

Rozšiřování kognitivních schopností: neuro-implantáty a HTM model neocortexu

Marek Otáhal, Jaroslav Vítků, Pavel Nahodil

ČVUT FEL, kat. Kybernetiky
Karlovo náměstí 13, Praha
Email: markotahal@gmail.com

Abstrakt

Přehled přístupů v oblasti *neuro-biologically inspired technologies*¹ a jejich rozdělení, ukázky využití a metody rozšiřování kognitivních schopností člověka.

Konkrétněji se pak zaměříme na *Hierarchical Temporal Memory-Cortical Learning Algorithm (HTM/CLA)* – paměťový model jehož princip vychází z biologické inspirace mozkovou kůrou savců - neopalliem, a který využíváme například pro rozpoznávání anomálií v sekvencích datech (EKG, video, ...) a výzkum senzomotorických signálů (robotika).

Třetí část je věnována metodám interakce člověka a stroje na úrovni centrální nervové soustavy (CNS) či přímo mozku (brain-computer interface - BCI) a rozvedeme diskusi o možných vylepšeních kognitivních schopností lidí využitím a propojením biologických a umělých neuronových sítí v jeden funkční celek a možné dopady na společnost a psychiku jedince.

Klíčová slova: neuromorphic HW, brain-computer interface, BCI, Hierarchical Temporal Memory, HTM/CLA, enhancement

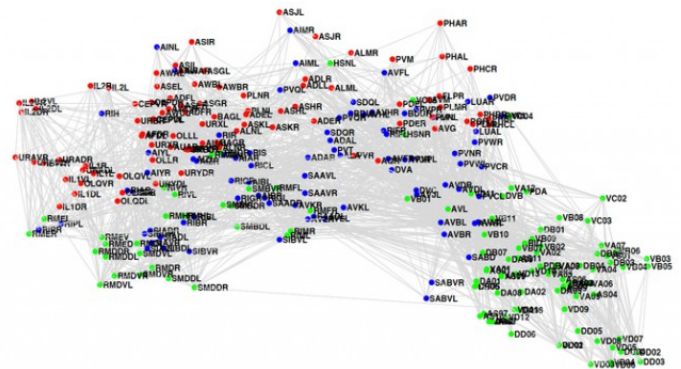
1 Úvod do současného stavu výzkumu a technologií v neurovědách

Po století páry a počítačů bývá to 21. mnohými označováno za století mozku. Umožnily to nové poznatky v neurovědách, rozvoj technologií pro sledování mozku, jako *fMRI*, *neuro-sondy apod.*, zvyšování výkonu procesorů zase přineslo renezanci neuronových sítí obecně a vznikají modely umělých neuronových sítí simulující svou biologickou předlohu do stále lepších detailů. To se projevuje i na "politické" úrovni - EU i USA podporují velké vědecké projekty zaměřené na mapování a simulaci funkce celého mozku, tzv. *connectomy*². Příkladem je např. connectome hádátka obecného na obr. 1. Tento pokrok prudce akceleruje zájem o obor, zapojují se mnohé komerční subjekty a vzniká specializovaný hardware -

¹neuromorphic hardware, biologické modely neuronových sítí, connectomy

²connectome = detailní mapa nervových spojení mozku, "diagram zadrátování", např.: The BlueBrain Project, The Connectome Project, The Human Brain Project... Gewaltig (2013), Frackowiak (1998)

*brain-chipy, retina implantáty, ...*Hsu (2014); Stingl a spol. (2015)



Obr. 1: Connectome hádátka obecného (*C. Elegans*), je v podstatě triviální v porovnání se snahou o model mozku myši či člověka, nicméně je kompletně hotový a umožňuje 3D simulaci nervové, svalové i trávicí soustavy hádátka ve virtuálním prostředí.

