Kognícia a umelý život 2023

zostavili

Igor Farkaš

Eva Ballová Mikušková

Martin Takáč

Kristína Malinovská

Matej Fandl

2023

Univerzita Komenského v Bratislave

Financovaný Európskou úniou

Smolenický zámok

Smolenice

1.6. – 3.6.2023

Kognícia a umelý život 2023
(recenzovaný zborník)

Tento zborník bol podporený Kultúrnou a edukačnou grantovou agentúrou (KEGA) MŠVVaŠ SR, projekt č. 022UK-4/2023, a Agentúrou na podporu výskumu a vývoja, projekt č. APVV-20-0335.

Recenzenti príspevkov:

PhDr. Eva Ballová Mikušková, PhD.
doc. Ing. Mgr. Jozef Bavoľár, PhD.
Ing. Ondřej Bečev, PhD.
prof. RNDr. Ľubica Beňušková, PhD.
RNDr. Zuzana Berger Haladová, PhD.
RNDr. Barbora Cimrová, PhD.
Mgr. Matej Fandl
prof. Ing. Igor Farkaš, Dr.
doc. Ing. Peter Lacko, PhD.
Mgr. Barbara Lášticová, PhD.
RNDr. Andrej Lúčny, PhD.
RNDr. Kristína Malinovská, PhD.
Mgr. Matej Pecháč
Mgr. Xenia Daniela Poslon, PhD.
doc. Ing. Branislav Sobota PhD.
prof. PhDr. Marián Špajdel, PhD.
doc. RNDr. Martin Takáč, PhD.
Mgr. Branislav Uhrecký, PhD.
Mgr. Michal Vavrečka, PhD.

Vydavateľ: Univerzita Komenského v Bratislave
Šafárikovo nám. 6
814 99 Bratislava

Tlač: Univerzita Komenského v Bratislave

Bratislava jún 2023

ISBN 978-80-223-5610-7 (online)

Predslov

Milé kolegyne a kolegovia,

vstupujeme už do tretej dekády organizovania česko-slovenskej konferencie Kognícia a umelý život (KUŽ), vďaka našej životaschopnej komunite, ktorá prispieva k rozvoju interdisciplinárnej kognitívnej vedy. Minulý rok sa KUŽ po dvojročnej prestávke spôsobenej pandémiou konal v Třešti v ČR, preto sme tradične prevzali štafetový kolík a 21. ročník konferencie organizujeme na Slovensku, tentokrát v krásnom historickom prostredí Smolenického zámku v Smoleniciach.

Ďakujeme za vaše zaujímavé príspevky, vďaka ktorým sme mohli zostaviť pestrý program. Dúfame, že v publikovaných dlhých i krátkych príspevkoch v elektronickom zborníku nájdete inšpiračné zdroje pre ďalšie skúmanie v oblastiach vášho záujmu.

Tradičným obohatením programu sú pozvané odborné prednášky, ktoré budú tento krát štyri. Prof. Ľubica Beňušková (Univerzita Komenského v Bratislave) nás uvedie do sveta výpočtovej neurovedy a neurogenetického modelovania mozgu. Dr. Martin Majerník (MindMed, Ltd.) sa bude vo svojej prednáške zaoberať skúmaním liekov inšpirovaných psychedelikami a digitálnou diagnostikou duševných porúch. Dr. Čeněk Šašinka (Masarykova univerzita v Brne) nás zasvätí do tajov imerzívnej virtuálnej reality ako prostriedku na extenziu ľudskej kognície, a privíta nás aj na súvisiacom workshope. Napokon, Dr. Alistair Knott (Victoria University of Wellington, Nový Zéland) vnesie do konferencie širší internacionálny rozmer a oboznámi nás s oblasťou búrlivo sa rozvíjajúcej umelej inteligencie a spektre debát týkajúcich sa potreby jej celosvetovej regulácie.

Na KUŽ 2023 budú prezentované tradične pestré témy príspevkov, ktoré prešli recenzným konaním, a to vďaka ochote a poctivej práci 19 oslovených recenzentov, za čo im srdečne ďakujeme. Opäť sme pripravili aj posterovú sekciu, ktorá umožňuje dlhšie interakcie autorov so záujemcami.

Okrem odborného programu máme v ponuke pripravenú aj spoločenskú aktivitu, exkurziu vo výrobni medoviny vo firme Včelco v Smoleniciach, spojenú s ochutnávkou rôznych druhov tohto sladkého moku. Veríme, že toto bude zaujímavým spestrením konferencie, na ktoré potom nadviaže večerný raut v priestoroch zámku.

Za všetkými týmito organizačnými krokmi je aktivita konkrétnych ľudí. Veľké poďakovanie patrí členom Centra pre kognitívnu vedu na Fakulte matematiky, fyziky a informatiky Univerzity Komenského v Bratislave. Osobitne oceňujeme nezištnú prácu Martina Takáča, Kristíny Malinovskej, Mateja Fandla a našej novej kolegyne Niny Zmajkovič Kučerákovej. Realizácia konferencie KUŽ 2023 bola podporená Kultúrnou a edukačnou agentúrou Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky (KEGA) v rámci projektu č. 022UK-4/2023, ako aj Slovenskou spoločnosťou pre kognitívnu vedu. Ďakujeme aj za sponzorskú podporu firme Gratex International, spol. s r.o.

Užite si príjemné momenty na KUŽ 2023 v tradičnej priateľskej atmosfére a kontexte podnetných rozhovorov týkajúcich sa nielen kognitívnej vedy.

Igor Farkaš
Centrum pre kognitívnu vedu, FMFI, Univerzita Komenského v Bratislave

Eva Ballová Mikušková
Centrum spoločenských a psychologických vied, v.v.i., Slovenská akadémia vied, Bratislava

Obsah

Pozvaní rečníci

Ľubica Beňušková

Prof. Beňušková vyštudovala biofyziku na FMFI UK v Bratislave. V tomto odbore ziskala aj titul PhD. Absolvovala aj magisterské štúdium výpočtovej neurovedy na Vanderbilt Univerzite v Nashville, TN, USA. V rámci tohto štúdia spolupracovala aj s nositeľom Nobelovej ceny prof. Leonom Cooperom. V rokoch 2003–2016 pôsobila na Novom Zélande, najskôr na Auckland University of Technology a potom na University of Otago, kde získala titul Associate Professor. Momentálne pracuje v Centre pre kognitívnu vedu FMFI UK v Bratislave. Výskumne sa venuje výpočtovej neurovede (najmä modelovaniu synaptickej plasticity) a analýze fMRI dát pomocou teórie grafov.

Prednáška: Výpočtová kognitívna neuroveda

Na úvod stručne predstavím publiku, kto to bol Peter Fedor, ktorému je venovaný tento ročník KUŽ. Neuróny v mozgu si vymieňajú informácie pomocou spajkov a neurotransmiterov. Ale neurón ako každá bunka má jadro a v ňom DNA. V prednáške pojednáme o tom ako táto DNA súvisí s činnosťou neurónov a ako môžme tento vplyv výpočtovo modelovať. Tento prístup sa zameriava na myšlienku tzv. génových regulačných sietí (GRN), čo sú siete interakcií medzi génmi a ich prostredím, ktoré ovplyvňujú činnosť každej bunky. Predstavíme ako môžu gény ovplyvňovať prenos signálov v neurónových sieťach a v konečnom dôsledku aj kogníciu. Prednáška je založená na monografii Computational neurogenetic modeling https://link.springer.com/book/10.1007/978-0-387-48355-9 autorov Beňušková a Kasabov.

Alistair Knott

Alistair Knott pracuje v oblasti kognitívnej vedy a umelej inteligencie (AI). Vyštudoval filozofiu a psychológiu na Oxfordskej univerzite, potom absolvoval postgraduálnu a postdoktorálnu prácu v oblasti umelej inteligencie na Edinburskej univerzite. Od roku 1998 pôsobí na Novom Zélande, najprv na Otago University a momentálne na Victoria University of Wellington. Ali sa vždy zaujímal o etiku a sociálny dopad umelej inteligencie. Na univerzite v Otagu spoluzakladal Centrum pre umelú inteligenciu a verejnú politiku, v rámci ktorého viedol výskum o vládnom využívaní UI na Novom Zélande a o vplyve UI na pracovné miesta a prácu na Novom Zélande. Vo Wellingtone koordinuje pravidelný seminár o UI a spoločnosti, na ktorom sa stretávajú odborníci z univerzitného prostredia so zainteresovanými stranami z vlády. Okrem Nového Zélandu je Ali členom pracovnej skupiny Global Partnership on AI's Responsible AI Working Group, kde spolu s Dinom Pedreschim vedie projekt o správe sociálnych médií, zameraný na odporúčacie systémy, klasifikátory škodlivého obsahu a veľké jazykové modely. Ali sa podieľal aj na pracovnom prúde Algoritmy v rámci výzvy Christchurch a zúčastňoval sa na činnosti pracovných skupín v rámci Globálneho internetového fóra pre boj proti terorizmu. Ali pracuje aj v odvetví umelej inteligencie. Pripojil sa k spoločnosti Soul Machines so sídlom na NZ, ktorá sa zaoberá umelou inteligenciou, ako akademický spolupracovník, keď bola v roku 2016 založená. V spoločnosti Soul Machines Ali pracuje so svojím významným kolegom Martinom Takáčom na rozsiahlom projekte na vytvorenie simulovaného stelesneného mozgu, ktorý sa používa v hlavnom produkte spoločnosti (systém dialógu medzi človekom a počítačom) a tiež v ich výskumnej platforme (simulované dieťa BabyX). Ali má tiež hlavnú zodpovednosť za etickú politiku spoločnosti.

Prednáška: How can we avoid confusion in the global conversation about AI regulation?

Discussions about AI and its oversight are happening everywhere - between friends in cafes, between colleagues in schools and workplaces, within local and national governments, in international conferences, in global citizens' groups, in big and small tech companies. AI/tech companies participate in multiple transparency initiatives in different jurisdictions, engage with external stakeholders of many kinds, and compete aggressively with one another. The large companies also run huge lobbying operations with multiple governments, and large international PR operations. On top of all this, the technologies at the centre of these discussions are progressing at a startling pace. It's important that the conversation about AI is broad, and stretches from high policymaking to grassroots. But how can we ensure that this broad national and international conversation is efficiently conducted, and leads to decisions in service of the common good? Lots of noise and activity is certainly not always a sign of progress. I will describe the state of the conversation as I have sampled it over the last few years, living in New Zealand, and participating in various international groups. I'll make a couple of proposals about how discussions can be made efficient, both within individual countries and multilaterally, between countries.

Martin Majerník

Martin Majerník odišiel do zahraničia prvýkrát už počas strednej školy. Pokračoval na Univerzite Komenského štúdiom aplikovanej psychológie, prácou v laboratóriu SAV UEP, na Univerzite v Cambridge a Univerzite v Herdfordshire štúdiom výpočtovej neurovedy a na Harvardskej škole biznisu štúdiom disrupčných stratégií. Po škole Martin pracoval s americkými investičnými fondmi, s ktorými sériovo zakladal inovatívne firmy v Európe. Po predaji jednej z nich je v súčasnosti riaditeľom produktového výskumu a vývoja, na burze obchodovanej americko-kanadsko-švajčiarskej firme MindMed, ktorá sa venuje vývoju novej generácie diagnostických nástrojov, terapií a liečiv duševných porúch. Je nositeľom niekoľkých patentov v oblasti digitálnej medicíny a členom Americkej spoločnosti digitálnej medicíny (DiMe).

Prednáška: Skúmanie liekov inšpirovaných psychedelikami a digitálna diagnostika duševných porúch

Táto prednáška preniká do vzrušujúceho sveta liekov inšpirovaných psychadelikami a digitálnej diagnostiky duševných porúch. Pokúsim sa poskytnúť vám pohľad do nášho súkromného výskumu a vývoja, pričom zdôrazním potenciálne výhody a výzvy, s ktorými sa stretávame. Od terapeutického potenciálu psychedelických látok, po použitie pokročilých digitálnych nástrojov pri diagnostikovaní a monitorovaní duševných stavov, preskúmame naše najnovšie poznatky a ich dôsledky pre starostlivosť o pacientov a všeobecné porozumenie zmenených stavov vedomia. Pripojte sa k nám na zaujímavú diskusiu, o priesečníku psychedelík a digitálnych technológií ktoré dúfame zmenia liečbu duševného zdravia.

Čeněk Šašinka

Primární výzkumnou oblastí Čeňka Šašinky je vizuální kognice, jíž se již přes 15 let na Masarykově univerzitě věnuje v rámci transdisciplinárních výzkumných projektů, často s využitím eye trackingu. Dlouhodobě úzce spolupracuje zjm. s kolegy z oblasti kartografie a informatiky. V posledních letech v jeho práci převládá základní i aplikovaný výzkum možností a limitů imerzivní virtuální reality a 3D zobrazování prostorových dat. V současnosti se jako řešitel či spoluřešitel podílí na výzkumných projektech s tématy sahajícími od psychologické diagnostiky, diagnostiky dyslexie pomocí eyetrackingu, přes vzdělávání v imerzivní VR až po studie interkulturních rozdílů v kognici.

Prednáška: Grafické reprezentace jako extenze kognitivního aparátu: cesta ze dvou dimenzí do imerzivní virtuální reality

První externí grafické reprezentace se objevily před několika deseti-tisíci lety. Od té doby jsou nedílnou součástí lidstva jako druhu a představují jeho externí kognitivní systém. Převážnou dobu historie lidstva byl způsob reprezentace vázán na dvourozměrné médium. V posledních letech, díky technologii imerzivní virtuální reality, je možné jevy zobrazovat ve třech rozměrech. Jaké možnosti nabízí tato technologická revoluce? A dokáže náš kognitivní aparát evolučně "zamrzlý" ve 2D zobrazeních efektivně pracovat s dalším rozměrem?

Trial-to-trial Contextual Adaptation in Sound Localization

Gabriela Andrejková, Stanislava Linková, Norbert Kopčo

P. J. Šafárik University in Košice
Perception and Cognition Laboratory, Institute of Computer Science
Jesenná 5, 04001 Košice
gabriela.andrejkova@upjs.sk, norbert kopco@upjs.sk

Abstract

Contextual plasticity (CP) is a localization aftereffect occurring on the time scale of seconds to minutes. It has been observed as a bias in horizontal sound localization of click target stimuli presented alone, when interleaved with contextual adapter-target trials in which the adapter was at a fixed location while the target location varied. The observed bias is always away from the contextual adapter location. This was confirmed for both real and virtual environments, for 1-click target sounds presented from different azimuths and interleaved with a 12-click adapter presented from a fixed position (Linková et al., 2022).

