of single-session motor cortex tDCS or single-session cerebellar tDCS on motor learning, or repeated motor cortex tDCS on upper extremity rehabilitation after stroke. Only for the very basic task of eyeblink conditioning did we replicate[61] a supporting role for cerebellar tDCS in people with a BDNF Val66Met polymorphism. Although this general conclusion might have been surprising in light of the early tDCS experiments which inspired this thesis, recent studies have increasingly found negative results on both an electrophysiological, a behavioral and a clinical level. First, a meta-analysis of the electrophysiological changes following motor cortex tDCS has shown that the increase in excitability has sharply declined over the last decade,[62] with recent studies failing to establish an effect.[63] Therefore, the basic finding which fueled tDCS research is now under question. Second, behavioral gains found for single-session cerebellar tDCS on visuomotor adaptation[64] and forcefield adaptation[65] and for single-session motor cortex tDCS on motor skill learning, have not been replicated by later studies,[66–68] although the results for motor cortex stimulation and multiple day skill learning have been more consistent.[38–41,69] Third, the 2016 Cochrane review on tDCS in the subacute phase after stroke found very small effects in activities of daily living but not in physical or cognitive functioning based on studies of very low to moderate quality.[70] In addition, theoretical concerns regarding interhemispheric inhibition as target for neuromodulation in stroke patients have been recently brought up. The concept behind interhemispheric inhibition is that following a lesion, the affected motor cortex is suppressed by the unaffected motor cortex to a larger extent than in a healthy brain.[57–60] Suppressing the unaffected cortex, with for example cathodal tDCS, and activating the affected cortex, with for example anodal tDCS, might therefore be effective for alleviating symptoms in stroke patients. Xu et al. have investigated interhemispheric inhibition in a longitudinal study of mildly to moderately affected stroke patients.[71] Their primary findings were that interhemispheric inhibition only appears in the chronic phase after stroke and is actually more apparent in people with better recovery.[71] This means the specific rationale for application of motor cortex tDCS after stroke, is no longer supported by empirical evidence.

Given the conflicting evidence on motor cortex and cerebellar tDCS, how do we proceed? For starters, it seems important to replicate in well-powered, placebo-controlled studies the most consistent and essential electrophysiological and behavioral findings. That is, motor cortex tDCS effects on (1) motor evoked potentials and (2) multiple day motor skill learning,[38–41,69] and cerebellar tDCS effects on eyeblink conditioning. Replication results could help steer the design of future tDCS studies. If the behavioral results are replicated, this would suggest that tDCS can

in fact improve learning, but only in tasks spanning multiple days or in very simple paradigms such as eyeblink conditioning. A failure to replicate these behavioral results would leave very little evidence for positive effects of tDCS on motor learning and invite questions on its usefulness. Replication of the electrophysiological results in addition to the behavioral results would suggest that motor cortex excitability and motor learning are interconnected and validate attempts to optimize tDCS for motor cortex excitability. A failure to replicate the motor cortex excitability findings in light of positive behavioral results, would necessitate a search for alternative electrophysiological markers of treatment success. For example, it might make sense to quantify motor excitability after combined stimulation and motor training or use alternative techniques such as fMRI, MRI spectroscopy or EEG. To address replicability of tDCS-induced increases in motor evoked potentials, we are currently finishing a placebo-controlled motor cortex excitability study in a group of 60 healthy individuals, with results expected at the end of 2020.