Zdroj: OpenWorm ProjectSzigeti a spol. (2014)

1.1 Rozdělení technologií v neurovědách

Přehledově bychom mohli pokrok rozdělit podle použitých technologií:

- hardware
 - senzory (oční, sluchové či hmatové implantáty)
 - měření aktivity mozku (sondy, pole, fMRI, EEG)
 - neuroprostetika (např. protéza ruky napojená na zbylá nervová zakončení)
- software
 - connectomy, mapy nervových HUBů, atlasy neuronů
 - biologicky inspirované umělé neuronové sítě (ANNs) - od deep learningu, přes HTM, spiking ANNs až po např. BlueBrain project, OpenWorm project 1

Případně dělit pokrok podle nasazení:

Lékařské poznatky zaměřené na člověka zahrnují vývoj léků (simulováno právě na reálných modelech CNS/mozku), neuro-stimulátory umožňují potlačit např. vliv Parkinsonovy choroby, nebo zmiňovaná neuro-prostetika (snaha o navrácení funkčnosti nejčastěji zraku (viz. obr. 7), sluchu či pohyblivosti končetin, ať už přímo, nebo přes jiný smyslový orgán). Kde nejsou dostupné implantáty, snažíme se pacientovi poskytnout alespoň lepší možnosti komunikace - přes rozhraní mezi člověkem a počítačem (*human-computer interface, HCI*), či přímo mezi mozkiem a strojem (*brain-machine interface, BMI*, Hsu (2015)). O poslední zmiňované má jistě zájem i armáda.

Na různé úrovni se projevuje pokrok i v běžně používaných technologiích: hlasové/zrakové ovládání, monitorování biologických signálů, asistované rozhodování, predikce, detekce anomálií, rozpoznávání, komunikace,... např. při řízení letadel (i modelů - dronů), chir. operacích nebo třeba modelování, tréninku lékařů i vojáků anebo obchodování na burze.

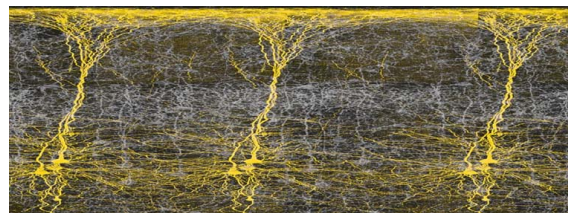
Za tím stojí i rozvoj obecné umělé inteligence (AI), která modeluje obrovská množství, často sekvenčních, streamovaných dat z různých zdrojů a dovoluje nám nad nimi uvažovat, či dovoluje simulovat mozek na dostatečně detailní úrovni (Gewaltig, 2013) a realizovat tak některé experimenty výrazně rychleji a z prostředí kanceláře.

Na jeden z těchto modelů se podíváme v další kapitole, v poslední části pak rozebereme teoretické možnosti vylepšování kognitivních schopností zdravých lidí v blízké budoucnosti.

2 Hierarchická temporální paměť (HTM)

Tato kapitola seznamuje s pojmem *hierarchické temporální paměti (hierarchical temporal memory, HTM)* (Hawkins a George, 2006; Hawkins, 2010; Mishkin a spol., 1997), což je biologicky inspirovaný model neuronových sítí, vycházející z inspirace mozkovou kůrou savců (neopallium, neocortex). Podle úrovně, o jak přesný biologický model jde, se HTM řadí mezi deep learning, spiking-NNs a connectomy, poměrně detailně modeluje důležité principy mozkové kůry (synapse, sloupce, inhibice, struktura, proces učení), ale zavádí některá zjednodušení (např. binární “váhy”, neurony stejného typu, zjednodušený synaptický potenciál) protože se snaží modelovat globální funkcionalitu - spatio-temporální paměť, detekce anomálií, online učení, reprezentace vzorů (ne tolik už například přenosy mezi synapsami).

Místo podrobného popisu HTM, které se dosti výrazně liší od “klasických” ANNs, bych zde rád představil hlavní myšlenky této teorie, používané koncepty a její výhody a uplatnění. Detaily lze prostudovat v Hawkins



Obr. 2: “Three Golden Columns” - Mozková kůra se skládá z tisíců sloupců tvořených jednotlivými neurony. Každý sloupec má průměr cca 0.5mm a obsahuje zhruba 10.000 neuronů. Kůru je také možné organizovat horizontálně do vrstev, žluté neurony na obrázku jsou velké pyramidové neurony v 5. vrstvě, na pozadí neurony dalších typů utvářející jeden sloupec kůry.

Zdroj: IBM/EPFL Blue Brain Project

(2010), případně kontaktovat komunitu NuPIC, která je velmi aktivní ve vývoji i výzkumu a jejímž jsem členem, algoritmus je implementovaný jako open-source a poměrně dobře nasaditelný v praxi.