In the current study, we investigated the short-term dynamics of the adaptation by analyzing the effect of the stimulus immediately preceding a given target. Because the adapter is in a fixed position and contains more energy than the target, it was expected to induce stronger biases. The results confirmed this expectation, but only in a virtual environment, while a small opposite trend was observed in a real environment. These results illustrate complex interactions between environ-mental factors and stimuli in spatial auditory plasticity.

1 Introduction

The adaptation induced by preceding stimulation has been studied by many authors. E.g., Freyman et al. (1991) examined the precedence buildup induced by repeated presentation of 'lead-lag' stimulus pairs. Other studies looked for the auditory localization aftereffect induced by prolonged presentation of an adapter (Carlile et al. (2001); Phillips and Hall (2005); Thurlow and Jack (1973)). The latter studies typically used a long continuous adapter immediately followed by a target. They observed a repulsion by the adapter, i.e., biases in the perceived target locations away from the adapter location. Kopčo et al. (2007) described a related effect, CP, in which a repeated presentation of a short 1-click distractor affects target localization in a reverberant and anechoic room. Recent CP studies showed that CP can be induced by passive listening in both real and virtual environments (Linková et al. (2022); Piková (2018)). In the current study we examined the short-term dynamics of CP by analyzing the influence of the immediately preceding trial on the localization of the subsequent target stimulus

2 Methods

Data from two experiments of Linková et al. (2022) were used. The target (T) was a 2-ms noise burst (click). The adapter (A) was a click train consisting of 12 such clicks. Six target locations were used, $\pm 33, \pm 22, \pm 11°$. Adapter locations were fixed within a block at $0, \pm 45$, or $\pm 90°$in Exp. 1 and 0 or $\pm 50°$in Exp. 2 (setup in Fig. 1). Baseline blocks contained no adapters. Subjects - 8 in Exp. 1, real reverberant (RRE), and 9 in Exp. 2, virtual reverberant (VRE), anechoic (VAE) environment - responded by using a numerical keypad while seated with their heads supported by a headrest.

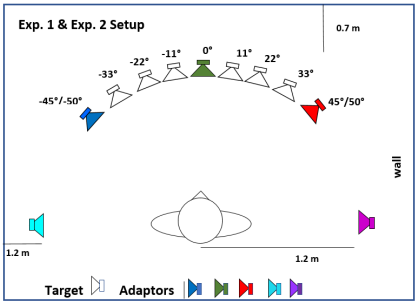


Fig. 1: Setup of experiments.

3 Results

We analyzed the biases in target responses separately for targets in adapter-target (AT) and target-target (TT) trial pairs. Fig.2 shows, for the three environments, the biases in responses (re. no-adapter baseline) as a function of target location after mirror-flipping the data assuming left-right symmetry in the effects. The thick line shows the deviations in AT and the thin line shows the deviations in TT data, with line color corresponding to different adapter locations. Because of the mirror-flip-ping, the red data are identical to blue after rotating around $0°$, and green data are also symmetrical around (0, 0).

Overall, the pattern of results is consistent across the environments and adapter locations. There are biases away from the adapter that decrease with separation between adapter and target (e.g., blue lines decrease from left to right; green lines are positive on the right-hand side). The immediately preceding stimulus has a modulator effect, mainly in virtual environments.

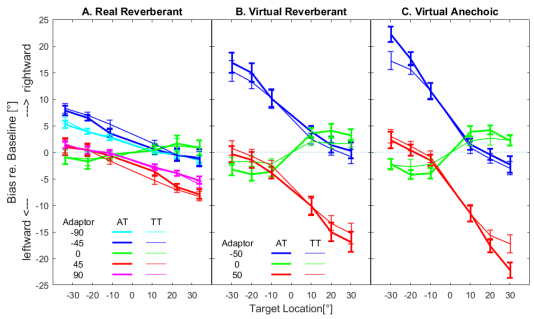


Fig. 2: Deviations in responses to target (T) relative to the baseline according to the previous type of target (TT) or adapter stimulus (AT).

The strongest effect of preceding trial is for the lateral adapter and nearby targets in VAE, where the difference between AT and TT data is $5°$(blue lines at $-30°$). Similar, but weaker effects are also observed in VRE and for the medial adapter (green lines). On the other hand, in RRE, the immediately preceding A causes, if anything, a smaller CP, only observed for the $45°$A and targets at $\pm 11°$. These results were confirmed by two ANOVAs in all environments. In ANOVA performed on virtual-environment data, the following interactions with the type of preceding stimulus were significant: *adapter x previous stimulus*: F(2, 16)=7.56, p=0.005, *environment x previous stimulus*: F(1, 8) =5.76, p=0.043, *target x previous stimulus*: F(2, 16)=5.10, p=0.02, *adapter x target x previous stimulus*: F(2, 16)=0.36, p = 0.703, *target x environment x previous stimulus*: F(4, 32)=4.41, p=0.006. In ANOVA performed on RRE data were significant *adapter*: F(4, 28) =38.42, p=0.00, *target*: F(2, 14) =6.57, p=0.009, *adapter x target*: F(8, 56) =5.45, p=0.00, *adapter x target x previous stimulus*: F(8, 56) =3.13, p=0.005.

Evaluation of standard deviations (SD) showed a complex pattern in RRE, while no significant effect was observed in VAE or VRE (data not shown). In RRE, SDs tended to be significantly higher when the previous stimulus was T vs. when it was A. However, that pattern only held for central targets and not for peripheral. ANOVA performed on the RRE data found significant main effect of the *previous trial type*: F(1, 7)=16.02, p=0.005, and *adapter*: F(4, 28)=4.17, p=0.009, a significant interaction *previous trial x target*: F(2, 14)=9.21, p=0.0028, and a nearly significant interaction: *previous trial x adapter*: F(4, 28)=2.60, p=0.058.

4 Discussion and conclusion

The results showed that the adapter affected target localization even in the short term. The biases increased immediately after A presentation in a virtual environment. This could be caused by the subjects' uncertainty in the localization of the stimuli, as a consequence of which the subjects responded mainly relatively with respect to the adapter which acted as an anchor. In terms of SD, the immediately preceding stimuli only affected responses in real environment. Again, it is likely that the 12 clicks of the A helped to stabilize spatial perception, making the responses to the subsequent T were more stable. On the contrary, if a target consisting of one click only was played from a varying location, the listener's spatial perception did not stabilize, and thus the variability in locating the subsequent target was high. These results show that CP has a fast component, observable on the time scale of several seconds, sup-porting the idea that CP is likely caused at least partially by a short-term suppression in spatial representation.

Acknowledgement

Work supported by VEGA 1/0350/22.

References

Carlile, S., Hyams, S. and Delaney, S. (2001). Systematic distortions of auditory space perception following prolonged exposure to broadband noise. *J. Acoust. Soc. Am.*, 110(1):416–424.

Freyman, R. L., Clifton, R. K. and Litovsky, R. Y. (1991). Dynamic processes in the precedence effect. *J. Acoust. Soc. Am.*, 90(2):874–884.

Kopčo, N., Best, V. and Shinn-Cunningham, B. G. (2007). Sound localization with a preceding distractor. *J. Acoust. Soc. Am.*, 121(1):420–432.

Linková, S., Andrejková, G. and Kopčo, N. (2022). Contextual plasticity in sound localization vs. source separation in real and virtual environments. V *Kognícia a umelý život XX, Třešť*, str. 162–163.

Phillips, D. P. and Hall, S. E. (2005). Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level. *Hearing Res.*, 202:188–199.

Piková, V. (2018). Mechanizmy kontextuálnej plasticity v lokalizácii zvukov. V *Bachalor Thesis, Faculty of Science, UPJŠ Košice*, str. 45.

Thurlow, W. R. and Jack, C. E. (1973). Some determinants of localization-adaptation effects for successive auditory stimuli. *J. Acoust. Soc. Am.*, 53(6):1573–1577.

Comparison of people with and without COVID-19 in their unfounded beliefs

Eva Ballová Mikušková, Peter Teličák

Ústav experimentálnej psychológie, Centrum spoločenských a psychologických vied SAV,
Dúbravská cesta 9, Bratislava
eva.ballova-mikuskova@savba.sk, peter.telicak@savba.sk

Abstract

The aim of the study was to examine changes in unfounded beliefs about COVID -19, powerlessness, and well-being during the pandemic and how people with varying severity of COVID -19 symptoms differed in these variables. A total of 1,420 adults aged 18-85 years (M=46.88±16.01) participated in two waves measuring demographics, severity of COVID -19 and powerlessness, well-being, and levels of unfounded beliefs. The unfounded beliefs did not change, powerlessness increased, and well-being decreased. There were no differences in unfounded beliefs between persons with no or mild symptoms and those with moderate or severe symptoms, although powerlessness was lower and well-being was higher in people with no or mild symptoms.

1 Introduction

Various forms of unfounded beliefs related to COVID-19 began to spread during the pandemic (WHO, 2020). At the same time, apart from health problems, the pandemic caused a mental health decrease (Brooks et al., 2020) through the increase of anxiety, stress, and depression (Wang, et al., 2020). And most importantly, it turned out that people experiencing anxiety (Chen et al., 2020) or powerlessness are more susceptible to unfounded beliefs (Abalakina-Paap, et al, 1999; Chen et al., 2020). The results of a recent longitudinal study indicated that beliefs in conspiracy theories are relatively stable over time (Williams, et al., 2022). The aim of the study was to examine changes in unfounded beliefs about COVID-19, powerlessness, and well-being during the pandemic and how people with varying severity of COVID-19 symptoms differed in these variables.

2 Methods

Data were collected in two waves (T1 - Fall 2021, T2 - Spring 2022) using an online questionnaire created in Qualtrics. A total of 1,420 adults (53.6% women) aged 18-85 years (*M*=46.81±15.95) completed demographic questions, COVID-19 *health status* (1=not infected, 2=mild symptoms, 3=moderate symptoms, 4=severe symptoms or hospitalisation), 4 items measuring *powerlessness* (Šrol et al., 2021), the *Satisfaction with life scale* (SWLS; Diener et al., 1985) and 18 items of *Scale of Covid-19 unfounded beliefs* (C19-UB; Teličák & Halama, 2022).

3 Results

The descriptive statistics are presented in Table 1.

Tab. 1: Descriptive statistics

	M	SD	min	max	α
age	46.88	16.01	18	85	-
T1 powerlessness	3.45	1.49	1	7	.87
T2 powerlessness	3.23	1.63	1	7	.91
T1 well-being	3.95	1.37	1	7	.90
T2 well-being	3.88	1.45	1	7	.91
T1 unfounded beliefs	2.41	1.02	1	5	.96
T2 unfounded beliefs	2.40	1.07	1	5	.96

Note: T1 – fall 2021, T2 – spring 2022, M – mean, SD – standard deviation, α – Cronbach's alpha

The level of unfounded beliefs did not change, powerlessness increased and well-being decreased from T1 – fall 2021 to T2 – spring 2022 (Table 2).

Tab. 2: Differences in powerlessness, well-being, and unfounded beliefs over time

	difference		t	p	d
	M	SD			
T1 powerlessness T2 powerlessness	0.22	1.60	5.35	<.001	0.14
T1 well-being T2 well-being	0.07	1.15	2.26	.024	0.06
T1 unfounded beliefs T2 unfounded beliefs	0.01	0.58	0.86	.391	0.02

Note: T1 – fall 2021, T2 – spring 2022, t – t-test value, p – the level of significance, d – Cohen's d

Participants were divided into two groups: those with no or mild symptoms from COVID-19 and those with moderate or severe symptoms. There were no

differences in unfounded beliefs between these groups, although powerlessness was lower and well-being was higher in those with no or mild symptoms (Table 3).

Tab. 3: Differences in powerlessness, well-being, and unfounded beliefs between people without or with mild symptoms and those with moderate or severe symptoms (Spring 2022)

	difference				
	M	SE	t	p	d
powerlessness	-0.34	0.10	0.567	.001	-0.21
well-being	0.23	0.09	3.256	.014	.0.16
unfounded beliefs	-0.04	0.07	2.449	.571	-0.04

Note: M – mean, SE – standard error, t – t-test value, p – the level of significance, d – Cohen's d

4 Discussion

We found a reduction in the perception of powerlessness during the pandemic. Although the effect was weak, a certain trend of reducing feelings of powerlessness may be explained by the gradual easing of pandemic measures. Nevertheless, the perception of well-being decreased during the pandemic. A possible explanation is the exhaustion of the population after the winter emergency measures were introduced. However, this effect needs to be reflected against the backdrop of a weak to negligible effect. Finally, beliefs in unfounded beliefs remained unchanged and similar trends were highlighted in the findings by Williams et al. (2022). The results further suggest that people with moderate to severe symptoms felt more powerless and had low well-being compared to people with no or mild symptoms (there were weak effects). However, the difference in belief in unfounded beliefs was statistically insignificant.

Acknowledgement

This research was supported by the Slovak Research and Development Agency under contract No. APVV-20-0387.

References

Abalakina-Paap, M., Stephan, W. G., Craig, T., & Gregory, W. L. (1999). Beliefs in conspiracies. *Political Psychology*, *20*(3), 637–647. https://doi.org/10.1111/0162-895X.00160

Asmundson, G. J. G., & Taylor, S. (2020). How health anxiety influences responses to viral outbreaks like COVID-19: What all decision-makers, health authorities, and health care professionals need to know. In *Journal of Anxiety Disorders* (Vol. 71). https://doi.org/10.1016/j.janxdis.2020.102211

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: rapid review of the evidence. In *The Lancet*, 395(10227). https://doi.org/10.1016/S0140-6736(20)30460-8

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, 49(1), 71–75. https://doi.org/10.1207/ s15327752jpa4901_13

Grzesiak-Feldman, M. (2013). The Effect of High-Anxiety Situations on Conspiracy Thinking. *Current Psychology*, *32*(1), 100–118. https://doi.org/10.1007/s12144-013-9165-6

Halama, P., & Teličák, P. (2022). Konštrukcia a psychometrická analýza vlastností škály Covid-19 nepodložených presvedčení (C19-NP). In P. Halama & V. Čavojová (Eds.), *Prežívanie a dôsledky pandémie COVID-19 na Slovensku*. Ústav experimentálnej psychológie, CSPV SAV.