References

1. Todorov, E. & Jordan, M.I. *Nat. Neurosci.* 5, 1226–35 (2002).
2. Shadmehr, R. & Krakauer, J.W. *Exp. brain Res.* 185, 359–81 (2008).
3. Åström, K.J. (Karl J. & Murray, R.M. (Princeton University Press: 2008).
4. Wolpert, D.M., Ghahramani, Z. & Jordan, M.I. *Science* 269, 1880–2 (1995).
5. Kalman, R.E. *J. Basic Eng.* 82, 35 (1960).
6. Shadmehr, R., Smith, M.A. & Krakauer, J.W. *Annu. Rev. Neurosci.* 33, 89–108 (2010).
7. Körding, K.P. & Wolpert, D.M. *Nature* 427, 244–247 (2004).
8. Todorov, E. *Nat. Neurosci.* 7, 907–15 (2004).
9. Anguera, J.A., Seidler, R.D. & Gehring, W.J. *J. Neurophysiol.* 102, 1868–1879 (2009).
10. Arrighi, P. et al. *PLoS One* 11, 1–27 (2016).
11. Torrecillos, F., Albouy, P., Brochier, T. & Malfait, N. *J. Neurosci.* 34, 4845–4856 (2014).
12. Prabhakaran, S. et al. *Neurorehabil. Neural Repair* 22, 64–71 (2008).
13. Winters, C., van Wegen, E.E.H., Daffertshofer, A. & Kwakkel, G. *Neurorehabil. Neural Repair* 29, 614–622 (2015).
14. Smith, M.-C., Byblow, W.D., Barber, P.A. & Stinear, C.M. *Stroke* 48, 1400–1403 (2017).
15. Veerbeek, J.M., Winters, C., van Wegen, E.E.H. & Kwakkel, G. *PLoS One* 13, e0189279 (2018).
16. Nijboer, T.C.W., Kollen, B.J. & Kwakkel, G. *Cortex* 49, 2021–2027 (2013).
17. Lazar, R.M., Speizer, A.E., Festa, J.R., Krakauer, J.W. & Marshall, R.S. *J. Neurol. Neurosurg. Psychiatry* 79, 530–534 (2008).
18. Kwakkel, G., Kollen, B. & Lindeman, E. *Restor. Neurol. Neurosci.* 22, 281–99 (2004).
19. Gresham, G.E. *Stroke* 17, 358–60
20. Ward, N.S. *Nat. Rev. Neurol.* 13, 244–255 (2017).
21. Boyd, L.A. et al. *Int. J. Stroke* 12, 480–493 (2017).
22. Bernhardt, J., Godecke, E., Johnson, L. & Langhorne, P. *Curr. Opin. Neurol.* 30, 48–54 (2017).
23. Byblow, W., Schlaug, G. & Wittenberg, G. *Ann. Neurol.* 80, 339–341 (2016).
24. Gladstone, D.J., Danells, C.J. & Black, S.E. *Neurorehabil. Neural Repair* 16, 232–240 (2002).
25. Krakauer, J. & Marshall, R. *Ann. Neurol.* 78, 845–847 (2015).
26. Feng, W. et al. *Ann. Neurol.* 78, 860–70 (2015).
27. Byblow, W.D., Stinear, C.M., Barber, P.A., Petoe, M.A. & Ackerley, S.J. *Ann. Neurol.* 78, 848–859 (2015).
28. Buch, E.R. et al. *Neurology* 86, 1924–1925 (2016).
29. Marshall, R.S. et al. *Ann. Neurol.* 65, 596–602 (2009).
30. Poreisz, C., Boros, K., Antal, A. & Paulus, W. *Brain Res. Bull.* 72, 208–14 (2007).
31. Nitsche, M.A. et al. *Brain Stimul.* 1, 206–23 (2008).
32. Nitsche, M.A. & Paulus, W. *Restor. Neurol. Neurosci.* 29, 463–92 (2011).
33. Rahman, A. et al. *J. Physiol.* 591, 2563–78 (2013).

34. Radman, T., Ramos, R.L., Brumberg, J.C. & Bikson, M. *Brain Stimul.* 2, 215–28, 228.e1–3 (2009).
35. Nitsche, M.A. & Paulus, W. *J Physiol* 527 Pt 3, 633–639 (2000).
36. Nitsche, M.A. & Paulus, W. *Neurology* 57, 1899–1901 (2001).
37. Wiethoff, S., Hamada, M. & Rothwell, J.C. *Brain Stimul.* 7, 468–475 (2014).
38. Waters-Metenier, S., Husain, M., Wiestler, T. & Diedrichsen, J. *J. Neurosci.* 34, 1037–50 (2014).
39. Prichard, G., Weiller, C., Fritsch, B. & Reis, J. *Brain Stimul.* 7, 532–40 (2014).
40. Reis, J. et al. *Proc Natl Acad Sci U S A* 106, 1590–1595 (2009).
41. Fritsch, B. et al. *Neuron* 66, 198–204 (2010).
42. Zimerman, M. et al. *Ann. Neurol.* 73, 10–5 (2013).
43. Vines, B.W., Cerruti, C. & Schlaug, G. *BMC Neurosci.* 9, 103 (2008).
44. Sriraman, A., Oishi, T. & Madhavan, S. *Brain Res.* 1581, 23–9 (2014).
45. Zimerman, M. et al. *Stroke* 43, 2185–2191 (2012).
46. Lefebvre, S. et al. *Front. Hum. Neurosci.* 6, 343 (2012).
47. Lefebvre, S. et al. *Brain* 138, 149–63 (2015).
48. Khedr, E.M. et al. *Neurorehabil. Neural Repair* 27, 592–601 (2013).
49. Allman, C. et al. *Sci. Transl. Med.* 8, (2016).
50. Fusco, A. et al. *Restor. Neurol. Neurosci.* 32, 301–12 (2014).
51. Lindenberg, R., Renga, V., Zhu, L.L., Nair, D. & Schlaug, G. *Neurology* 75, 2176–84 (2010).
52. Nair, D.G., Renga, V., Lindenberg, R., Zhu, L. & Schlaug, G. *Restor Neurol Neurosci* 29, 411–420 (2011).
53. Stagg, C.J. et al. *J. Neurosci.* 29, 5202–6 (2009).
54. Bachtiar, V. et al. *Curr. Biol.* 4, 1023–1027 (2015).
55. Nitsche, M.A. et al. *J Physiol* 553, 293–301 (2003).
56. Stagg, C.J., Bachtiar, V. & Johansen-Berg, H. *Curr Biol* 21, 480–484 (2011).
57. Di Pino, G. et al. *Nat. Rev. Neurol.* 10, 597–608 (2014).
58. Di Lazzaro, V. et al. *Brain Stimul.* 7, 841–848 (2014).
59. Ward, N.S. & Cohen, L.G. *Arch. Neurol.* 61, 1844–8 (2004).
60. Murase, N., Duque, J., Mazzocchio, R. & Cohen, L.G. *Ann. Neurol.* 55, 400–9 (2004).
61. Zuchowski, M.L., Timmann, D. & Gerwig, M. *Brain Stimul.* 7, 525–31 (2014).
62. Horvath, J.C., Forte, J.D. & Carter, O. *Neuropsychologia* 66, 213–236 (2015).
63. Horvath, J.C., Vogrin, S.J., Carter, O., Cook, M.J. & Forte, J.D. *Exp. Brain Res.* 234, 2629–2642 (2016).
64. Galea, J.M., Vazquez, A., Pasricha, N., de Xivry, J.J. & Celnik, P. *Cereb Cortex* 21, 1761–1770 (2011).
65. Herzfeld, D.J. et al. *Neuroimage* 98, 147–58 (2014).
66. Jalali, R., Miall, R.C. & Galea, J.M. *J. Neurophysiol.* jn.00896.2016 (2017).doi:10.1152/jn.00896.2016
67. Mamlins, A., Hulst, T., Donchin, O., Timmann, D. & Claassen, J. *J. Neurophysiol.* 121, 2112–2125 (2019).
68. Dehem, S. et al. *Int. J. Rehabil. Res.* 41, 138–145 (2018).
69. Waters, S., Wiestler, T. & Diedrichsen, J. *J. Neurosci.* 37, 7500–7512 (2017).
70. Elsner, B., Kugler, J., Pohl, M. & Mehrholz, J. *Cochrane Database Syst. Rev.* (John Wiley & Sons, Ltd: Chichester, UK, 2016).doi:10.1002/14651858.CD009645.pub3
71. Xu, J. et al. *Ann. Neurol.* 85, 502–513 (2019).