2.1 Teorie neocortexu

Mozková kůra u savců má několik velmi zajímavých funkcí a vlastností, a právě z tohoto důvodu byla zvolena za biologickou inspiraci pro HTM teorii. Mezi důvody patří:

- **Modeluje “vyšší” kognitivní funkce**

V lidské mozkové kůře je realizováno zpracování řeči, vidění, slyšení, uchování epizodické paměti, detekce anomálií v sekvenčních datech, spojení informací z různých smyslových vjemů i senzomotorické chování. Tyto “úlohy” patří i v dnešní době k velmi obtížně algoritmicky zpracovávaným. (Pulvermüller a spol., 2014)

- **Jednotný princip a struktura**

Přes poměrně velké rozdíly ve výše popisovaných aktivitách zpracovávaných kortexem je velmi zajímavé, že neocortex má v podstatě homogenní strukturu, což vedlo k myšlence, že i všechny jeho oblasti pracují na stejném principu a mělo by tedy být možné je řešit stejným algoritmem - to vedlo ke vzniku teorie HTM.

Neocortex je tvořen neurony ve vertikálních sloupcích, a také horizontálními vrstvami a zpětnými spojeními mezi nimi (viz obr. 2). Zjednodušeně, sloupce vykonávají stejnou funkcionalitu a vrstvy pracují na stejné úrovni abstrakce. (Creutzfeldt, 1977)

- **Hierarchie, abstrakce a stabilní reprezentace**

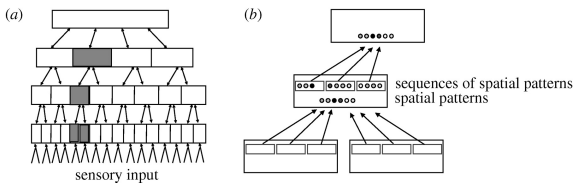
Hierarchické uspořádání, stejně jako u deep-NN, vytváří abstraktnější reprezentace. Neocortex má také hierarchickou strukturu, jak vidíme na obr. 2. Důležité jsou zpětné vazby jdoucí z vyšších vrstev

(stabilnější) do nižších, tím vznikají stabilní reprezentace na vyšších vrstvách pro celé sekvence. (Fuster, 2009)

- **Online učení, predikce a detekce anomálií**

HTM se učí na proudech dat (stream), online a bez učitele; zpracování sekvencí je důležité, protože mnohá smyslová data (krom zraku) nejdou rozpoznat z jednoho “doteku”, je potřeba si věc “osahat” sekvencí akcí a vytvořit si tak její model v paměti. To lze uvést na příkladu rozpoznávání písničky podle not, z jedné noty nepoznáme nic, ale už z pár not je znát celá písnička.

Stejně tak neustále probíhá detekce anomálií, které plní funkci pozornosti (*attention*), kde v pokoji si nevšímáme věcí postupně, ale ihned si všimneme chybějící vázy. (Hawkins, 2013)



Obr. 3: (a) Hierarchická struktura regionů v HTM. (b) Vznik stabilních reprezentací sekvencí na vyšších úrovních hierarchie.

Zdroj: Hawkins & George, Sequence memory for prediction, inference and behaviour

2.2 Koncepty v HTM

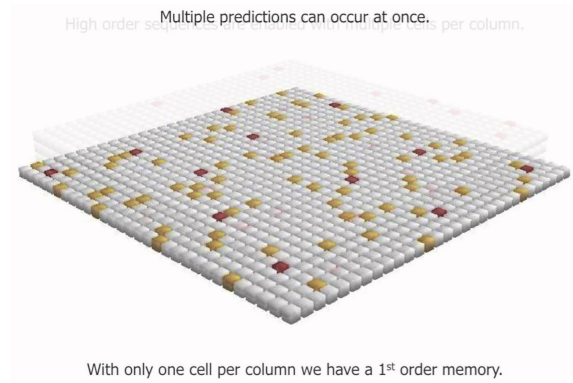
Tato sekce seznamuje s termíny zaběhlými v HTM, které usnadní případné čtení dalších článků.