Chen, X., Zhang, S. X., Jahanshahi, A. A., Alvarez-Risco, A., Dai, H., Li, J., & Ibarra, V. G. (2020). Belief in Conspiracy Theory about COVID-19 Predicts Mental Health and Well-being: A Study of Healthcare Staff in Ecuador. *JMIR Public Health and Surveillance*. https://doi.org/10.2196/20737

Klofstad, C. A., Uscinski, J. E., Connolly, J. M., & West, J. P. (2019). What drives people to believe in Zika conspiracy theories? *Palgrave Communications*, *5*(1), 1–8. https://doi.org/10.1057/s41599-019-0243-8

Šrol, J., Mikušková, E. B., & Čavojová, V. (2021). When we are worried, what are we thinking? Anxiety, lack of control, and conspiracy beliefs amidst the COVID-19 pandemic. *Applied Cognitive Psychology*, 35, 720–729. https://doi.org/10.1002/acp.3798

Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., Ho, C. S., & Ho, R. C. (2020). Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *International Journal of Environmental Research and Public Health*, *17*(5), 1729–1764. https://doi.org/10.3390/ijerph17051729

Williams, M., Ling, M., Kerr, J. R., Hill, S., Marques, M. D., Mawson, H., & Clarke, E. J. R. (2022). To what extent do beliefs in conspiracy theories change over time? (psyarxi.com)

WHO. (2020). Coronavirus Disease 2019(COVID-19)Situation Report-86. *World Health Organization Bulletin*.

Zkoumání Sense of Agency metodami neinvazivní mozkové stimulace

Bečev, O.[1,*], Laskov, O.[1,2], Bakštein E.[1], Biačková, N.[1,2], Novák, T.[1,2], Mohr, P.[1,2], Klírová, M.[1,2]

[1] Národní ústav duševního zdraví, Klecany, Česko
[2] 3. lékařská fakulta, Univerzita Karlova, Praha, Česko

*Korespondenční autor: ondrej.becev@nudz.cz

Abstrakt

Sense of agency je pocit, že jsem tím, kdo způsobil danou akci. V této studii jsme zkoumali efekt nízkofrekvenční (1 Hz) a vysokofrekvenční (10 Hz, 20 Hz) rTMS stimulace pravého dolního parietálního laloku na sense of agency a potenciální roli gamma oscilací v tomto procesu. Participanti (N=16) podstoupili rTMS stimulaci, po které následoval úkol na sense of agency zahrnující pohybování kurzorem a externí manipulaci zpětné vazby.
Aplikace 20 Hz rTMS stimulace vedla ke snížení přesnosti rozpoznání vlastních a cizích akcí. Tento efekt byl způsoben poklesem kapacity rozpoznat externí manipulaci pohybu.

Úvod

Sense of agency je pocit, že jsem tím, kdo generuje akci: tedy, že způsobuji, že se něco pohne, nebo způsobím myšlenku v mém proudu vědomí (Gallagher, 2000), kromě inicování akce zahrnuje i pocit kontroly akce a skrze to i událostí ve vnějším světě (Haggard & Tsakiris, 2009). Za prožitkem sense of agency patrně leží multifaktoriální systém, který podle kontextu dynamicky mění váhy jednotlivých vstupních signálů (Synofzik et al., 2008, 2013), mezi které patří např. motorické intence (Vosgerau & Synofzik, 2012), senzorické signály a časové predikce (Gentsch et al., 2012; SanMiguel et al., 2013), afektivní valence (Gentsch & Synofzik, 2014; Majchrowicz & Wierzchoń, 2021), nebo interocepce (Bečev et al., 2022; Marshall et al., 2018; Seth et al., 2012).
Neurozobrazovací studie nasvědčují, že sense of agency je spojené s parieto-frontálním okruhem (Zito, Wiest, et al., 2020). Zejména oblast identifikovaná jako rIPL nebo TPJ[1] bývá spojována s reprezentací motorických akcí (Bečev et al., 2021; Gallese et al., 2002) a s komparací motorického plánu vůči výsledkům akce (Gallese, 2005). Pozitivní sense of agency byl spojen se zvýšeným gamma coupling mezi rIPL a preSMA[2] (Ritterband-Rosenbaum, Nielsen, et al., 2014), což koresponduje s teoriemi o klíčové funkci gamma oscilací v parietálním laloku pro sense of agency (Guggisberg et al., 2011; Guggisberg & Mottaz, 2013).
Provázanost rIPL se sense of agency potvrdila také řada neurostimulačních studií, jejich nálezy jsou však do jisté míry rozporuplné. Aplikace *excitatorní* 10 Hz rTMS stimulace zvýšila reportování vlastních akcí jako externě manipulovaných (Ritterband-Rosenbaum, Karabanov, et al., 2014), stejný efekt mělo ale také aplikování jediného pulzu během self/other úlohy (Preston & Newport, 2008) nebo dokonce aplikace *inhibitorní* stimulace na stejnou oblast (Zito, Anderegg, et al., 2020). Intenzivní přímá elektrická stimulace vedla k iluzornímu pocitu provedení akce bez faktického pohybu (Desmurget et al., 2009).
V této studii jsme zkoumali úlohu pravého IPL na sense of agency prostřednictvím aplikace rTMS stimulace a sledování změn výkonu v behaviorální úloze s on-line reportováním agence. Kromě placeba byly aplikovány jak nízkofrekvenční (1 Hz), tak vysokofrevekvenční (10 Hz) protokoly vycházející ze starší literatury (Preston & Newport, 2008; Ritterband-Rosenbaum, Karabanov, et al., 2014). Kromě toho byla aplikována také vysokofrekvenční 20 Hz stimulace, u které se předpokládá indukce gamma oscilací v cílové oblasti (Honda et al., 2021; Liu et al., 2022; Noda et al., 2017), a doposud nebyl její efekt na sense of agency zkoumán.

Materiály a metody

1.1 Participanti a experimentální design

Zdraví participanti, praváci (N=16, 10 žen, průměrný věk 31.3 let; SD =10.1) se účastnili experimentu, ve kterém byla aplikována repetitivní transkraniální stimulace (rTMS) a následně participanti plnili senzomotorickou úlohu založenou na ovládání kurzoru po hrací ploše s občasnou externí manipulací pohybu. Efekt stimulace a úlohy byl měřen pomocí encefalografie (EEG). Experiment byl navržen jako zaslepená, placebem kontrolovaná, vnitrosubjektová (within-subject) studie s designem 2x4x4 a faktory *typ úlohy* (sense of agency úloha, kontrolní úloha), *externí manipulace* (ano, ne) a *rTMS podmínka* (placebo, 1 Hz,

[1] TPJ - temporo-parietální junkce; rIPL – dolní parietální lalok, pravostranný.
[2] preSMA – presuplementární motorická oblast.

10 Hz, 20 Hz). Každý subjekt se účastnil 4 návštěv s nejméně týdenním odstupem, přiřazení podmínek k pořadí návštěvy bylo kontrabalancované.

1.2 Procedura

Každá návštěva zahrnovala zjišťování aktuálního emočního naladění na začátku a na konci session zkrácenou škálou POMS (Shacham, 1983; Stuchlíková et al., 2005), nahrávku baseline klidové EEG aktivity a dva sety zahrnující: aplikaci rTMS stimulace, behaviorální úlohu (sense of agency úloha, kontrolní úloha) a nahrávku post-intervenční klidové EEG aktivity při zavřených očích.

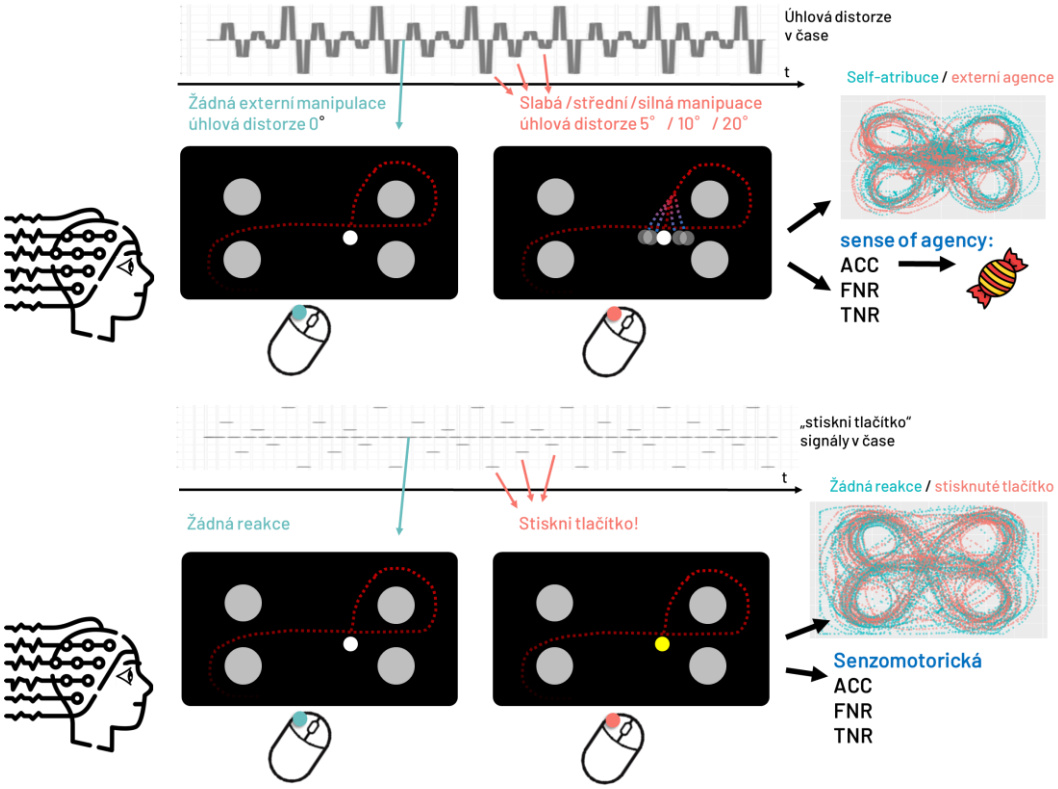
1.3 Sense of agency úloha a kontrolní úloha

Sense of agency úloha je adaptací úlohy dle (Kozáková et al., 2020) a zapojuje jak agentní, tak senzomotorické procesy. Subjekt kontroloval plynulý pohyb kurzoru na obrazovce a snažil se vyhýbat rozestaveným překážkám. Participanti byli instruováni, že výzkumníci jim mohou vzdáleně čas od času do pohybu zasahovat, což mají poznačit podržením tlačítka na myši. Hlavním úkolem participantů bylo snažit se dávat pozor a co nejpřesněji rozeznávat, kdy mají kurzor plně pod kontrolou a kdy je jim do pohybu zasahováno. Zásahy však byly ve skutečnosti řízeny počítačem podle specifického rozvrhu úhlových výchylek. Rozvrh střídal bloky bez zásahu (0° deviace) a bloky s různou měrou deviace (5°,10° a 20°) – viz Obr. 1. Kromě peněžní odměny po absolvování všech 4 návštěv byli účastníci motivováni sladkou odměnou podle jejich výkonu v absolvované sense of agency úloze.

Kontrolní senzomotorická úloha byla navržena jako co nejpodobnější k sense of agency úloze, ovšem bez manipulace agence. Participanti také ovládali kurzor mezi překážkami, vnější zásahy však nebyly přítomné a místo nich bylo úkolem participantů reagovat co nejrychleji na změnu barvy kurzoru.

1.4 Aparát

Obě úlohy byly prováděné na 15“ laptopu, ovládaném pomocí myši. Experimentální úlohy byly implementovány v prostředí Processing (Reas & Fry, 2006). RTMS stimulace byla prováděna prostřednictvím systému MagPro R30 (Magventure®, Inc., Denmark) s cívkou Cool D-B80 A/P (dvojitý kužel). RTMS bylo aplikována na oblast dolního parietálního laloku (IPL) pravostranně (na místo elektrody 173 dle pozic 256 kanálové čepici Magstim



Obr. 1: Schéma experimentální procedury. Sense of agency úloha (nahoře) a kontrolní senzomotorická úloha (dole). Vpravo je ukázka trajektorie pohybu participanta kolem překážek.

EGI HCGSN 100), s intenzitou 100% individuálního motorického prahu. EEG bylo pořizováno přístrojem Magstim EGI Net Amps 400 Geodesic EEG (Magstim EGI, Eugene, OR, USA) s 256 kanálovou čepicí. EEG data a výsledky v tomto textu dále nejsou reportována.

2 Výsledky

2.1 Sledované proměnné

Analýza sledovala tři hlavní závislé proměnné v sense of agency úloze:

a) **False negative rate** (*FNR*) je míra nesprávného reportování vlastních akcí jakožto externích.

b) **True negative rate** (*TNR*): správné reportování externích zásahů, non-agency

c) **Accuracy** (*ACC*) – celková přesnost klasifikace určení self- vs non-self událostí.