Chapter 6. Summary

In this thesis, we aimed to integrate recent insights on motor learning, stroke recovery and neuromodulation with the ultimate goal to improve upper limb rehabilitation after stroke.

Optimal control models of movement

In Chapter 2, we considered the optimal control model of movement to quantify individual differences in motor learning ability.

In Chapter 2.1, we investigated the relation between components of motor noise and visuomotor adaptation rate across individuals. If adaptation approximates optimal learning from movement error, it can be predicted from Kalman filter theory that planning noise correlates positively and execution noise negatively with adaptation rate.[1] To test this hypothesis, we performed a visuomotor adaptation experiment in 69 subjects and extracted planning noise, execution noise and adaptation rate using a state-space model of trial-to-trial behavior. Indeed, we found that adaptation rate correlates positively with planning noise and negatively with execution noise. In addition, the steady-state Kalman gain calculated from planning and execution noise correlated positively with adaptation rate. Therefore, individual differences in adaptation rate can be understood to a large extent from an individual's motor noise which means any effort to identify determinants of motor learning ability should include a decomposition of motor noise.

In Chapter 2.2, we found that, in the context of visuomotor adaptation, frontal midline theta activity (FMθ) does not act as a 'top-down teaching signal', but rather as 'bottom-up alarm signal'. The EEG analysis showed that the feedback-related FMθ in each trial was better explained by the absolute error size in the corresponding trial, than by the correction in the following trial, or a combination of both variables. The positive relation between frontal midline EEG activity and the absolute error size (EEG-error sensitivity) corroborates earlier work. This study expands on that earlier work in two ways. First, this study shows that EEG-error sensitivity is also present in the absence of external perturbations i.e. in response to small self-made errors during natural movements. Furthermore, this study shows that FMθ is directly involved in error detection, but not directly involved in error correction.

Proportional recovery models of stroke

In Chapter 3, we investigated statistical models of recovery in the subacute phase after stroke.

In Chapter 3.1, we developed a longitudinal mixture model of motor impairment recovery which describes the time course of the Fugl-Meyer assessment of the upper extremity (FM-UE) after a first-ever ischemic stroke and does not suffer from mathematical coupling.[6,7] Based on this model, we analyzed a large FM-UE dataset of 412 first-ever ischemic stroke patients collected in prospective cohorts. Subsequently, we identified five subgroups, which we organized in three clinically relevant clusters of poor, moderate and good recovery. Using cross-validation, our paper provides first-ever estimates of predictability of endpoint FM-UE between three and six months poststroke, as well as subgroup assignment as a function of time poststroke.

In Chapter 3.2, we amended the longitudinal mixture model of FM-UE recovery to account for participation in a stroke rehabilitation trial. Using this amended model, we simulated different randomized controlled studies and estimated study power. The longitudinal mixture model has a much higher power to detect intervention effects than a Mann-Whitney U-test applied to the endpoint FM-UE at 26 weeks poststroke. More specifically, based on a study design

with a limited number of repeated measurements (at one week and 26 weeks poststroke), without between-patient variability in timing of measurements or treatment start, we found a study sample of 70 patients to be sufficient for obtaining 90% power to detect a 4.25 point difference versus a study sample of 510 for the cross-sectional analysis.

Electrophysiology, genetics and neuromodulation

In Chapter 4, we investigated transcranial direct current stimulation as a neuromodulator for improving motor learning and rehabilitation after stroke.

In Chapter 4.1, we developed a new method for generating motor maps with transcranial magnetic stimulation. The golden standard for motor generation relied on a counting analysis of motor evoked potentials acquired with a predefined grid. However, with the development of digital reconstruction methods, it should now be possible to acquire motor maps with a much faster pseudorandom procedure. We compared the absolute reliability of the reconstruction methods with the golden standard by performing both grid and pseudorandom acquisition on two subsequent days in 21 healthy subjects. The standard error of measurement was at least equal using digital reconstructions. Pseudorandom acquisition and digital reconstruction can therefore be used in intervention studies without sacrificing reliability.