- **Sparse, Distributed Reprezentace (SDR)**

V HTM jsou “myšlenky”, vzory reprezentovány výstupem sloupců, tedy binárními vektory. Těto reprezentaci se říká “sparse distributed” - SDR - Ahmad a Hawkins (2015). “Distributed” je něco mezi úplným kódováním a “grandmother cell”³, je v něm jeden neuron (sloupec) přepoužit ve více reprezentacích, a přitom jedna reprezentace je vždy skupinou několika aktivních neuronů. “Sparse” znamená, že je aktivní vždy stabilní (a malé, cca 2%) procento buněk. Příklad takové reprezentace na obr. 4. SDR mají další zajímavé vlastnosti: podobné koncepty mají podobné reprezentace a existuje “aritmetika”, významová suma konceptů se vyjádří sumou (log. OR) reprezentací - např. “ovoce” = “jablka” + “hrušky” + ...

- **Síť, region, sloupec, vrstva, neuron**

HTM tvoří síť regionů s dopřednými a zpětnozab-



Obr. 4: Ukázka konceptu v sparse, distributed reprezentaci - SDR. Koncept je tvořen sjednocením aktivních buněk z bottom-up vstupu (červeně) a predikovaných buněk z horizontálních synapsí (zeleně). Jeden koncept v SDR může “fuzzy” reprezentovat několik konceptů najednou.

Zdroj: Numenta, Modelling Data Streams workshop

nými vstupy, region je skupina sloupců, které vykonávají stejnou funkci (např. V1 region pro vidění), tyto sloupce jsou tvořeny neurony. Sloupce šíří informaci vertikálně, neurony horizontálně v rámci regionu. Regiony mohou tvořit horizontální vrstvy s podobnou úrovní abstrakce. Pak se vyskytují synapse v segmentech, to jsou nejmenší modelované jednotky.

- **Spatial Pooler, Temporal Pooler, Temporal Memory**

Spatial Pooler (SP) je funkční (algoritmický) koncept, který realizuje prostorovou paměť, vstupy do systému převádí na SDR. *Temporal Pooler (TP)* - sekvencí paměť, vytváří kontext. Pro každý vstup generuje predikce pro další krok a udává anomálnost aktuálního vstupu. *Temporal Memory (TM)* obdobně provádí temporální pooling - abstrahuje časté sekvence do nového konceptu SDR.

- **Učení - Cortical Learning Algorithm (CLA)**

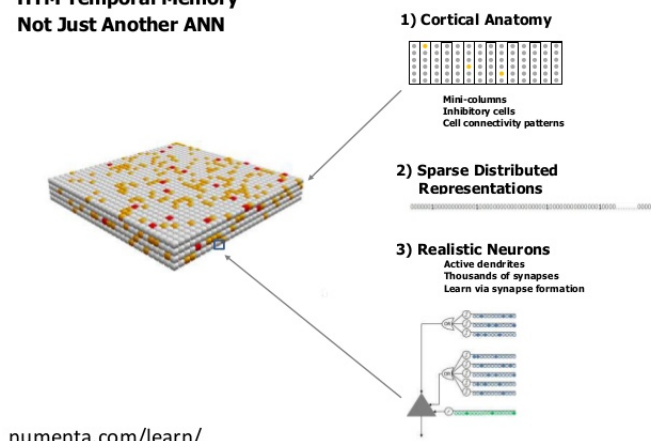
Učení probíhá online na principu Hebbovského učení, pokud neurony správně predikovaly v čase $T-1$ aktivitu, posílí se váha spojení s neurony aktivními v čase T . Schematické znázornění učení v HTM na obr. 3 a 5, blíže v Hawkins (2010); Hawkins a George (2006).

2.3 Vhodné úlohy

Z principu je HTM vhodná pro podobné úlohy, které řeší neocortex. Naopak nevýhody vyplývají z online učení bez učitele, nevhodné je učit se z malého množství labe-
lovaných dat, dobře dokáže detekovat anomálie ve streamovaných datech Hawkins (2013), v datech “s dírami”, sjednocovat data s různou modalitou nebo reprezentovat

³Úplné kódování - max hustota informace, bin. kód; Grandmother-cell - opak, jeden neuron reprezentuje jeden koncept.