2.2 20 Hz rTMS na rIPL snižuje přesnost self-other distinkce

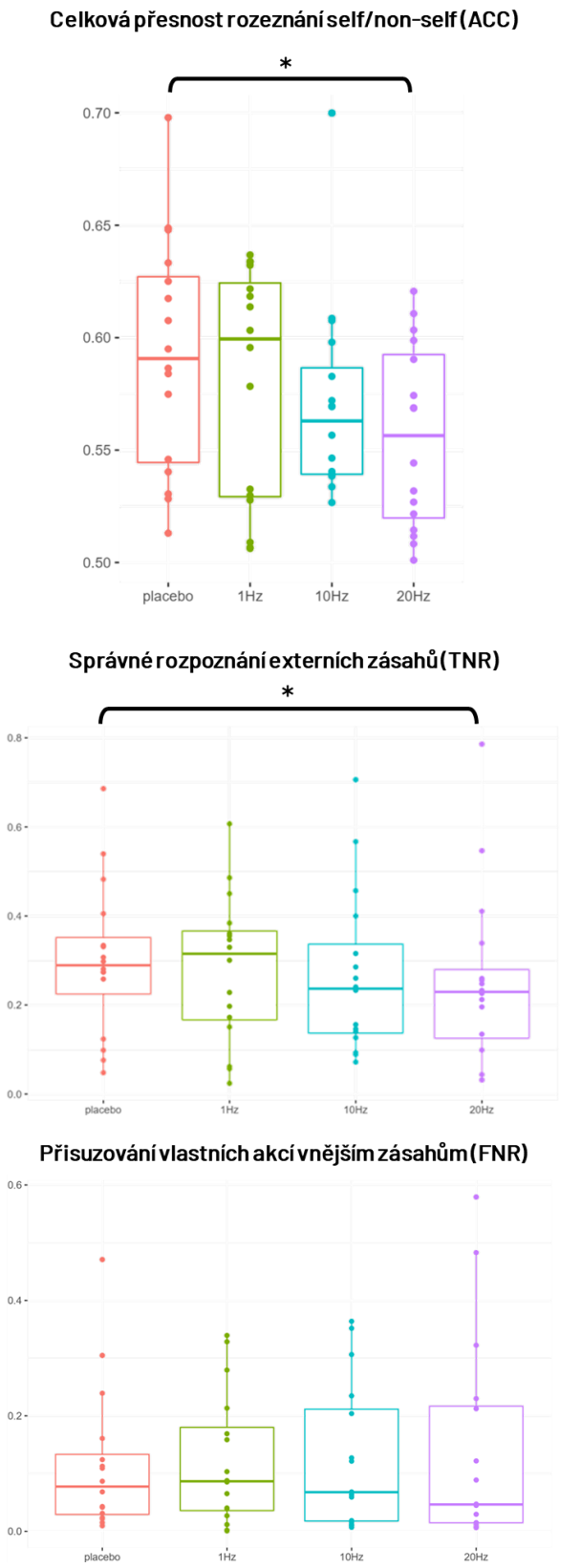
Hlavním zjištěním je, že po aplikaci 20 Hz rTMS stimulace subjekty vykazovaly vůči placebu výrazně sníženou přesnost (ACC) v rozeznávání self/other, $t(16)=-4{,}275$, $p=0{,}001$ (viz Obr. 2.). Tento pokles v přesnosti čerpal zejména ze signifikantního poklesu správného rozpoznání vnějších intruzí (TNR), $t(16)=-2{,}461$, $p=0{,}026$ a patrně částečně také ze simultánního nárůstu chybných označení vlastních akcí za cizí (FNR), $t(16)=1{,}314$, $p=0{,}209$. Tyto nálezy přitom není možné vysvětlit pouhým efektem senzomotorických schopností - párové srovnání efektu 20 Hz stimulace a placeba na přesnost u kontrolní úlohy není signifikantní $t(16)=0{,}531$, $p=0{,}603$.

2.3 Slabé zásahy jsou příliš náročné na postřehnutí

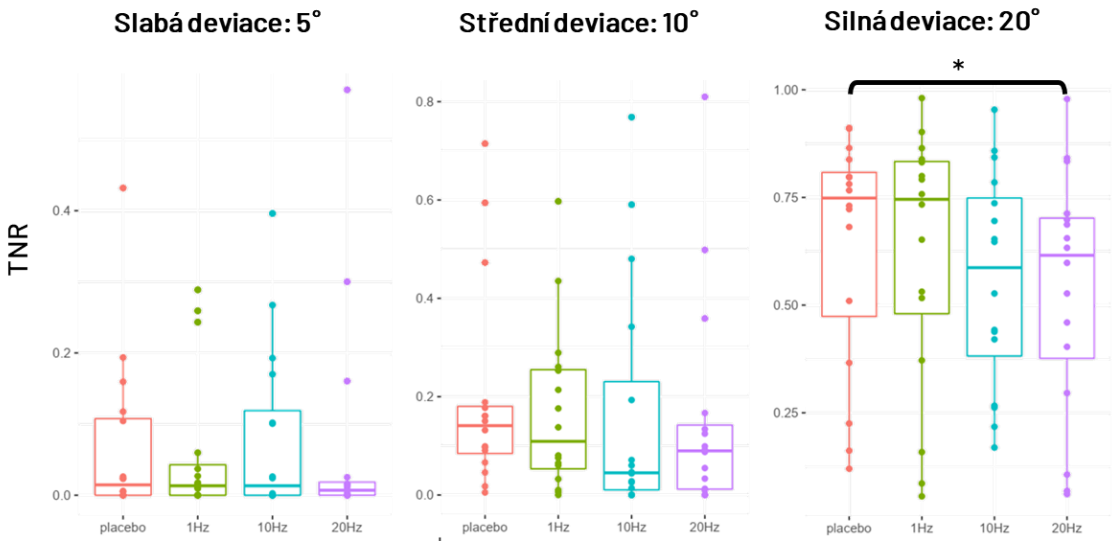
Efekt 20 Hz rTMS na TNR se projevuje nejvýrazněji na silných perturbacích (viz Obr. 3). Nejpravděpodobnějším důvodem je podlahový efekt: pro většinu participantů je příliš náročné zaregistrovat 5° deviaci a jejich reporty byly na míře náhody.

2.4 Data se rozchází s předchozími studiemi

Naše behaviorální výsledky, získané doposud na největším vzorku participantů, nereplikovaly po aplikaci vysokofrekvenční rTMS (tj. 10 Hz či 20 Hz) zvýšení FNR pozorované Ritterband-Rosenbaum et al. (2014) po vysokofrekvenční 10 Hz rTMS stimulaci. Kromě toho jsme nepozorovali ani efekt inhibitorní stimulace (1 Hz) na přesnost (ACC), reportovaný MacDonald & Paus (2003).



Obr. 2: Nahoře: srovnání ACC u Sense of Agency task mezi čtyřmi rTMS podmínkami. Uprostřed a dole: srovnání správného rozpoznání externích zásahů (TNR) a přisuzování vlastních akcí vnějším zásahům (FNR) mezi jednotlivými rTMS podmínkami.



Obr. 3: Porovnání ACC u Sense of Agency task mezi čtyřmi rTMS podmínkami. Podlahový efekt u 5° a 10° deviace je zřetelný.

2.5 Diskuse

Naše nálezy podporují hypotézu, že pravostranný IPL je klíčovou oblastí zapojenou do zpracování *sense of agency* a senzomotorické integrace.

Excitatorní 20 Hz rTMS na rIPL narušuje sense of agency, nicméně faktory, které k tomu vedou, jsou výrazně odlišné od starší studie Ritterband-Rosenbaum, Karabanov, et al. (2014) – dalo by se dokonce říci, že přímo opačné. V našich výsledcích je efekt saturovaný především narušením schopnosti detekovat cizí akce a je spojený s poklesem TNR - true negative rate, zatímco Ritterband-Rosenbaum, Karabanov, et al. (2014) po podobné stimulaci pozorovali častější označování vlastních akcí jako cizí (FNR). V naší studii 10 Hz ani 20 Hz stimulace stejné oblasti tento efekt nenavozuje.

První otázka se týká tohoto rozporu. Příčinou může být jiný časový rámec reportu agence: zatímco v naší studii participanti reportovali agenci kontinuálně, ve studii Ritterband-Rosenbaum participanti prováději reporty po dokončení pohybu. V naší studii jsme také aplikovali stimulaci off-line (před úlohou), zatímco Ritterband-Rosenbaum et al. ji aplikovali on-line (během pohybu).

Další otázkou je, zda použitá intervence narušila přímo proces *sense of agency*, nebo spíše jeho *klíčový vstupní faktor* – senzomotorickou komparaci. Miele et al. (2011) přichází s experimentálně podloženým tvrzením, že IPL je spíše zodpovědná za nízkoúrovňové, automatické zpracování „příznaků" agence a aktivuje se, když dojde k diskrepanci mezi vizuálními, proprioceptivními a motorickými vstupy. Podle nich IPL nejspíše není zapojená do vyšších kognitivních procesů zakládajících sense of agency, jako je vědomá reflexe, zda byla akce způsobena někým jiným, nebo mnou.

Pokud přijmeme, že funkce rIPL je spojená spíše se senzomotorickou komparací, nabízí se otázka, zda je výstupní informace z komparátoru spojená s detekcí *diskrepance* (mismatch), nebo naopak s detekcí senzomotorické *korespondence/kongruence*. Starší fMRI studie spojují aktivitu v rIPL s *diskrepancí* mezi zamýšlenou akcí a konsekvencemi pohybu (Farrer et al., 2008), s čímž souhlasí také nedávná metastudie Zito, Wiest, et al. (2020) připravená na neurozobrazovacích datech. Konstatování spojitosti aktivity rIPL s diskrepancí však může být důsledkem nízkého temporálního rozlišení neurozobrazovacích metod a z toho plynoucích problémů se zachycením rychlé dynamiky meziregionální komunikace. Ritterband-Rosenbaum, Nielsen, et al. (2014) kritizují reduktivní pohled na neurální podklady sense of agency, spočívající v hledání změn na úrovni jednotlivých míst v mozku (typické pro fMRI studie) a argumentují, že neurální signatury sense of agency mají spíše povahu změn síťového spárování. S použitím EEG a dynamického kauzálního modelování ukázali, že preSMA vytváří predikci předpokládaného výstupu akce, která je poté předávána do IPL, který následně komparuje tento předpokládaný stav s průběžnou vizuální a proprioceptivní zpětnou vazbou během provádění akce. Teprve v pozdní fázi zpracování dojde v případě senzomotorické korespondence v IPL k mezifrekvenčnímu spárování (coupling) v pásmu vysoké gammy, iniciovaného z IPL do preSMA. Prožitek sense of agency tedy patrně vzniká až v tomto kroku, když IPL indikuje do preSMA kongruenci a to formou přechodné modulace síly gamma aktivity v preSMA gamma aktivitou v rIPL.

Aktuálních poznatky tedy potvrzují dřívější předpoklad Synofzik et al., (2008), že rIPL komplexně vyhodnocuje korespondenci senzorických vstupů (reaference) se stavem predikovaným na základě eferentní kopie.

Poděkování

Tento příspěvek vznikl za podpory grantové agentury GA ČR, v rámci projektu č. 20-24782S. Díky patří také Nicol Schlezingerové a Lucii Krejčové, které se podílely na sběru dat a Janu Rydlovi a Janu Hubenému za technické zajištění.

Literatura

Bečev, O., Kozáková, E., Sakálošová, L., Mareček, R., Majchrowicz, B., Roman, R., & Brázdil, M. (2022). Actions of a Shaken Heart: Interoception Interacts with Action Processing. *Biological Psychology*, *169*(December 2021), 108288. https://doi.org/10.1016/j.biopsycho.2022.108288

Bečev, O., Mareček, R., Lamoš, M., Majchrowicz, B., Roman, R., & Brázdil, M. (2021). Inferior parietal lobule involved in representation of "what" in a delayed-action Libet task. *Consciousness and Cognition*, *93*(July 2020), 103149. https://doi.org/10.1016/j.concog.2021.103149

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science (New York, N.Y.)*, *324*(5928), 811–813. https://doi.org/10.1126/science.1169896

Farrer, C., Frey, S. H., Van Horn, J. D., Tunik, E., Turk, D., Inati, S., & Grafton, S. T. (2008). The angular gyrus computes action awareness representations. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(2), 254–261. https://doi.org/10.1093/cercor/bhm050

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21.

Gallese, V. (2005). The 'Shared Manifold' Hypothesis. *Journal of Consciousness Studies*, *8*, 33–50.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (2002). Action Representation and the inferior parietal lobule. In W. Prinz & B. Hommel (Eds.), *Common Mechanisms in Perception and Action: Attention and Performance XIX* (pp. 334–355). Oxford University Press.

Gentsch, A., Kathmann, N., & Schütz-Bosbach, S. (2012). Reliability of sensory predictions determines the experience of self-agency. *Behavioural Brain Research*, *228*(2), 415–422. https://doi.org/10.1016/j.bbr.2011.12.029

Gentsch, A., & Synofzik, M. (2014). Affective coding: the emotional dimension of agency. *Frontiers in Human Neuroscience*, *8*(August), 608. https://doi.org/10.3389/fnhum.2014.00608

Guggisberg, A. G., Dalal, S. S., Schnider, A., & Nagarajan, S. S. (2011). The neural basis of event-time introspection. *Consciousness and Cognition*, *20*(4), 1899–1915. https://doi.org/10.1016/j.concog.2011.03.008

Guggisberg, A. G., & Mottaz, A. (2013). Timing and awareness of movement decisions: does consciousness really come too late? *Frontiers in Human Neuroscience*, *7*(July), 385. https://doi.org/10.3389/fnhum.2013.00385

Haggard, P., & Tsakiris, M. (2009). The Experience of Agency. *Current Directions in Psychological Science*, *18*(4), 242–246.

Honda, Y., Nakamura, S., Ogawa, K., Yoshino, R., Tobler, P. N., Nishimura, Y., & Tsutsui, K. I. (2021). Changes in beta and high-gamma power in resting-state electrocorticogram induced by repetitive transcranial magnetic stimulation of primary motor cortex in unanesthetized macaque monkeys. *Neuroscience Research*, *171*, 41–48. https://doi.org/10.1016/j.neures.2021.02.002

Kozáková, E., Bakštein, E., Havlíček, O., Bečev, O., Knytl, P., Zaytseva, Y., & Španiel, F. (2020). Disrupted Sense of Agency as a State Marker of First-Episode Schizophrenia: A Large-Scale Follow-Up Study. *Frontiers in Psychiatry*, *11*(December), 1–9. https://doi.org/10.3389/fpsyt.2020.570570

Liu, C., Han, T., Xu, Z., Liu, J., Zhang, M., Du, J., Zhou, Q., Duan, Y., Li, Y., Wang, J., Cui, D., & Wang, Y. (2022). Modulating Gamma Oscillations Promotes Brain Connectivity to Improve Cognitive Impairment. *Cerebral Cortex*, *32*(12), 2644–2656. https://doi.org/10.1093/cercor/bhab371

MacDonald, P. A., & Paus, T. (2003). The role of parietal cortex in awareness of self-generated movements: A transcranial magnetic stimulation study. *Cerebral Cortex*, *13*(9), 962–967. https://doi.org/10.1093/cercor/13.9.962

Majchrowicz, B., & Wierzchoń, M. (2021). Sensory attenuation of action outcomes of varying amplitude and valence. *Consciousness and Cognition*, *87*(May 2020). https://doi.org/10.1016/j.concog.2020.103058

Marshall, A. C., Gentsch, A., & Schütz-Bosbach, S. (2018). The interaction between interoceptive

and action states within a framework of predictive coding. *Frontiers in Psychology*, *9*(FEB), 1–14. https://doi.org/10.3389/fpsyg.2018.00180

Miele, D. B., Wager, T. D., Mitchell, J. P., & Metcalfe, J. (2011). Dissociating neural correlates of action monitoring and metacognition of agency. *Journal of Cognitive Neuroscience*, *23*(11), 3620–3636. https://doi.org/10.1162/jocn_a_00052

Noda, Y., Zomorrodi, R., Saeki, T., Rajji, T. K., Blumberger, D. M., Daskalakis, Z. J., & Nakamura, M. (2017). Resting-state EEG gamma power and theta–gamma coupling enhancement following high-frequency left dorsolateral prefrontal rTMS in patients with depression. *Clinical Neurophysiology*, *128*(3), 424–432. https://doi.org/10.1016/j.clinph.2016.12.023

Preston, C., & Newport, R. (2008). Misattribution of movement agency following right parietal TMS. *Social Cognitive and Affective Neuroscience*, *3*(1), 26–32. https://doi.org/10.1093/scan/nsm036

Reas, C., & Fry, B. (2006). Processing: programming for the media arts. *Ai & Society*, *20*(4), 526–538.