In Chapter 4.2, we undertook two cerebellar tDCS studies in subjects genotyped for BDNF Val66Met. Subjects performed an eyeblink conditioning task and received sham, anodal or cathodal tDCS or a vestibulo-ocular reflex adaptation task and received sham and anodal tDCS. For the eyeblink conditioning task, we found distinct groups of learners and non-learners. Carriers of the BDNF Val66Met polymorphism were more likely to be learners. Within the group of learners, anodal tDCS supported eyeblink conditioning in BDNF Val66Met non-carriers, but not in carriers. For the vestibulo-ocular reflex adaptation task, we found no effect of BDNF Val66Met or cerebellar tDCS. Therefore, the BDNF Val66Met polymorphism is important for some, but not all, cerebellar-dependent components of motor learning. Furthermore, cerebellar tDCS supports eyeblink conditioning only in non-carriers of the BDNF Val66Met polymorphism who have genetically determined slower conditioning rates.

In Chapter 4.3, we aimed to replicate the result that cathodal stimulation of the right cerebellum improves task performance on a verb generation task.[9] In contrast with the between-subject design study of the original study, we used a cross-over within-subject design, in order to reduce the impact of individual variability.[10] Participants had to complete two visits, with half of the group receiving cathodal c-tDCS the first time and half of the group receiving sham c-tDCS the first time. However, our results did not show a facilitating effect of cathodal c-tDCS on verb generation, either in terms of verbal response times or variability.

In Chapter 4.4, we investigated the role of BDNF Val66Met and motor cortex tDCS on motor skill learning of a circuit tracing task for which favorable effects of stimulation had been found in a similar chronic stroke patient group.[11,12] First, we were interested if the BDNF Val66Met polymorphism affects motor skill learning in patients with chronic stroke as observed in healthy subjects[13,14] and could serve as a mediator of motor cortex tDCS effects.[13] Indeed, non-carriers (no Met alleles) outperformed carriers (at least one Met allele) on day nine of the study. This result indicates activity-dependent release of BDNF is important for motor skill learning after stroke and could potentially mediate motor cortex tDCS effects. Second, we addressed if motor cortex tDCS affects motor skill learning and whether these effects depend on timing of stimulation relative to training. More specifically, we compared motor skill learning in patients receiving conventional stimulation during training (hypothesis “direct effects”), long-lasting

stimulation one day before training (hypothesis "aftereffects") or conventional stimulation one day before training (hypothesis "intermediate effects"), with sham stimulation. However, none of the tDCS protocols affected motor skill learning. In addition, we found no effect of any of the tDCS protocols on manual dexterity or maximum grip force.

In Chapter 4.5, we investigated the effects of long-lasting offline motor cortex tDCS on upper extremity motor recovery in 48 subacute ischemic stroke patients. No difference was found between the sham tDCS and long-lasting offline tDCS groups in motor impairment (FM-UE) or in any of the secondary outcomes on: (1) functional activity, (2) walking ability (10-meter walk test), (3) dependence in activities of daily living (Barthel index), (4) mood disorders.

References

1. Kalman, R.E. *J. Basic Eng.* 82, 35 (1960).
2. Anguera, J.A., Seidler, R.D. & Gehring, W.J. *J. Neurophysiol.* 102, 1868–1879 (2009).
3. Vocat, R., Pourtois, G. & Vuilleumier, P. *Neuropsychologia* 49, 360–367 (2011).
4. Torrecillos, F., Albouy, P., Brochier, T. & Malfait, N. *J. Neurosci.* 34, 4845–4856 (2014).
5. Arrighi, P. et al. *PLoS One* 11, 1–27 (2016).
6. Hope, T.M.H. et al. *Brain* 306514 (2018).doi:10.1093/brain/awy302
7. Hawe, R.L., Scott, S.H. & Dukelow, S.P. *Stroke* 50, 204–211 (2019).
8. Boyd, L.A. et al. *Int. J. Stroke* 12, 480–493 (2017).
9. Pope, P.A. & Miall, R.C. *Brain Stimul.* 5, 84–94 (2012).
10. Wiethoff, S., Hamada, M. & Rothwell, J.C. *Brain Stimul.* 7, 468–475 (2014).
11. Lefebvre, S. et al. *Front. Hum. Neurosci.* 6, 343 (2012).
12. Lefebvre, S. et al. *Brain* 138, 149–63 (2015).
13. Fritsch, B. et al. *Neuron* 66, 198–204 (2010).
14. McHughen, S.A. et al. *Cereb. Cortex* 20, 1254–62 (2010).

Samenvatting

In dit proefschrift hebben we recente inzichten over motorisch leren, revalidatie en neuromodulatie gecombineerd om het herstel van arm- handfunctie na een beroerte te verbeteren.

Modellen van bewegingssturing uit de meet- en regeltechniek

In hoofdstuk 2 hebben we optimale regeltheorie gebruikt om individuele verschillen in motorische foutcorrectiesnelheid te kwantificeren.