HTM Temporal Memory Not Just Another ANN

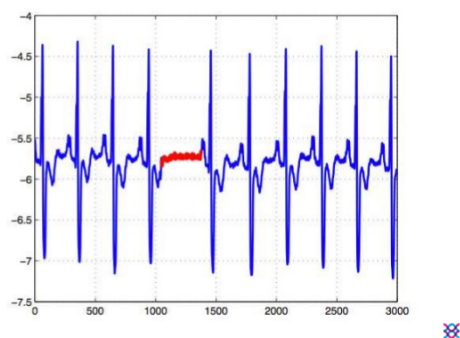


Obr. 5: Tvorba kontextu pomocí více buněk ve sloupci (zde 4), kde různé buňky mají různá synaptická spojení mezi sebou, čímž dokáží odlišit stejnou věc v různých kontextech. Výsledný vzor pak odpovídá sjednocení aktivity všech buněk ve sloupci.

Zdroj: Numenta, Modelling Data Streams workshop

celé sekvence (akce, např. pohyb ruky robota) novým konceptem. Tématicky je HTM vhodné pro predikci v sekvencích datech a detekci anomálií (na obr. 6 detekce anomálií v signálu EKG.), ve stádiu výzkumu je NLP, zpracování video a audio dat a motor-control. (Valentin, 2014)

Temporal Anomalies



Obr. 6: Detekce anomálií v EKG signálu pomocí HTM.

Zdroj: Numenta

3 Návrhy zlepšení kognitivních schopností člověka za použití neuronových sítí

Předchozí část seznámila s hierarchickou temporální pamětí (HTM) a popsala její výhody a nasazení, tato kapitola se zaměřuje na možnosti rozšiřování kognitivních schopností zdravého jedince využitím speciálních sen-

zorů, biologických neuronových sítí a brain-computer rozhraní (BCI).

Metod (s různou mírou důvěryhodnosti) pro zlepšení kognitivních schopností existuje mnoho druhů - pomocí podpůrných léků (nootropik), zdravý životní styl a cvičení (prokrvení, dostatek spánku), meditace (pozitivní nálada, uklidnění se prokazatelně “přepíná” mozek do jiných stavů, kde se lépe učí), sankalpa⁴, laterální myšlení (hledání kreativních řešení v době, kdy je mozek mimo soustředění, tedy unaven), udržování si neuroplasticity (různé kognitivní hry, předkládání stále nových problémů k řešení), učení se cizím jazykům (vytváří jiné reprezentace pro koncepty slov a tím “jiné” myšlení)

Na biologické úrovni tyto metody mají společné: výživu neuronů (strava, spánek - odplavování nečistot, cvičení - okysličení), posilování a růst nových spojení (mentální aktivita).

V následujících sekcích se podíváme na různé možnosti zlepšení mozkové aktivity - *neuro-enhancementu*.

3.1 Neuro-implantáty, human-machine rozhraní

Jedná se o připojování nových senzorů (častější) a aktuatorů k člověku. Nejčastěji jde o nějaký neuromorfický hardware, který se pomocí HMI/BCI připojuje k centrální nervové soustavě (to je jednodušší) nebo přímo do mozku. Díky neuroplasticitě a díky velké redundanci (jak ukazuje HTM model) je možné se přizpůsobit novým vjemům i z jiné modality (druhu), ovšem “přeučení se” na pouze pozměněný typ je snažší.

Příkladem je např. *infra-červené vidění* - napojení kamery s jiným spektrem snímání než má lidské oko na optický nerv (obr. 7), *3. oko* - přídavný vizuální senzor napojený na jiný sensorický vstup (jazyk), “*Magnetický-smysl*” v prstu napojený na hmatové vnímání, *tepelné čidlo*, nebo exotičtější - třeba “vnímání hodnoty akcií” v čase přes haptický nerv.

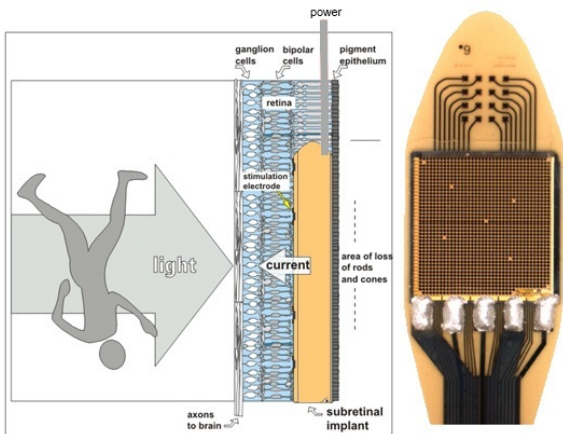
Možností je opravdu mnoho, v podstatě lze zvolit jakýkoli fyzikální senzor, či libovolný informační zdroj, vhodně ho převést do “neuronové reprezentace”, připojit na (existující) smyslový vstup a naučit se ho používat - rozlišovat.