Ritterband-Rosenbaum, A., Karabanov, A. N., Christensen, M. S., & Nielsen, J. B. (2014). 10 Hz rTMS over right parietal cortex alters sense of agency during self-controlled movements. *Frontiers in Human Neuroscience*, *8*(June), 1–7. https://doi.org/10.3389/fnhum.2014.00471

Ritterband-Rosenbaum, A., Nielsen, J. B., & Christensen, M. S. (2014). Sense of agency is related to gamma band coupling in an inferior parietal-preSMA circuitry. *Frontiers in Human Neuroscience*, *8*(July), 510. https://doi.org/10.3389/fnhum.2014.00510

SanMiguel, I., Todd, J., & Schröger, E. (2013). Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. *Psychophysiology*, *50*(4), 334–343. https://doi.org/10.1111/psyp.12024

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *3*(JAN), 1–16. https://doi.org/10.3389/fpsyg.2011.00395

Shacham, S. (1983). A Shortened Version of the Profile of Mood States. *Journal of Personality Assessment*, *47*(3), 305–306. https://doi.org/10.1207/s15327752jpa4703_14

Stuchlíková, I., Mana, F., & Hagtvet, K. (2005). Dotazník k měření afektivních stavů: konfirmačně faktorová analýza krátké české verze. *Československá Psychologie*, *49*(5), 459–467.

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and Cognition*, *17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: an interplay between prediction and postdiction. *Frontiers in Psychology*, *4*(March), 1–8. https://doi.org/10.3389/fpsyg.2013.00127

Vosgerau, G., & Synofzik, M. (2012). Weighting models and weighting factors. *Consciousness and Cognition*, *21*(1), 55–58. https://doi.org/10.1016/j.concog.2011.09.016

Zito, G. A., Anderegg, L. B., Apazoglou, K., Müri, R. M., Wiest, R., Holtforth, M. G., & Aybek, S. (2020). Transcranial magnetic stimulation over the right temporoparietal junction influences the sense of agency in healthy humans. *Journal of Psychiatry and Neuroscience*, *45*(4), 271–278. https://doi.org/10.1503/jpn.190099

Zito, G. A., Wiest, R., & Aybek, S. (2020). Neural correlates of sense of agency in motor control: A neuroimaging meta-analysis. *PLoS ONE*, *15*(6), 1–17. https://doi.org/10.1371/journal.pone.0234321

The relationship between education and non-normative political behaviour: The mediating effect of conspiracy beliefs

Juliána Bujňáková, Eva Ballová Mikušková

Ústav experimentálnej psychológie, Centrum spoločenských a psychologických vied SAV, v.v.i.
Dúbravská cesta 9, Bratislava
juliana.bujnakova@savba.sk, eva.ballova-mikuskova@savba.sk

Abstract

Endorsement of conspiracy beliefs – associated with lower education level – could have consequences that can take many forms, one of which is non-normative political behaviour. The aim of the study was to examine whether conspiracy beliefs and conspiracy mentality could mediate the relationship between education and non-normative political behaviour. Study sample consisted of 1682 participants (903 women) aged 18–85 years from Slovakia. They answered questions about their beliefs in conspiracies, conspiracy mentality, and possible non-normative political behaviours. Although there was some mediating effect of conspiracy mentality on non-normative political behaviour, the mediating effect of conspiracy beliefs appears to be even stronger in this study.

1 Introduction

Conspiracy theories, described as attempts to explain events as the results of secret plots of powerful forces (Douglas & Sutton, 2008; Goertzel, 1994) have always been part of our history and seem to be more prominent in times of social crisis (Douglas et al., 2019), such as recent COVID- 19 pandemic.

Conspiracies tend to blame political groups for negative political and economic events, which puts to question their legitimacy and, when combined with conspiracy mentality – tendency to see the world as driven by secret plots (Bruder et al., 2013) – can lead to increased non–normative political behaviour (Imhoff et al., 2021). Because research shows that higher education is linked to lower endorsement of conspiracy theories (Douglas et al., 2016), and lower political violence (Østby et al. 2019), the aim of our study was to examine whether conspiracy beliefs and conspiracy mentality can mediate the relationship between education and non – normative political behaviour.

2 Methods

A total of 1682 participants (903 women) aged 18–85 years (*M*=46.1, *SD*=16.1) were recruited by the agency to be representative of Slovak population in terms of age, sex and education completed the online survey (data were collected as the first wave of a larger longitudinal study). Participants answered demographic questions, 18 items of *Scale of Covid-19 unfounded beliefs* (C19-UB; Halama & Teličák 2022; *M*=2.2; *SD*=1.11), *Conspiracy Mentality Questionnaire* (CMQ, Bruder et al. 2013; *M*=4.01; *SD*=1.82), and whether they would be willing to engage in some forms of non–normative political behaviour (Imhoff et al., 2021 *M*=1.32; *SD*=0.57).

3 Results

First, the mediation analysis shows that conspiracy mentality has statistically significant ($p < .001$) effect on the relationship between education and non – normative political behaviour (Figure 1, Table 1). Indirect effect explains 37.8 % of the relationship, but direct effect is even stronger (62.2 %) in this analysis.

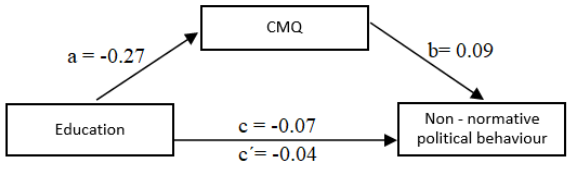


Figure 1: Mediation analysis of the relationship between education and non – normative political behaviour.

Effect	Estimate	SE	Z	p	% Mediation
Indirect	-0.03	0.01	-7.78	< .001	37.8
Direct	-0.04	0.01	-4.98	< .001	62.2
Total	-0.07	0.01	-7.82	< .001	100.0

Table 1: Indirect, direct and total effects of mediation analysis between education and non – normative political behaviour.

Secondly, the analysis shows that mediating effect of conspiracy beliefs is statistically significant ($p < .001$). Conspiracy beliefs explain 60.1 % of the relationship

between education and non – normative political behaviour (Figure 2, Table 2).

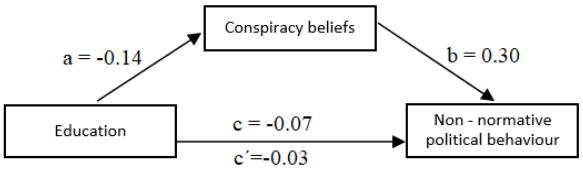


Figure 2: Mediation analysis of the relationship between education and non – normative political behaviour.

Effect	Estimate	SE	Z	p	% Mediation
Indirect	-0.04	0.01	-9.93	< .001	60.1
Direct	-0.03	0.01	-3.24	0.001	39.9
Total	-0.07	0.01	-7.82	< .001	100.0

Table 2: Indirect, direct and total effects of mediation analysis between education and non – normative political behaviour.

Discussion

The aim of this study was to explore possible mediating effect of conspiracy mentality and conspiracy beliefs on the relationship between education and non–normative political behaviour. As shown by the results, conspiracy mentality can explain 37.8 % of the model's total effect. Conspiracy beliefs proved to be even stronger mediator in this study, explaining 60.1 % of the relationship. These findings are in line with previous studies, that found a relationship between conspiracy beliefs and non – normative political behaviour (e.g. Douglass, 2021; Imhoff & Bruder, 2014). Our results also suggests that belief in specific conspiracy claims may have stronger effect on non – normative political behaviour than general conspiracy worldview. Future research could investigate whether similar results can be obtained by examining different conspiracies, given that COVID-19 related conspiracy theories were utilized in this study. There is also a question, whether education truly is a predictor in this relationship, or should be considered its different role in this relationship.

Acknowledgement

This research was supported by the Slovak Research and Development Agency under contract No. APVV-20-0387.

References

Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy Mentality Questionnaire. *Frontiers in Psychology, 4(225), 1–15.* https://doi.org/10.3389/fpsyg.2013.00225.

Douglas, Karen M. (2021).“Are Conspiracy Theories Harmless?” *The Spanish Journal of Psychology 24: e13*. doi:10.1017/SJP.2021.10.

Douglas, K.M., & Sutton, R.M. (2008). The hidden impact of conspiracy theories: Perceived and actual influence of theories surrounding the death of Princess Diana. *Journal of Social Psychology, 148, 210–222.*

Douglas, K. M., Sutton, R. M., Callan, M. J., Dawtry, R. J., & Harvey, A. J. (2016). Someone is pulling the strings: Hypersensitive agency detection and belief in conspiracy theories. *Thinking & Reasoning, 22(1),* 57–77. https://doi.org/10.1080/13546783.2015.1051586.

Douglas K.M., Uscinski J.E., Sutton R.M., Cichocka A., Nefes T., Ang C.S. & Deravi F. (2019). Understanding conspiracy theories. *Advanced Political Psychology, 40(3), 35. https://doi.org/10.1111/pops.12568.*

Goertzel, T. (1994). Belief in conspiracy theories. *Political Psychology, 15, 731–742.*

Halama, P., & Teličák, P. (2022). Konštrukcia a psychometrická analýza vlastností škály Covid-19 nepodložených presvedčení (C19-NP). In P. Halama & V. Čavojová (Eds.), Prežívanie a dôsledky pandémie COVID-19 na Slovensku. Ústav experimentálnej psychológie, CSPV SAV.

Imhoff, R., & Bruder, M. (2014). Speaking (un-) truth to power: Conspiracy mentality as a generalised political attitude. *European Journal of Personality,28,25–43.* http://dx.doi.org/10.1002/per.1930.

Imhoff, R., Dieterle, L., & Lamberty, P. (2021). Resolving the Puzzle of Conspiracy Worldview and Political Activism: Belief in Secret Plots Decreases Normative but Increases Nonnormative Political Engagement. *Social Psychological and Personality Science, 12(1), 71–79.* https://doi.org/10.1177/1948550619896491

Østby, G., Urdal, H., & Dupuy, K. (2019). Does Education Lead to Pacification? A Systematic Review of Statistical Studies on Education and Political Violence. Review of Educational Research, 89(1), 46–92.

Etika letálních autonomních vojenských robotů

David Černý

Centrum Karla Čapka pro studium hodnot ve vědě a technice
Ústav informatiky AV ČR, v. v. i.
Pod Vodárenskou věží 2, 18207 Praha, Česká republika
Email: david.cerny@ilaw.cas.cz

Abstrakt

V tomto příspěvku se zaměřuji na stručnou etickou reflexi důvodů pro a proti nasazování autonomních letálních vojenských robotů v reálných podmínkách válečných konfliktů. Tyto důvody jsou primárně konsekvencialistické, tj. věnují se benefitům a rizikům spojeným s touto moderní vojenskou technologií.

1 Úvod

Možná prvním vojenským robotem v naší západní kulturní imaginaci byl měděný obr Tálos, kterého na Diův příkaz vytvořil bůh Héfaistos, tvůrce mnoha mechanických divů, které dnes můžeme považovat za roboty (Mayor, 2018). Od doby řeckých mýtů však uplynulo mnoho času a možnost vytvořit plně autonomního vojenského robota, který dokáže plnit všechny bojové úlohy bez zásahu člověka, je již realitou. Mnozí lidé to z morálního hlediska považují za chybný krok, který bychom měli napravit úplným mezinárodním zákazem jejich vývoje a nasazování v reálných bojových podmínkách. Takové zákazy jsou však nerealistické; těžko můžeme čekat, že tak strategicky významné technologie by se státy skutečně vzdaly. Je proto velmi důležité snažit se o její etickou reflexi a nastavení podmínek jejich využívání, které budou v souladu s moderní doktrínou spravedlivé války.

V tomto příspěvku se nemohu věnovat celé problematice etické reflexe autonomních robotů ve službě armády. Představím proto jen některé, primárně konsekvencialistické, důvody pro jejich nasazování a některé důvody v jejich neprospěch. Komplexní uchopení celé problematiky by si vyžádalo zpracování formou monografie.

2 Letální vojenští roboti

Ani mezi odborníky neexistuje jednoznačná definice robota. Jak říká americký robotik Illah Nourbakhsh (Nourbakhsh, 2013):

> Nikdy se robotika neptejte, co je to robot. Odpověď se mění příliš rychle. V okamžiku, kdy vědci ukončí nejnovější diskusi o tom, co je a co není robot, zrodí se zcela nová interakční technologie a hranice se posune dál.

K určité pracovní definici robotů se nejlépe dopracujeme prostřednictvím pojmu agenta. Budeme ho definovat následujícím způsobem:

- **Agent**. Agentem rozumíme systém, jenž vnímá své okolí prostřednictvím senzorů a interaguje se světem pomocí aktuátorů.

Nejjednoduššího agenta si můžeme představit jako systém, který je vybaven senzory, aktuátory, a specifickým programem (*agent program*). Senzory zprostředkovávají kognitivní kontakt s prostředím formou perceptů, jejichž úplná historie se označuje jako sekvence perceptů. Agentský program je konkrétní implementací agentské funkce, jež zobrazuje percepty na množinu jednání. Agent může disponovat velmi jednoduchou agentskou funkcí ve formě „jestliže… pak“, např. „jestliže je před tebou překážka, zahni vpravo“., její podoba však může být výrazně sofistikovanější.

V zájmu jasnosti je třeba rozlišovat mezi vojenskými roboty a vojenskými systémy bez posádky (*unmmaned systems*).; nadále je budeme označovat jako US. Některé US mohou být chápány jako vojenští roboti, jiné zase ne. Dron, který je po celou dobu letu řízený nějakým operátorem, není robotem; pokud však v některých částech své mise jedná na operátorovi nezávisle, lze ho v nich považovat za robota.