In hoofdstuk 2.1 onderzochten we de relatie tussen verschillende componenten van bewegingsruis in een reiktaak en de foutcorrectiesnelheid. Als foutcorrectiesnelheid de optimale leersnelheid benadert, voorspelt Kalmanfiltertheorie dat ruis in de planning van reikbewegingen positief correleert met foutcorrectie en ruis in de uitvoering van bewegingen negatief correleert met foutcorrectie.[1] Om deze hypothese te testen hebben we bij 69 proefpersonen reikbewegingen gemeten en met een specifiek leermodel (state-space model) de plannings- en uitvoerruis en de foutcorrectiesnelheid bepaald. Inderdaad vonden we een positieve correlatie tussen planningsruis en foutcorrectie en een negatieve correlatie tussen uitvoerruis en foutcorrectie. Verder berekenden we de optimale leersnelheid (Kalman gain) uit de twee ruistermen en stelden een positieve correlatie met de gemeten foutcorrectiesnelheid vast. De conclusie is dat verschillen in foutcorrectiesnelheid tussen individuen begrepen kunnen worden uit variaties in bewegingsruis. Voor studies naar individuele determinanten van leersnelheid is het daarom van belang een gedetailleerde analyse van de bewegingsruis uit te voeren.

In hoofdstuk 2.2 hebben we met behulp van elektro-encefalografie geconstateerd dat bepaalde hersenactiviteit (thetagolven frontaal in de middellijn) niet fungeert als een direct leersignaal, maar eerder als een toezichthouder voor het optreden van onverwachte fouten. Theta-activiteit gerelateerd aan foutterugkoppeling werd namelijk beter verklaard door de absolute afwijking van de huidige beweging dan door de correctie van de volgende beweging of een combinatie van de twee. De positieve correlatie tussen thetagolven frontaal in de middellijn en de absolute bewegingsfout (aangeduid als foutgevoeligheid van elektro-encefalografie) is in overeenstemming met resultaten uit eerder onderzoek. Wij voegen hier twee bevindingen aan toe. Ten eerste dat zelfs de kleinste, natuurlijk optredende bewegingsfouten terug te vinden zijn in thetagolven. Ten tweede dat thetagolven betrokken zijn bij herkenning maar niet bij correctie van fouten.

Modellen van proportioneel herstel na een beroerte

In hoofdstuk 3 hebben we statistische modellen van herstel in de subacute fase na een herseninfarct onderzocht.

In hoofdstuk 3.1 hebben we een longitudinaal groepsmodel van herstel van arm-handfunctie na een herseninfarct ontwikkeld. Dit model beschrijft het beloop van de Fugl-Meyer score (een maat voor beperking van arm- en handfunctie) na het optreden van een eerste herseninfarct en kampt niet met de wiskundige problemen van eerdere modellen.[6,7] Om de parameters van het model te schatten hebben we een bestand van 412 patiënten met een eerste herseninfarct samengesteld uit eerder uitgevoerde prospectieve cohortstudies. We vonden vijf subgroepen die we verder hebben gegroepeerd in voor ons klinisch relevante clusters van slecht, matig en goed herstel. Met behulp van kruisvalidatie hebben we voor het eerst de

voorspelbaarheid van toekomstig herstel op de Fugl-Meyerschaal, uitgedrukt als de absolute score of als het herstelcluster, uitgerekend.

In hoofdstuk 3.2 hebben we het longitudinale groepsmodel van herstel na een herseninfarct uitgebreid met een term die de effecten van deelname aan een klinische studie beschrijft. Met behulp van dit gewijzigde model hebben we gerandomiseerde klinische studies met verschillende onderzoeksprotocollen gesimuleerd. We vonden dat het longitudinale groepsmodel veel gevoeliger is in het vaststellen van interventie-effecten dan een gangbare cross-sectionele (Mann-Whitney U) test. Bijvoorbeeld, voor een onderzoeksopzet met een beperkt aantal metingen (op één en 26 weken na een beroerte) berekenden we een minimum van 70 patiënten voor het longitudinale model en 510 patiënten voor het cross-sectionele model om met 90% zekerheid een interventie-effect van 4.25 punten terug te kunnen vinden.

Elektrofysiologie, genetica en neuromodulatie

In hoofdstuk 4 hebben we onderzocht of een specifieke vorm van neuromodulatie, transcraniële gelijkstroomstimulatie, motorisch leren en revalidatie na een beroerte kan verbeteren.

In hoofdstuk 4.1 hebben we een nieuwe methode ontwikkeld voor het in kaart brengen van het hersenoppervlak dat betrokken is bij de aansturing van een enkele spier op basis van transcraniële magnetische stimulatie. De gouden standaard was gebaseerd op een rasteranalyse waarbij per vakje werd bepaald of de meerderheid van een vast aantal stimulatiepulsen wel of niet tot spieractiviteit leidde om zo tot een oppervlak te komen. Met de ontwikkeling van digitale reconstructiemethoden zou het nu echter mogelijk moeten zijn om het oppervlak te verkrijgen met een veel snellere pseudo-willekeurige procedure. We vergeleken de absolute betrouwbaarheid van de twee reconstructiemethoden door zowel raster- als pseudowillekeurig transcraniële magnetische stimulatie uit te voeren op twee opeenvolgende dagen in 21 gezonde proefpersonen. De standaardmeetfout van pseudowillekeurige stimulatie met digitale reconstructie was ten minste gelijk aan de gouden standaard. Pseudowillekeurige stimulatie in combinatie met digitale reconstructie kan daarom gebruikt worden in interventiestudies zonder aan betrouwbaarheid in te boeten.