3.1.1 Asistované učení

Po připojení nového smyslového senzoru se na něj musí nervová soustava adaptovat, to otevírá novou zajímavou větev výzkumu - *asistované neuro-učení*. Jde o to se co nejdříve naučit správně interpretovat signály jdoucí z/do senzoru. To je případ učení s učitelem (*supervised-learning*), kde využijeme spatio-temporální paměti a Hebbovského učení⁵ Můžeme stimulovat výstup ze senzoru, vstup na připojených nervech, nebo přímo i mozkové

⁴Sankalpa - technika práce s podvědomím před spánkem pomocí pozitivní motivace a navození myšlenek

⁵Hebb. pravidlo - “Who fire together, wire together”, blízké události mají kauzální linku.



Obr. 7: Neuroimplantát umělé sítnice připojený na optický nerv, nový úspěch - 30x40pixelů! Umožňuje znovu vidět pacientům se ztrátou zraku.
Zdroj: Retina Implant AG, 2010

regiony.

Dalším využitím by bylo zrychlování reflexů⁶ pro sportovce či armádu (trénování sešlápnutí brzdy v (před-připravené) krizové situaci, ještě dříve, než ji mozek vyhodnotil jako nebezpečnou a signál došel zpět do nervů (ruky), ...)

3.2 Brain-brain interface, Brain-computer interface, rozšiřování paměti

Následující postupy by vyžadovaly výrazně složitější hardware pro I/O rozhraní - muselo by být přímo v mozku a napojovat skupinu neuronů, tomu se říká *electrode array* a pokusy s nimi byly provedeny zatím jen na zvířatech. (Weyand a spol., 2015; Grandchamp a Delorme, 2009)

Brain-brain interface: pokud by se zdálo, že existuje univerzální formát reprezentace informace v mozku (např. SDR), šlo by se naučit mapování mezi SDR1 pro člověka1 a SDR2 pro člověka2, pak by Č1 a Č2 mohli komunikovat posíláním zpráv přímo mezi svými mozky - *telepaticky*.

Podobným způsobem přes mapování reprezentací v mozku, by bylo možné realizovat *rozšíření paměti*, ať už sdílením zkušeností někoho jiného (viz. případ výše), i odhadnutím vlastního mapování (např. pro doménu rozpoznávání hub), a příslušného dotrénování v umělém paměťovém modulu (HTM). Při vyhodnocování by se pak vstupní signál posílal do obou center současně (biologického i umělého) a to, které by vracelo výstup s větší jistotou by bylo použito jako dominantní.

⁶Podle jedné studie jsou naše reflexy výrazně pomalejší než u lidí žijících v 18. století. Pravděpodobně je to dáno způsobem obživy. I v dnešní populaci jsou trénování jedinci s výrazně menšími reakčními časy

3.2.1 Brain-AI interface, uvažování asistované AI

Navázáním na témata výše je rozšíření funkcionality mozku o nové/jiné výpočetní schopnosti - upravením HTM dokážeme dělat různé "transformace" s procházejícími myšlenkami. Jednalo by se o nový NN modul zapojený do cest (neural pathway) v mozku, vstupem je SDR, uvnitř transformační model HTM, výstupem zase SDR.

Například bychom mohli detekovat anomálie v signálech, pomáhat při plánování (predikce N-kroků vpřed) nebo myslet abstraktněji - HTM namapované jen na podčást prostoru se SDR provádí kompresi informace (ale co nejvíce "bezeztrátovou") a tím efektivně generalizuje; či kreativně - provede kompresi směrem nahoru v hierarchii, ale pak i zpětnou dekompresi s náhodným šumem - dochází k nepřesnému re-kódování vstupu a tím vzniku alternativní/kreativní reprezentace konceptu.