Některé roboty lze chápat jako US (nemají-li posádku), jiné zase ne. A některé US jsou roboty (alespoň v některé části svého pracovního cyklu), jiné zase ne. Obě třídy – robotů a US – mají neprázdný průnik, vojenské roboty však nelze definovat jako systémy umělé inteligence bez posádky. Předběžná definice vojenských robotů může vypadat následovně:

- **Vojenský robot.** Vojenský robot je robot, tj. ve fyzickém světě existující agent, který
  1. vnímá své prostředí;
  2. zpracovává své percepty a vytváří si plány jednání;
  3. využívá své aktuátory k aktualizaci vybraného jednání;

4. je alespoň v určité části svého operačního cyklu autonomní;

5. plní vojenské cíle.

Podle toho, jaké cíle vojenští roboti plní lze rozlišit celou řadu robotů, nás zde ale budou zajímat plně autonomní letální vojenští roboti (AVR), tj. takoví roboti, kteří dokáží splnit svůj úkol, včetně výběru a likvidace cíle, bez zásahu člověka.

Definovat autonomii robotů není snadný úkol. K bližší definici autonomie musíme rozlišit mezi systémem S, úkolem t, systémem S^*, vůči němuž je S v nějaké vztahu závislosti či nezávislosti, a konečně prostředím e, v němž má S splnit svůj úkol t (Tamburini, 2016). Smyslem zavedení systému S^* je rozlišit situace, kdy S závisí na různých typech systémů, např. na jiném systému umělé inteligence, či na lidském operátoru. Autonomie vojenských robotů je tedy čtyřčlenou strukturou $< S, t, S^*, e >$. Například robotický systém SGR-A1 společnosti Samsung může nahradit stráže na hranicích mezi Severní a Jižní Koreou a operovat v částečně či plně autonomním režimu. Pokud by operoval v plně autonomním režimu, dokázal by identifikovat cíle, sám se rozhodnout je zneškodnit a posléze tak učinit. Neexistoval by tedy žádný systém kontroly S^* (lidský či jiný), jenž by na systém S (SGR-1A) během jeho činnosti jakkoli působil. Tato autonomie však závisí na prostředí: pokud tyto stroje budou nasazeny na střežení hranic, kam není nikomu dovolený přístup, není identifikace cílů obtížná; každý objekt určený jako člověk může být zneškodněný. V jiném prostředí by však plná autonomie možná nebyla; systém by jednoduše nedokázal dobře rozlišovat mezi legitimními a nelegitimními cíli. Byla by proto nezbytná existence systému S^* (lidské kontroly), který by v nějaké fázi operace systému SGR-1A intervenoval a rozhodoval o tom, zda smí či nesmí provést určitou činnost. Plně autonomní vojenští roboti nevylučují dohled člověka, ten však nemusí být strategicky vhodný a zajišťující dostatečně rychlou reakci na měnící se podmínky na bojišti.

3 Důvody pro zavádění vojenských robotů

Existuje celá řada dobrých důvodů, proč by vojenští roboti měli postupně nahrazovat lidské kombatanty (Arkin, 2009; Galliot, 2015) První a zřejmě nejevidentnější spočívá v tom, že když namísto lidí budou bojovat roboti, budou války méně krvavé; ušetříme mnoho životů a zdraví. I když je vedení moderních válek stále sofistikovanější, vojáci zůstávají těmi, kdo riskují své životy a zdraví na bitevních polích. Nehrozí jim smrt pouze z přímé konfrontace s živým nepřítelem. Bombardování, využívání dronů, miny, útoky raketami apod. umožňují efektivní zabíjení a mrzačení vojáků. Čím bude vojáků na bitevním poli méně, tím více lidských životů může být ušetřeno. A to je nepochybně důležitým pozitivním faktorem.

Životy a zdraví vojáků však nemusí být ušetřeny pouze tím, že se jejich přímé zapojení do válečné vřavy bude snižovat. Bojoví roboti by měli být schopni reagovat bez emocí a zasahovat mnohem přesněji než lidé. Vzhledem k tomu, že usmrcení nepřátel není nutným cílem k naplnění cílů spravedlivé války, mohli by roboti namísto zabíjení zraňovat. Zřejmě by se muselo jednat o vážnější zranění, aby vojáci skutečně nepokračovali v boji a minimalizovala se šance, že se do války zapojí po svém uzdravení, zranění je stále lepší než smrt.

Vojáci nejsou vystaveni pouze riziku smrti a fyzického zranění. Mnozí si odnášejí psychické problémy, obtížně se vracejí do starého života, nemohou spát, mají těžké sny a trpí postraumatickou stresovou poruchou. Nelze se tomu divit, válka je děsivá zkušenost, a i když vojáci přežijí ve zdraví, jen těžko se vyrovnávají s tím, že museli zabíjet, že jejich přátelé a spolubojovníci umírali, často děsivými a bolestivými způsoby. Nemálo vojáků spáchá po návratu do své vlasti sebevraždu. Veteráni válek často potřebují odbornou pomoc a někteří autoři odhadují, že kdyby jim byla skutečně poskytnuta, finančně by to ohrozilo systémy veřejného zdravotnictví. Nasazování vojenských robotů by mohlo omezit tyto negativní důsledky a tím i ve válce vzniklou, nicméně mnohá léta po válce existující újmu.

Využívání vojenských robotů by však mohlo přispět ještě jinak k minimalizaci ztrát na životech a újmě na zdraví nejen kombatantů, ale i civilistů. Vojáci pocházejí z různých společenských skupin, liší se svou výchovou, vzděláním, ale také motivací, proč slouží v armádě (je-li služba vojenská služba dobrovolná). Mohou mít rozdílné názory na morálku a jejich pohled na nepřátelské kombatanty a civilisty může být zatížený předsudky, rasismem či jinými negativními postoji. Válka je navíc velmi stresující událost, která může výraznou měrou ovlivňovat jejich úsudek a jednání. Citujeme zprávu Surgeon General Office (USA), která zkoumala duševní zdraví a etiku na bojišti u vojáků účastnících se vojenské operace v Iráku. Výzkum zjistil, že (Galliot, 2015)

1. 10 % vojáků a příslušníků americké Námořní pěchoty uvedlo, že se nekorektně chovalo k civilistům a jejich majetku (bití, kopání, ničení obydlí apod.).

2. Pouze 47 % vojáků a 38 % příslušníků americké Námořní pěchoty souhlasilo s tím, že k civilistům je třeba se chovat s úctou a respektem.

3. Více než třetina vojáků a příslušníků americké Námořní pěchoty souhlasila s mučením, pokud by mohlo zachránit životy či umožnilo získat důležité informace.

4. 17 % vojáků a příslušníků americké Námořní

pěchoty se domnívalo, že s civilisty by se mělo jednat jako s povstalci.

5. Třetina příslušníků americké Námořní pěchoty a více než čtvrtina vojáků se domnívá že jim jejich velitelé nedali jasný pokyn, aby se k civilistům chovali korektně.

6. Třebaže všichni vojáci absolvovali kurzy etiky, 28 % vojáků a 31 % příslušníků americké Námořní pěchoty uvedlo, že se ocitli v situaci, kdy nevěděli, jakým způsobem ji eticky řešit.

Tyto výsledky jsou alarmující, byť nejsou úplně překvapující. Smyslem etiky spravedlivé války je minimalizovat újmu na lidských životech a zdraví. To samozřejmě předpokládá, že jsou její principy skutečně aplikovány v praxi. Ze strany armádního velení tedy musíme očekávat, že jejich strategické plány a konkrétní rozkazy vždy berou principy spravedlivé války v potaz, zatímco u vojáků požadujeme, aby se jimi ve svém činnosti řídili. K tomu je však třeba, aby byli dostatečně trénováni, aby zvládali své emoce, aby dokázali poněkud abstraktní principy aplikovat v konkrétních situacích, aby se nenechávali strhávat svými předsudky, negativními postoji, žalem, strachem a dalšími faktory ovlivňujícími jejich rozhodování a konání. Je pochopitelné, nicméně z hlediska spravedlivé války nepřípustné, pokud se jim to nedaří dokonale. Důvodů tohoto nezdaru je hned několik:

1. Ztratili ve válce přátelé a ovládla je touha po pomstě.

2. Dehumanizují si své nepřátele, například tím, že je soustavně označují urážlivými termíny.

3. Jejich trénink nebyl dostatečný a/či nemají dostatek bojových zkušeností.

4. Propadají depresím a rostoucímu pocitu frustrace.

5. Mají potěšení z pocitu nadvlády a ze zabíjení.

Lze předpokládat, že v případě využívání vojenských robotů se podaří tyto negativní faktory ovlivňující jednání vojáků odstranit. Roboti nebudou mít emoce, nebudou se chtít mstít, nebudou plakat nad ztrátou svých „druhů“, nebudou dehumanizovat nepřítele, budou důsledně rozlišovat mezi kombatanty a civilisty, mezi legitimními cíli a těmi nelegitimními, neovládne je deprese či frustrace, nebudou mít potěšení ze zabíjení. V důsledku toho bude válka humánnější v tom smyslu, že ubudou zbytečné útoky a zabíjení, nebude docházet k zabíjení a mučení civilistů, ženy nebudou znásilňovány, nikdo nebude ničit majetek civilistů, zranění vojáci budou v bezpečí před případnou potřebou pomsty či vybitím si frustrace. To vše představuje další důležitý pozitivní faktor.

Mezi další pozitivní faktory se uvádí nižší finanční náročnost vojenských robotů v porovnání s vojenskými stroji s posádkou na palubě či nižší dopady na životní prostředí (stroje bez posádky nemusí mít tolik bezpečnostních prvků a konzumují méně pohonných hmot). Za nejdůležitější faktory hovořící ve prospěch využívání vojenských robotů, včetně AVR, však považuji ty spojené s minimalizací újmy, ať již na životech, zdraví a majetku, a umožňující důslednější aplikaci principů *ius in bello*.

4 Důvody proti zavádění vojenských robotů

Kritici zavádění a využívání vojenských robotů ve válečných konfliktech nejčastěji argumentují tak, že domnělé výhody se velmi snadno mohou změnit v nevýhody; z pozitivních faktorů na negativní. Kromě toho se mohou objevit i další problémy, které hovoří v neprospěch těchto moderních vojenských technologií. Uvedu zde jen některé z nich.

Moderní umělá inteligence, zvláště rozpoznávání obrazu ze senzorů, řeči a obecně nějakých vzorů a vztahů mezi nimi, využívá neuronové sítě a celou řadu algoritmů. Tyto sítě se dokáží naučit činnost, kterou od nich člověk očekává, kvalita jejich výkonu však závisí na množství a variabilitě dat. Jestliže se například neuronové sítě autonomních aut budou učit rozpoznávat lidi na vzorku, který bude obsahovat málo zástupců jiné barvy pleti, může mít později problémy rozpoznat černocha či Hispánce jako člověka, což samozřejmě povede k diskriminačnímu jednání. Podobné problémy se mohou objevit i u vojenských robotů, zvláště jejich plně autonomní varianty. Jedním z klíčových požadavků na vojenské roboty musí být, aby dokázali bezchybně rozlišovat legitimní a nelegitimní cíle, tedy minimálně kombatanty své armády od těch nepřátelské strany, a samozřejmě civilisty od kombatantů. Opomenutí v tréninkové sadě dat může vést k tomu, že se vojenští roboti budou cíle plést; chybně identifikují civilistu jako kombatanta či vlastního vojáka jako nepřátelského. Tyto chyby mohou mít vážné dopady a vojenští roboti, místo aby přispívali k naplňování podmínek i*us in bello*, mohou přispívat k nespravedlnosti a zvyšovat utrpení.

Další problémy s využíváním vojenských robotů spočívají v tom, že hluboké neuronové sítě jsou náchylné k tzv. *adversarial attack*. Některé studie ukazují, že na obrázku postačí změnit pouhý jeden pixel a neuronová síť bude psa klasifikovat jako kočku. Kdyby se nepřátelskému vojsku podařilo proniknout do kamerových systémů robotů a pozměnit percepty (vizuální obrazy přicházející z kamer ke zpracování), mohlo by to mít nedozírné následky. Roboti by mohli zcela chybně vyhodnocovat percepty pocházející z kamer a vést útoky na vlastní vojáky či dokonce civilisty.

Útok na vojenské roboty by mohl být globálnější

a pokusit se je ovládnout zcela (hacknout). Ti by se obrátili proti vlastním vojákům a způsobili by nemalé škody; nacházeli by se mezi „svými“, blízko potenciálním cílům, navíc by mohli útočit zcela nečekaně a zastihnout své oběti zcela nepřipravené. Teroristické organizace či některé státy by dokonce mohly využít hacknuté bojové roboty k útokům na civilisty, vznášet mezi ně zmatek, paniku a podkopat tak bojovou morálku celého národa a ochotu se i nadále angažovat ve spravedlivém válečném konfliktu.

Nepodceňujeme tyto výhrady proti využívání vojenských robotů, zvláště AVR, nezdá se nám však, že převažují nad pozitivními faktory. Důvodů je několik, hlavní však spočívá v tom, že se vlastně nejedná o negativní faktory jako spíše o rizika. Rizika jsou spojená s každou moderní technologií. Například u autonomních vozidel hrozí, že se do jejich systému někdo naboural a pokusí se je využít jako prostředek zabíjení. Pozitiva autonomní dopravy však převažují negativa + rizika. Kromě toho, identifikace rizik je důležitým nástrojem jejich předcházení. Víme-li, že nedostatečné sady trénovacích dat mohou vést k chybné identifikaci legitimních cílů, nemusíme to chápat jako důvod zákazu autonomních vojenských robotů, ale jako podmínku, která stanoví způsob trénování neuronových sítí a klade podmínky na jejich využívání v praxi.

Jednou z důležitých námitek proti vývoji AVR je riziko, že povede k novým závodům ve zbrojení (Asaro, 2019).[1] Současně by mohlo být snadnější válečné konflikty zahájit (Sharkey, 2008). Pokud ve válkách nebudou umírat lidé, nebo jich bude umírat výrazně méně, mohl by poklesnout případný odpor veřejnosti vůči válkám a tlak na vlády, aby se jim snažily vyhnout a využívaly jiné prostředky řešení mezinárodních sporů. Tato námitka je v podstatě konsekvencialistická a snaží se upozornit na riziko, že vyvíjení AVR povede k neblahým důsledkům. Těmi může být buď to, že posílí závody ve zbrojení a poroste napětí mezi státy, nebo rostoucí počet válečných konfliktů. I kdyby potom platilo, že AVR mají potenciál ušetřit lidské životy, tento pozitivní faktor by byl převážený počtem těchto konfliktů. V konečném důsledku by mohlo umírat stejné, nebo dokonce i větší množství lidí.