In hoofdstuk 4.2 hebben we twee cerebellaire transcraniële gelijkstroomstimulatie-experimenten uitgevoerd bij personen die zijn geanalyseerd voor dragerschap van het BDNF Val66Met polymorfisme. Een deel van de proefpersonen nam deel aan een klassiek conditioneringsexperiment van de oogknipperreflex en ontving placebo, anodale of kathodale gelijkstroomstimulatie. Een ander deel onderging een adaptatie-experiment voor de vestibulo-oculaire reflex en ontving placebo of anodale gelijkstroomstimulatie. In het klassiek conditioneringsexperiment vonden we een duidelijke tweedeling in een groep responsieve en een groep niet-responsieve proefpersonen. Dragers van het BDNF Val66Met polymorfisme behoorden vaker tot de responsieve groep. Binnen de responsieve groep verhoogde anodale gelijkstroomstimulatie de snelheid van conditioneren in niet-dragers van het BDNF Val66Met polymorfisme maar niet in dragers van het polymorfisme. In het adaptatie-experiment van de vestibulo-oculaire reflex vonden we helemaal geen effect van dragerschap van het BDNF Val66Met polymorfisme of van transcraniële gelijkstroomstimulatie. De conclusie is daarom dat BDNF Val66Met belangrijk is voor sommige, maar niet voor alle vormen van cerebellumafhankelijk motorisch leren. Verder vonden we een bescheiden rol voor cerebellaire gelijkstroomstimulatie in conditionering van de oogknipperreflex in individuen die door hun genetische achtergrond trager leren.

In hoofdstuk 4.3 wilden we het prestatiebevorderende effect van kathodale stimulatie van het rechter cerebellum op werkwoordproductie reproduceren.[9] In tegenstelling tot de oorspronkelijke studie maakten we in ons experiment gebruik van een binnen-proefpersoon opzet om de impact van individuele verschillen te verminderen.[10] Deelnemers voerden de oefening twee keer uit en werden tijdens het eerste of het tweede bezoek gestimuleerd met kathodale gelijkstroom over het rechter cerebellum en tijdens het andere bezoek met placebo. In tegenstelling tot de oorspronkelijke studie toonden onze resultaten geen bevorderend effect van kathodale gelijkstroomstimulatie op het genereren van werkwoorden, noch in responstijden noch in variabiliteit.

In hoofdstuk 4.4 hebben we de rol van BDNF Val66Met en gelijkstroomstimulatie van de motor cortex op het aanleren van een traceerbeweging onderzocht. Eerder onderzoek had bij deze zelfde taak een gunstig effect van stimulatie gevonden in een vergelijkbare groep patiënten met een al langer geleden doorgemaakte beroerte.[11,12] Ten eerste wilden we uitzoeken of dragerschap van het BDNF Val66Met polymorfisme invloed zou kunnen hebben op de leersnelheid, net zoals eerder is vastgesteld in gezonde proefpersonen[13,14] en of BDNF betrokken zou kunnen zijn bij de positieve effecten van gelijkstroomstimulatie.[13] Inderdaad presteerden niet-dragers (geen Met-allel) beter dan dragers (minstens één Met-allel) op dag negen van de studie. Dit resultaat geeft aan dat activiteitafhankelijke afgifte van BDNF belangrijk is voor het leren van motorische vaardigheden na een beroerte en daarmee berokken zou kunnen zijn bij positieve effecten van gelijkstroomstimulatie. Ten tweede hebben we onderzocht of gelijkstroomstimulatie van de motor cortex het aanleren van een motorische vaardigheid beïnvloedt en of deze effecten afhangen van de timing van stimulatie in relatie tot training. Meer specifiek vergeleken we motorische leersnelheid bij patiënten die conventionele stimulatie tijdens training (hypothese "directe effecten"), langwerkende stimulatie een dag voor training (hypothese "na-effecten") of conventionele stimulatie een dag voor training (hypothese "indirecte effecten") kregen met een placebogroep. Geen van de gelijkstroomstimulatieprotocollen had echter invloed op het aanleren van de motorische vaardigheid. Bovendien vonden we geen effect van gelijkstroomstimulatie op handvaardigheid of maximale grijpkracht.

In hoofdstuk 4.5 onderzochten we de effecten van langwerkende tDCS op motorisch herstel van de bovenste extremiteit bij 48 patiënten met een recent doorgemaakt herseninfarct. Er werd geen verschil gevonden tussen placebo en langwerkende gelijkstroomstimulatie in arm-handvaardigheid (FM-UE) of in een van de secundaire uitkomstmaten: (1) functionele arm-handactiviteit, (2) loopvaardigheid (10-meter looptest), (3) afhankelijkheid in het dagelijks activiteiten (Barthel-index) of (4) stemmingsstoornissen.