4 Diskuse

Myšlenky zmiňované v poslední kapitole představují výzkumné výzvy, co se neurobiologických modelů mozku týče, především je potřeba prozkoumat možnosti propojení mezi umělými a biologickými sítěmi. Na poli hardware je hlavně otázka (pro *neuronová pole*), jak co nepřesněji a nejšetrněji stimulovat a číst nervová zakončení či přímo neurony/synapse v mozku. Pro experimenty by zatím bylo vhodné použít modely *connectomů*, případně opatrně testovat na zvířatech v laboratoři. Zajímavou otázkou je, jak by se mozek vyrovnal s tak výraznou změnou - přidáním nového modulu/smyslu, za povšimnutí ovšem stojí nedávno provedená transplantace hlavy na úplně jiné tělo! Krom dopadů na psychiku jsou zde i zajímavé problémy etické a sociologické.

Poděkování

Tato práce byla podpořena z výzkumného grantu No. SGS14/144/OHK3/2T/13.

Literatura

- Ahmad, S. a Hawkins, J. (2015). Properties of Sparse Distributed Representations and their Application to Hierarchical Temporal Memory.
- Creutzfeldt, O. D. (1977). Generality of the Functional Structure of the Neocortex. *Naturwissenschaften*, 64:507–517.
- Frackowiak, R. (1998). Recreating the Brain Online. *Science*, 279(5355):p. 1320.
- Fuster, J. M. (2009). Cortex and memory: emergence of a new paradigm. *J Cogn Neurosci*, 21(11):2047–72.

- Gewaltig, M.-O. (2013). The Human Brain Project: Chances and Challenges for Computer Science. Horbach, M. (zost.), V *GI-Jahrestagung*, vol. 220 z *LNI*, str. 16. GI.
- Grandchamp, R. a Delorme, A. (2009). NeuroTRIP: A Framework for Bridging between Open Source Software. Application to Training a Brain Machine Interface. V *Signal-Image Technology Internet-Based Systems (SITIS), 2009 Fifth International Conference on*, str. 451–457.
- Hawkins, D. D. J. (2010). Hierarchical Temporal Memory including HTM Cortical Learning Algorithms. Technická správa.
- Hawkins, F. a George, D. (2006). Hierarchical Temporal Memory: Concepts, Theory, and Terminology. Technická správa, Numenta Inc.
- Hawkins, J. (2013). Online Learning from Streaming Data. V *Proceedings of the 22Nd ACM International Conference on Conference on Information & Knowledge Management, CIKM '13*, str. 1915–1916, New York, NY, USA. ACM.
- Hsu, J. (2014). IBM's new brain [News]. *Spectrum, IEEE*, 51(10):17–19.
- Hsu, W.-Y. (2015). Brain-computer interface: The next frontier of telemedicine in human-computer interaction. *Telematics and Informatics*, 32(1):180–192.
- Mishkin, M., Suzuki, W. A., Gadian, D. G. a Vargha-Khadem, F. (1997). Hierarchical Organization of Cognitive Memory. *Philosophical Transactions: Biological Sciences*, 352(1360):1461–1467.
- Pulvermüller, F., Garagnani, M. a Wennekers, T. (2014). Thinking in Circuits: Toward Neurobiological Explanation in Cognitive Neuroscience. *Biol. Cybern.*, 108(5):573–593.
- Stingl, K., Bartz-Schmidt, K.-U., Besch, D., Chee, C. K., Cottrill, C., Gekeler, F., Groppe, M., Jackson, T. L., MacLaren, R. E., Koitschev, A., Kusnyerik, A., Neffendorf, J., Nemeth, J., Naeem, M. A. N., Peters, T., Sachs, H., Simpson, A., Singh, M. S., Wilhelm, B., Wong, D. a Zrenner, E. (2015). Subretinal visual Implant Alpha IMS - Clinical trial interim report. *Vision Res.*
- Szigeti, B., Gleeson, P., Vella, M., Khayrulin, S., Palyanov, A., Hokanson, J., Currie, M., Cantarelli, M., Idili, G. a Larson, S. (2014). OpenWorm: an open-science approach to modeling *Caenorhabditis elegans*. *Front Comput Neurosci*, 8:137.
- Valentin, K. (2014). Invariant Visual Object Recognition Using Hierarchical Temporal Memory. *thesis*.
- Weyand, S., Takehara-Nishiuchi, K. a Chau, T. (2015). Weaning off Mental Tasks to Achieve Voluntary Self-Regulatory Control of a Near-Infrared Spectroscopy Brain-Computer Interface. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, PP(99):1–1.