Připouštíme, že tato námitka je poměrně vážná. Pokud by skutečně vývoj a využívání AVR vedly k masivním závodům ve zbrojení a jisté řekněme lehkovážnosti v zahajování válečných konfliktů, potom by negativní důsledky využívání této moderní technologie skutečně mohly převážit pozitivní faktory. Bylo by však nerealistické předpokládat, že tyto obavy zastaví vývoj a posléze využívání AVR ve válečných konfliktech (Scholz, 2016). Některé země, jako je např. Rusko, vyvíjí AVR zcela otevřeně a všem státům je jasné, že zastavit jejich vývoj by v situaci nedůvěry a soupeření mezi státy nebylo moudré.

[1] Závody ve zbrojení zde rozumím soupeření mezi dvěma či více státy ve vývoji vojenských technologií. Cílem závodu ve zbrojení je dosáhnout kvalitativní (typy a kvalita zbraní) či kvantitativní (množství zbraní) výhody oproti soupeřícím státům.

5 Závěr

Je zjevné, že existují poměrně dobré (pro někoho přesvědčivé) důvody k využívání AVR v moderním válčení. Neznamená to, že s touto technologií nejsou spojena vážná rizika, ale jak jsme již uvedli, rizika nemusíme chápat jako překážky, ale spíše jako vodítka, která dokáží nasměřovat naši pozornost správným směrem a umožnit nám přijmout opatření k jejich minimalizaci.

Poděkování

Tento příspěvek vznikl za podpory programu Strategie AV21 „Filozofie a umělá inteligence“.

Literatúra

Arkin, R. (2009). *Governing Lethal Behavior in Autonomous Robots*. CRC Press, Boca Raton.

Asaro, P. (2019). What Is an Artificial Intelligence Arms Race Anyway? *A Journal of Law and Policy for the Information Society*, (15):45–64.

Galliot, J. (2015). *Military Robots. Mapping the Moral Landscape*. Routledge, London.

Lucas, G. (2023). *Law, Ethics and Emerging Technologies. Confronting Military Technologies*. Routledge, New York.

Mayor, A. (2018). *Gods and Robots. Myths, Machines, and Ancient Dreams of Technology*. Princeton University Press, Princeton.

Nourbakhsh, I. R. (2013). *Robot Futures*. The MIT Press, Cambridge, Mass.

Scholz, J. a kol. (2016). Ethical Weapons. A Case for AI in Weapons. V *Moral Responsibility in Twenty-First Century Warfare. Just War Theory and the Ethical Challenges of Autonomous Weapons Systems*, str. 181–213. SUNY Press, Albany.

Sharkey, N. (2008). Cassandra or the False Prophet of Doom: AI Robots and War. *IEEE Intelligent Systems*}, (23):14–17.

Tamburini, G. (2016). On Banning Autonomous Weapons Systems. V *Autonomous Weapons Systems. Law, Ethics, Policy*, str. 122–141. Cambridge University Press, Cambridge.

Dizajn nového učiaceho pravidla pre moderné Hopfieldove siete

Matej Fandl, Martin Takáč

Centrum pre kognitívnu vedu FMFI
Univerzita Komenského v Bratislave
{matej.fandl,martin.takac}@fmph.uniba.sk

Abstrakt

Asociatívna pamäť je vo výpočtových systémoch využiteľná na rekonštrukciu nekompletných, zašumených vzorov. Moderné Hopfieldove siete sú jej neurálny model, ktorý má veľkú kapacitu a je použiteľný samostatne, ale aj ako súčasť architektúr hlbokého učenia. V našej práci skúmame nové spôsoby učenia týchto sietí. Tento príspevok je vhľadom do prebiehajúcej exploratívnej fázy tvorby nového učiaceho pravidla s ambíciou kombinovať efektívnosť učenia s efektívnosťou samotnej rekonštrukcie.

1 Úvod

Moderné Hopfieldove siete (Krotov a Hopfield, 2016; Ramsauer a spol., 2020) sú trieda modelov asociatívnej pamäte s vysokou kapacitou. Vidíme príležitosť pre efektívne trénovanie týchto sietí tvorbou receptívnych polí neurónov na skrytej vrstve tak, aby využili pravidelnosti v trénovacej množine a rozdelili si prácu. Biologicky inšpirované metódy učenia sa receptívnych polí bez učiteľa vzorov opísali napríklad Krotov a Hopfield (2019), alebo Ravichandran a spol. (2021), nie však pre potreby rekonštrukcie vzorov, ktoré sú oproti klasifikácii špecifické v tom, že ich kombináciou musia vznikať kompletné vzory. Náš výskum sa zameriava na tvorbu nového učiaceho pravidla pre spojité moderné Hopfieldove siete (SMHS) vhodného pre rekonštrukciu.

2 Model

Neuróny v SMHS majú spontánnu aktivačnú dynamiku určenú pravidlom

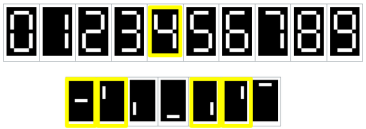
$$\boldsymbol{\xi}^{t+1} = \boldsymbol{W} softmax(\beta \boldsymbol{W}^T \boldsymbol{\xi}^t), \qquad (1)$$

kde $\boldsymbol{\xi}$ je vektor stavov neurónov o dĺžke rovnej dimenzionalite vstupného priestoru d, $\boldsymbol{W}$ je matica s rozmermi $d \times N$, ktorej N stĺpcov predstavuje jednotlivé zapamätané reprezentácie a β je inverzná teplota.

SMHS je s týmto rekonštrukčným pravidlom interpretovateľná ako sieť s jednou skrytou vrstvou (Krotov a Hopfield, 2020). Rekonštruované vzory sú afinnou kombináciou váhových vektorov (receptívnych polí) neurónov skrytej vrstvy.

3 Trénovanie

Potrebu nového spôsobu trénovania spojitých moderných Hopfieldových sietí sme popísali v našej predchádzajúcej práci (Fandl a Takáč, 2022). V nej sme dosiahli deľbu práce neurónov na skrytej vrstve, ale nie požadované distribuované reprezentácie. V tejto sekcii predstavíme nový prístup, ktorým ich chceme dosiahnuť.



Obr. 1: Dátová množina digitálnych číslic a predstava o možnej podobe distribuovaných reprezentácií. Zdroj obrázka: (Fandl a Takáč, 2022)

3.1 Doúčanie sa chýbajúcich častí

Nové pravidlo je postavené na princípe doúčania sa chýbajúcich častí v rekonštrukcii práve predkladaného vzoru. Doúčanie sa je navrhnuté vzhľadom na aktivačnú dynamiku v rovnici (1) a jeho fungovanie opisuje pseudokód jednej epochy.

for all $vzor \leftarrow vzory$ **do**
  $rez = kopiruj(vzor)$
  while $\|rez\| > \epsilon$ **or** $!max\_iteracii$ **do**
    $vaha = najblizsia\_k\_rez$
    $vaha = \frac{vaha + \alpha \cdot rez}{\|vaha + \alpha \cdot rez\|}$
    $rez = rez - p_{vaha} \cdot vaha$
  end while
end for

V algoritme je ϵ prahová hodnota zanedbateľnosti rezídua, α je rýchlosť učenia a p_{vaha} je miera pravdepodobnosti, do akej receptívne pole reprezentuje rezíduum. Táto hodnota je získaná ako príslušný komponent vektora $\boldsymbol{p} = softmax(\beta \boldsymbol{W}^T \boldsymbol{rez})$, kde $\boldsymbol{W}$ je matica váh SMHS. Vzorec pre výpočet $\boldsymbol{p}$ je adaptovaný z rovnice (1).

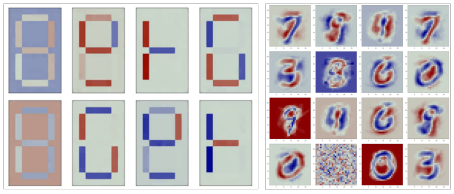
4 Predbežné výsledky

Prezentované výsledky exploratívnej fázy sú z experimentov na jednoduchej dátovej množine 13 $\times$ 8 px

digitálnych číslic (DČ) a na dátovej množine MNIST (Deng, 2012).

4.1 Receptívne polia

Podobne ako pri predchádzajúcom prístupe (Fandl a Takáč, 2022) sa nám darí dosiahnuť deľbu práce medzi neurónmi, avšak tentoraz aj s distribuovanými reprezentáciami. Pre obe množiny sú výsledné receptívne polia vidieť na Obr. 2.



Obr. 2: Receptívne polia pri trénovaní na dátových množinách DČ a MNIST. Červená - negatívne hodnoty, modrá - pozitívne hodnoty, biela - hodnoty blízke nule.

V oboch prípadoch sa tvorí „priemerný neurón“, ktorého váhový vektor je blízko ku všetkým vzorom, a páry „opačných neurónov“, ktoré sa vzájomne vylučujú.

Tieto páry sú spôsobené osciláciami zapríčinenými odpočítavaním normalizovaných váhových vektorov od nenormalizovaného rezídua pri povolenom opakovanom výbere neurónov. Biele časti receptívnych polí slúžia vďaka použitej metrike – kosínusovej podobnosti – ako maska. Väčšina neurónov teda slúži ako detektory príznakov.

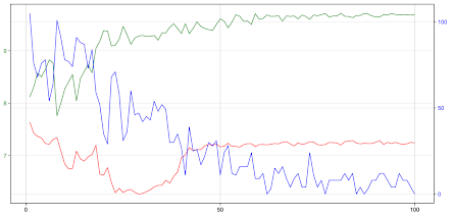
4.2 Kvalita rekonštrukcie

Meriame rozdiel medzi vstupom a rekonštrukciou kompletných, nezašumených vzorov. Konkrétne kumulatívnu kosínusovú podobnosť, euklidovskú vzdialenosť a počet komponentov s odlišným znamienkom. Pre dátovú množinu DČ sa nám podarilo dosiahnuť nulovú chybu pri poslednej metrike, avšak spojitá rekonštrukcia je neuspokojivá. Pre MNIST sa nám zatiaľ nepodarilo dosiahnuť uspokojivé výsledky v žiadnej zo spomínaných metrík.

Na Obr. 3 je vidieť úspešný beh. Za pozornosť stojí zhoršovanie rekonštrukcie po úspešnej 60. epoche.

5 Záver

Predbežné výsledky naznačujú, že pokračovať v skúmaní pravidla má zmysel, ale aj poukazujú na nedostatky a otvorené otázky. Dosahujeme deľbu práce medzi neurónmi aj distribuované reprezentácie. Lepšie porozumenie vplyvu parametrov a stratégie selekcie váhových vektorov na dynamiku učiaceho pravidla by mohlo viesť k jeho potrebným úpravám pre zlepšenie



Obr. 3: Vývoj rekonštrukčnej chyby v epochách trénovania na množine digitálnych číslic. Zelená - kosínová podobnosť, červená - euklidovská vzdialenosť, modrá - počet komponentov s odlišným znamienkom.

kvality rekonštrukcie. Receptívne polia formované trénovaním na dátovej množine MNIST pôsobia podobne ako tie, ktoré dosiahli Krotov a Hopfield (2019). Jedným z plánov je otestovať vhodnosť tých našich aj pri klasifikácii.

Poďakovanie

Tento príspevok vznikol s podporou grantu VEGA 1/0373/23 a KEGA 022UK-4/2023.

Literatúra

Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142.

Fandl, M. a Takáč, M. (2022). Zvyšovanie efektivity trénovania a kapacity v atraktorovom neurálnom modeli asociatívnej pamäte. V *Kognice a umělý život XX*.

Krotov, D. a Hopfield, J. (2020). Large associative memory problem in neurobiology and machine learning. *arXiv preprint arXiv:2008.06996*.

Krotov, D. a Hopfield, J. J. (2016). Dense associative memory for pattern recognition. *Advances in Neural Information Processing Systems 29 (2016), 1172-1180*.

Krotov, D. a Hopfield, J. J. (2019). Unsupervised learning by competing hidden units. *Proceedings of the National Academy of Sciences of the United States of America*, 116(16):7723–7731.

Ramsauer, H. a spol. (2020). Hopfield networks is all you need. *arXiv preprint arXiv:2008.02217*.

Ravichandran, N. B., Lansner, A. a Herman, P. (2021). Brain-like approaches to unsupervised learning of hidden representations – a comparative study. Farkaš, I. a spol. (zost.), V *Artificial Neural Networks and Machine Learning - ICANN 2021*, str. 162–173, Cham. Springer International Publishing.

Dôveryhodnosť výpočtových modelov v umelej inteligencii a robotike*

Igor Farkaš

Fakulta matematiky, fyziky a informatiky
Univerzita Komenského v Bratislave
igor.farkas@fmph.uniba.sk

Abstrakt

S búrlivým rozvojom umelej inteligencie v rôznych oblastiach súvisí aj zvyšovanie kritérií na navrhované a využívané riešenia. Dôveryhodnosť ako pojem prevzatý z psychológie rezonuje v súčasnej umelej inteligencii a robotike. Týka sa nielen programov implementovaných v počítači, ktoré produkujú požadované výstupy, napr. na displeji, ale aj programov ovládajúcich robotické systémy, ktoré priamo zasahujú do prostredia alebo interagujú s človekom. Požiadavka dôveryhodnosti sa javí samozrejmá z pohľadu človeka, ktorý sa potrebuje na umelý systém spoľahnúť. Zložitejšou otázkou je, čo tento pojem znamená v matematickom zmysle a ako dôveryhodnosť zabezpečiť a merať. Na tieto aspekty poukazujeme v príspevku, kde spomenieme dôležité koncepty, v kontexte súvisiacich pojmov, a so zvláštnym dôrazom na humanoidné roboty interagujúce s človekom.

1 Úvod

Umelá inteligencia (UI) sa teší v ostatnej dekáde vysokej popularite v rôznych aplikačných oblastiach. Úspešné modely UI typicky fungujú vďaka strojovému učeniu založenému na hlbokých neurónových sieťach, ktoré umožnili nachádzať úspešné riešenia rozmanitých úloh ako spracovanie obrazových dát (klasifikácia do tried), úlohy v prirodzenom jazyku či sekvenčné rozhodovacie úlohy (napr. hranie hier proti ľudským súperom) (Schmidhuber, 2015).