Referenties

1. Kalman, R.E. *J. Basic Eng.* 82, 35 (1960).
2. Anguera, J.A., Seidler, R.D. & Gehring, W.J. *J. Neurophysiol.* 102, 1868–1879 (2009).
3. Vocat, R., Pourtois, G. & Vuilleumier, P. *Neuropsychologia* 49, 360–367 (2011).
4. Torrecillos, F., Albouy, P., Brochier, T. & Malfait, N. *J. Neurosci.* 34, 4845–4856 (2014).
5. Arrighi, P. et al. *PLoS One* 11, 1–27 (2016).
6. Hope, T.M.H. et al. *Brain* 306514 (2018).doi:10.1093/brain/awy302
7. Hawe, R.L., Scott, S.H. & Dukelow, S.P. *Stroke* 50, 204–211 (2019).
8. Boyd, L.A. et al. *Int. J. Stroke* 12, 480–493 (2017).
9. Pope, P.A. & Miall, R.C. *Brain Stimul.* 5, 84–94 (2012).
10. Wiethoff, S., Hamada, M. & Rothwell, J.C. *Brain Stimul.* 7, 468–475 (2014).

11. Lefebvre, S. et al. *Front. Hum. Neurosci.* 6, 343 (2012).
12. Lefebvre, S. et al. *Brain* 138, 149–63 (2015).
13. Fritsch, B. et al. *Neuron* 66, 198–204 (2010).
14. McHughen, S.A. et al. *Cereb. Cortex* 20, 1254–62 (2010).

Chapter 7. Epilogue

Publications

This thesis

1. Individual differences in motor noise and adaptation rate are optimally related. R. van der Vliet, M.A. Frens, L. de Vreede, Z.D. Jonker, G.M. Ribbers, R.W. Selles, J.N. van der Geest and O. Donchin. Eneuro 5 (4) (2018)
2. Frontal midline theta activity acts as a bottom-up alarm signal and not as a top-down teaching signal in the context of motor adaptation. Z.D. Jonker, R. van der Vliet, G. Maquelin, J. van der Cruijsen, G.M. Ribbers, R.W. Selles, O. Donchin and M.A. Frens. In preparation.
3. Predicting upper limb motor impairment recovery after stroke: a mixture model. R. van der Vliet, R.W. Selles, E.-R. Andrinopoulou, R. Nijland, G.M. Ribbers, M.A. Frens, C. Meskers and G.Kwakkel. Annals of Neurology 87 (3), 383-393 (2020)
4. Improving statistical power of subacute upper limb motor rehabilitation trials. R. van der Vliet, G. Kwakkel, E.-R. Andrinopoulou, R. Nijland, G.M. Ribbers, M.A. Frens, E.E.H. van Wegen, C. Meskers and R.W. Selles. Submitted.
5. TMS motor mapping: Comparing the absolute reliability of digital reconstruction methods to the golden standard. Z.D. Jonker, R. van der Vliet, C.M. Hauwert, C. Gaiser, J.H.M. Tulen, J.N. van der Geest, O. Donchin, G.M. Ribbers, M.A. Frens and R.W. Selles. Brain stimulation 12 (2), 309-313 (2019)
6. Cerebellar transcranial direct current stimulation interacts with BDNF Val66Met in motor learning. R. van der Vliet, Z.D Jonker, S.C. Louwen, M. Heuvelman, L. de Vreede, G.M. Ribbers, C.I. De Zeeuw, O. Donchin, R.W. Selles, J.N. van der Geest and M.A. Frens. Brain stimulation 11 (4), 759-771 (2018)
7. Cerebellar cathodal transcranial direct stimulation and performance on a verb generation task: a replication study. K. Spielmann, R. van Der Vliet, W.M.E. van de Sandt-Koenderman, M.A. Frens, G.M. Ribbers, R.W. Selles, S. Van Vugt, J.N. van der Geest and P. Holland Neural plasticity (2017)
8. BDNF Val66Met but not transcranial direct current stimulation affects motor learning after stroke. R. van der Vliet, G.M. Ribbers, Y. Vandermeeren, M.A. Frens and R.W. Selles. Brain stimulation 10 (5), 882-892 (2017)
9. Long-lasting tDCS in the subacute phase after stroke: double-blind randomized clinical trial. R. van der Vliet, Z.D. Jonker, M.A. Frens, R.W. Selles and G.M. Ribbers. In preparation.

Additional publications

10. Recovery of arm-hand capacity after stroke: a dynamic prediction approach. R.W. Selles, E.-R. Andrinopoulou, R.H.M. Nijland, R. van der Vliet, J. Slaman, C. Meskers, D. Rizopoulos, G. Ribbers and G. Kwakkel. In preparation.
11. The role of the BDNF Val66Met polymorphism in recovery of aphasia after stroke. R.G.A. de Boer, K. Spielmann, M.H. Heijenbrok-Kal, R. van der Vliet, G.M. Ribbers and W.M.E. van de Sandt-Koenderman. Neurorehabilitation and neural repair 31 (9), 851-857

12. Eye movement behaviour and silent reading: motor and cognitive correlates in a population-based study of pre-adolescents. S.C. Louwen, R. van der Vliet, H.H. Adams, S.P.C. Koenraads, M.C.J.P. Franken, M.H.J. Hillegers, J.N. Van der Geest and H. Tiemeier. In preparation.
13. Cortical Thickness and Surface Area correlates of Reading Profiency in pre-adolescents: A Population-Based Study. S.C. Louwen, R. van der Vliet, M.H.J. Hillegers, M.A. Frens, H. Tiemeier and R. Muetzel. In preparation.

PhD portfolio

Name PhD student	Rick van der Vliet
Erasmus MC Departments	Neuroscience Rehabilitation Medicine
PhD period	May 2013 April 2020
Promotors	Prof.Dr. M.A. Frens Prof.Dr. G.M. Ribbers
Co-promotor	Dr. R.W. Selles

1. PhD training	Year	Workload
General courses		
Master of Science in Medicine	2013-2019	180
Premaster Biomedical Engineering	2011-2014	67
Master of Science in Biomedical Engineering	2014-2017	120
BROK course	2014	2
Conferences		
Neuromodulation conference New York	2013	1
Biomedical Engineering conference	2015	1
Motor control conference Be'er Sheeva	2015	1
Neurorehabilitation and neural repair conference Maastricht	2015	1
Biomedical Engineering	2016	1
Biomedical Engineering	2017	1
Workshops, meeting and symposia		
Summer school Chicago	2016	2

2. Teaching activities	Year	Workload
Supervision		
Suzanne Louwen (master thesis)	2013-2014	3
Linda de Vreede (master thesis)	2014-2015	3
Suzy Margaretha (master thesis)	2014-2015	3
Zeb Jonker (master thesis)	2014-2016	3
Anne Geelhoed (bachelor thesis)	2015	2
Noor Gieles (bachelor thesis)	2015	2
Yiyi Zhang (bachelor thesis)	2015	2
Annelot van der Meulen (bachelor thesis)	2016	2
Christopher Hauwert (master thesis)	2016-2018	3
Marco Hoog (master thesis)	2016-2017	2
Charlotte Viëtor (master thesis)	2016-2018	3

Guido Maquelin (master thesis)	2016-2018	3
Junior Med School	2015	1
Junior Med School	2016	1
Lecturing		
Linear systems	2014-2016	6
Minor Medicine	2015-2017	2

Dankwoord

Iets minder dan zeven jaren scheiden het einde van deze thesis van het begin. Ongetwijfeld heeft dat alles te maken met de ingewikkelde keuze voor het juiste moment van afronden tegenover de veel sneller genomen beslissing voor het starten met onderzoek. Ieder hoofdstuk biedt ruimte voor nieuwe vragen en met de uitwerking van sommige van deze vragen zijn we ook nu nog bezig. Enfin, in ieder geval geeft het aan dat ik hier altijd met veel plezier gewerkt heb onder leiding van Gerard, Maarten en Ruud.

Maarten kende ik al voor mijn promotie. Van Junior Med School, nog op de middelbare school. En van de Honours Class, tijdens de bachelor geneeskunde. Altijd heb ik van Maarten de ruimte gekregen om creatieve vragen te bedenken en naar antwoorden te zoeken met alle in ons lab beschikbare onderzoeksopstellingen. Sommige opstellingen waren al een tijd buiten gebruik maar dan kreeg ik ze met Jos, Suzanne en Zeb weer werkend door te ontwerpen, bouwen en programmeren. Met Claire was het altijd gezellig, door haar volume leek het lab op thuis. Opher, die net was verhuisd naar Rotterdam toen ik aan mijn master neurowetenschappen begon, hield ons scherp met kritische vragen die in het begin frustreerden, maar onze onderzoeksvoorstellen altijd beter maakten. Alle complexe statistiek uit hoofdstuk drie heb ik van Opher geleerd: misschien wel de leukste tijd in het lab. Voor het gestructureerd opzetten van studies en opschrijven van resultaten kon ik altijd rekenen op Ruud. Het hardst hebben we samen met Gert gewerkt aan het duidelijk uitleggen van de modellen uit hoofdstuk drie aan een klinisch publiek. Tenslotte waren de twee projecten met patiënten uit hoofdstuk vijf nooit tot stand gekomen zonder Gerard, die als ervaren clinicus ook de fundamentele wetenschap van motor leren beheerst. Enorm inspirerend.

Het meest genoot ik van de studie geneeskunde tijdens de Honours Class. Met Hieab, Lennard, Sirwan en Zeb heb ik gereisd, geskied, gegeten en gedronken en eindeloos veel plannen gesmeed.

Het laatste deel van mijn promotie heb ik gecombineerd met klinisch werk op de afdeling neurologie van het Ikazia ziekenhuis, de afdeling van Laus en Dick. Met veel respect en plezier heb ik geleerd van hun oprechte betrokkenheid, humor en klinische ervaring.

Dit boek is voor mijn ouders en mijn broertje, die mij hebben geholpen drie studies en een promotieonderzoek af te ronden. Omdat ze erop vertrouwden dat ik dat kon. En omdat ze wisten dat ik met die kennis kan gaan doen wat ik interessant vind, op het snijvlak van techniek en geneeskunde, als arts en als ingenieur. En dit boek is voor Anne, die mij met haar kenmerkende niet-aflatende energie aanmoedigde om dit onderzoek toch eindelijk eens af te ronden. Om zo ook weer aan nieuwe dingen te kunnen beginnen. Nou ja. Oké. Goed dan.