Napriek svojim úspechom majú mnohé súčasné UI systemy aj rôzne nedostatky, hlavne to, že sú zraniteľné voči nepostrehnuteľným, tzv. adverzariálnym útokom, nedávajú objektívne výsledky (sú zaujaté voči nedostatočne zastúpeným triedam dát v klasifikačných úlohách), no najmä, sú málo zrozumiteľné, a to nielen pre bežných ľudí ale aj odborníkov. Tieto nedostatky zhoršujú používateľskú skúsenosť a narúšajú dôveru ľudí vo všetky systémy UI. K negatívnemu povedomiu o UI prispieva aj zábavný priemysel, ktorý už desiatky rokov chrlí filmy s emocionálne nabitou tematikou, že svet ovládne umelá inteligencia, najčastejšie v stelesnená v neohroziteľných humanoidných robotoch alebo iných ničivých agresoroch.

*podporené projektmi VEGA 1/0373/23 a APVV-21-0105.

2 Dôveryhodnosť systémov UI

Je mimoriadne dôležité podnikať účinné kroky opačným smerom a zvyšovať objektívnu informovanosť v spoločnosti. Budovanie dôveryhodných systémov UI je zložitý proces, ktorý bude musieť podchytiť rôzne súvisiace aspekty ako sú robustnosť, dostatočná generalizácia, vysvetliteľnosť, transparentnosť, reprodukovateľnosť a ďalšie. Li a spol. (2023) ponúkajú zjednocujúci pohľad na v súčasnosti dostupné, ale roztrieštené prístupy k dôveryhodnej UI, a organizujú ich systematicky. Identifikujú aj kľúčové príležitosti a výzvy pre budúci vývoj dôveryhodných systémov UI.

Koncept dôveryhodnosti je človeku intuitívne zrejmý, nakoľko ju posudzujeme u ľudí podľa ich vonkajšieho správania a konania, hoci berieme do úvahy aj názory iných ľudí. Najspoľahlivejšie však dospejeme k názoru na základe vlastných skúseností. Limitáciou tohto prístupu je, že človeku do jeho mysle nevidíme, preto nás môže niekedy jeho správanie prekvapiť alebo človek oklamať.

V prípade požadovania dôveryhodnosti u systémov UI platia podobné princípy, avšak sú tu aj rozdiely: (1) vlastnosti systému UI môžeme ovplyvniť pri jeho dizajne, a (2) do vnútra systému môžeme nahliadnuť, keďže ide o výpočtový model.

Podľa existujúcej literatúry, dôveryhodný systém musí spĺňať viacero vlastností. Nezávislá expertná skupina na vysokej úrovni pre umelú inteligenciu, zriadená Európskou komisiou (EK) v roku 2018 vypracovala Etické usmernenia pre dôveryhodnú UI, ktorá by mala byť:

1. **zákonná**, čiže by sa mala riadiť celým platným právom a právnymi predpismi;
2. **etická**, čiže by mala zabezpečiť súlad s etickými zásadami a hodnotami, a
3. **odolná**, aby nebola zraniteľná voči zneužitiu, a bola bezpečná.

Každá zložka je sama o sebe potrebná, ale nestačí na dosiahnutie dôveryhodnej UI. V ideálnom prípade pôsobia všetky tri zložky vo vzájomnom súlade a prekrývajú sa vo svojom fungovaní. Ak medzi týmito zložkami v praxi vzniknú rozpory, spoločnosť by sa mala pokúsiť o ich zosúladenie.

Zákonný aspekt sa zdá byť zrejmý, spomenieme len, že EK navrhla štyri stupne regulácie UI (podľa miery rizika; tzv. AI Act), podľa ktorých sa bude musieť nasadzovanie jednotlivých systémov UI riadiť. Samozrejme možno očakávať pribúdanie zákonov týkajúcich sa UI.

Etický aspekt zahŕňa viacero ďalších požiadaviek, ako sú transparentnosť, vysvetliteľnosť, bezpečnosť a férovosť. Tieto pojmy už stihli získať v literatúre dosť pozornosti, najviac asi požiadavka vysvetliteľnosti (Barredo a spol., 2020), ktorá vnáša dôveru v systém, ak vieme vysvetliť jeho správanie. V prípade hlbokých neurónových sietí sú vysvetlenia v princípe zložité, a úroveň vysvetlenia závisí aj od toho, komu je určené (užívateľovi, pacientovi, expertovi). Expert rozumie aj matematickým vzorcom, zatiaľ čo bežný človek uprednostní vysvetlenie v prirodzenom jazyku alebo na obrázkoch.

Odolnosť (robustnosť) neurónových sietí je relatívne nový problém, ktorý bol zistený pred necelými 10-imi rokmi. (Szegedy a spol., 2014) zistili, že minimálna cielená perturbácia vstupov (teda útok) prezentovaných dobre natrénovanej a fungujúcej neurónovej sieti spôsobí jej spoľahlivé oklamanie (čiže sieť si bude "myslieť", že vidí niečo úplne iné). Jedným zo smerov, ktoré by mohli pomôcť riešiť tento problém je, že zavedieme do modelu mechanizmus pozornosti (Farkaš a spol., 2022). Existujúce modifikácie učenia pomocou tzv. adverzariálneho tréningu trochu pomáhajú liečiť symptómy, že takto zdokonalené modely vedia odolať aspoň niektorým útokom, nie sú však (aspoň zatiaľ) principiálnym riešením (Zhao a spol., 2022).

3 Dôveryhodnosť UI v robotike

Dôveryhodnosť UI implementovanej vo fyzických robotoch je veľmi dôležitá, pretože narastá význam úlohy robotov, nielen humanoidných, ktoré budú interagovať s človekom, napr. v priemysle, v domácnosti a pod. (Kok a Soh, 2020). V tomto kontexte ešte viac akcentuje aspekt bezpečnosti človeka (vyjadrený už prvým Asimovým zákonom robotiky), preto sa častejšie začína využívať pri vývoji systémov virtuálna realita na interakciu robota a človeka (Dianatfar a spol., 2021), a na počítačové simulácie robotov.

Výskum v tejto oblasti je v počiatkoch. Jedným zo smerov je požadovaná transparentnosť robota, ktorú môžeme navonok posudzovať podľa jeho správania, keďže každý robot koná vo svojom prostredí. Človek sa prirodzene snaží predikovať správanie iných ľudí, za čím sú (vedomé i nevedomé) mechanizmy prebiehajúce na viacerých úrovniach súčasne (Bubic a spol., 2010). Na motorickej úrovni sa človek snaží predikovať napríklad pohyb ruky alebo tela, na mentálnej úrovni zase úmysel druhého človeka (využívajúc teóriu mysle). Pri skúmaní UI v robotoch máme výhodu, že máme prístup k obom úrovniam - viditeľnej (motorickej) i skrytej (mentálnej), pretože môžeme analyzovať bežiace algoritmy UI.

V prípade motorickej úrovne, ktorá často prezrádza zámery človeka, je jedným zo zaujímavých konceptov čitateľnosť (legibility) pohybu, čo znamená ako rýchlo vieme správne predpovedať cieľ pohybu pozorovaním počiatočnej trajektórie, napríklad robotického ramena. K tomu sa pripája sprevádzajúci pohľad očí, ktorý napomáha predikcii cieľa motorického pohybu. V snahe o bezproblémovú a hladkú interakciu medzi človekom a robotom je dôležité vybaviť roboty náležitými perceptuálnymi vlastnosťami ako schopnosť sledovať a predikovať pohyby ruky človeka a jeho pohľadu. Je známe, že antropomorfné vlastnosti robota napomáhajú lepšej interakcii medzi človekom a robotom. Všetky tieto procesy možno chápať ako výpočtové na oboch stranách (Thomaz a spol., 2013) a ich analýza, modelovanie a testovanie sú otázkou operacionalizácie a behaviorálnych experimentov.

Literatúra

Barredo, A. a spol. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82–115.

Bubic, A., von Cramon, D. Y. a Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4.

Dianatfar, M., Latokartano, J. a Lanz, M. (2021). Review on existing VR/AR solutions in human–robot collaboration. *Procedia CIRP*, 97:407–411.

Farkaš, I., Cimrová, B., Štefan Pócoš a Bečková, I. (2022). Pozornosť ako biologicky inšpirovaný koncept pre vysvetliteľné, robustné a efektívne strojové učenie. V *Kognícia a umelý život XXI*, str. 34–38.

Kok, B. a Soh, H. (2020). Trust in robots: Challenges and opportunities. *Current Robotics Rep.*, 1:297–309.

Li, B. a spol. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, 55(9):1–46.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117.

Szegedy, C. a spol. (2014). Intriguing properties of neural networks. V *International Conference on Learning Representations*.

Thomaz, A., Hoffman, G. a Cakmak, M. (2013). Computational human-robot interaction. *New Foundations and Trends*, 4(2-3):105–223.

Zhao, W., Alwidian, S. a Mahmoud, Q. H. (2022). Adversarial training methods for deep learning: A systematic review. *Algorithms*, 15(8).

Viacvrstvové neurónové siete s učiacimi sa produktovými neurónmi

Slavomír Holenda, Kristína Malinovská a Ľudovít Malinovský

Katedra aplikovanej informatiky, FMFI,
Univerzita Komenského v Bratislave
Mlynská dolina, 84248 Bratislava
Email: {holenda1,malinovska}@fmph.uniba.sk

Abstrakt

Klasické neurónové siete dosahujú obmedzené výsledky pri probléme parity alebo dvoch špirál pri malom počte neurónov. V tomto článku prezentujeme experimenty s nami navrhnutým konceptom nového neurónu, ktorý používa násobenie namiesto sumácie. Špecifický je tým, že je možné ho učiť a zároveň ho netreba ďalej obmedzovať alebo stabilizovať. Taktiež ho vieme zapojiť do architektúr s klasickými neurónmi. Testovali sme ho na probléme dvoch špirál a porovnávali jeho rôzne zapojenia na probléme parity 7. stupňa. Naše výsledky preukázali efektívnosť použitia produktových vrstiev.

1 Úvod

Napriek tomu že je násobenie v neurónových sieťach teoreticky preskúmané, v terajšom state-of-the-art výskume ho nájdeme iba zriedka. Vo väčšine prípadov sa totiž nepodarilo násobenie zovšeobecniť a prekonať problémy implementácie. V predchádzajúcom príspevku (Malinovský, Ľ. and Malinovská, K., 2022) sme práve na konferencii KUŽ prvýkrát predstavili náš model siete s produktovými neurónmi, ktoré majú trénovateľné váhy.

V histórii neurónových sietí sa už v 60-tych rokoch objavili neuróny vyššieho rádu, kde sa vstupný vektor rozšíril o vynásobené a umocnené už existujúce vstupy (Nilsson, 1965). Pri aplikovaní spätného šírenia chyby na takéto neuróny hovoríme o tzv. *sigma-pí* sieťach (Rumelhart a spol., 1986). Náš model vychádza z modelu, ktorý navrhli Ghosh a Shin (1992). Ten najprv vstupy lineárne kombinuje a až následne ich násobí. Nelineárna aktivačná funkcia sa použije až na poslednej vrstve, ktorá sa ako jediná učí pomocou gradientovej metódy. Váhy neurónov predošlých vrstiev sú nastavené na hodnotu 1 a nemenia sa. Tento model vo všeobecnosti prekonáva silu klasického perceptrónu, dokonca je porovnateľný so sieťami s neurónmi vyššieho rádu, avšak neumožňuje trénovanie váh produktových neurónov.

Potenciál násobenia v neurónových sieťach analyzuje Schmitt (2002), ktorý porovnáva rôzne typy sietí s násobením. Aj v oblasti hlbokých neurónových sietí nájdeme modely, ktoré využívajú násobenie, ako napríklad sum-product siete (Poon a Domingos, 2011; Delalleau a Bengio, 2011) alebo modely, ktoré využívajú násobenie na vybraných miestach (Zhu a spol., 2018; Diba a spol., 2017).

2 Náš model

Náš model (Malinovský, Ľ. and Malinovská, K., 2022) sme predstavili ako klasickú pi-sigma sieť (Ghosh a Shin, 1992) s dvoma vrstvami - klasickou sumačnou a následne produktovou. Vrstvu produktových neurónov, tak ako sme ju navrhli my, je však možné zapájať v sieti ľubovoľne na požadované miesta a adaptovať jej váhy pri trénovaní modelu. Ďalšou výhodou je, že náš model netreba pri a po trénovaní stabilizovať alebo obmedzovať.

Pre vstup $\boldsymbol{x}$, bez trénovateľného prahu (bias), a váhy w počítame aktiváciu neurónov ako

$$y_i = \prod_j \left(1 - \sigma(w_{ij})(1 - x_j)\right), \tag{1}$$

kde σ je logistická funkcia. Váhy našich produktových neurónov nie sú fixné ako pri pôvodnej pi-sigma sieti. Keďže na ne aplikujeme sigmoidálnu funkciu, pri výpočtoch sa ich hodnoty pohybujú v rozmedzí $\langle 0, 1\rangle$. Následne je použitá analógia produktového neurónu s tým rozdielom, že váhy vstupujú do špeciálnej aktivačnej funkcie, ktorá má skoro totožné vlastnosti s umocnením na čísla z intervalu $\langle 0, 1\rangle$. Narozdiel od umocnenia, je táto funkcia spojitá a spojito derivovateľná, takže vieme vrstvu takýchto neurónov učiť pomocou gradientových metód. Váhu produktového neurónu si vieme predstaviť ako percento dôležitosti. Určuje v akej miere sa hodnota z presynaptického neurónu použije v násobení.

Pre výstupnú vrstvu môžeme pomocou metódy gradientového zostupu pre strednú kvadratickú chybu odvodiť nasledovné pravidlo pre učenie výstupných váh:

$$\frac{\partial E}{\partial w_{ij}^{\text{out}}} = (d_i - y_i)\left(\prod_{k\neq j} 1 - \sigma(w_{ik}^{\text{out}})(1 - h_k)\right) \tag{2}$$
$$(h_j - 1)\sigma(w_{ij}^{\text{out}})(1 - \sigma(w_{ij}^{\text{out}}))$$

Ak sa vo výpočtoch nevyskytne delenie nulou, vieme výpočet nahradiť delením nasledovne:

$$\prod_{k \neq j} 1 - \sigma(w_{ik}^{\text{out}})(1 - h_l) = \frac{y_i}{1 - \sigma(w_{ij}^{\text{out}})(1 - h_j)} \quad (3)$$

3 Viacvrstvová sieť a experimenty

V tejto časti prezentujeme experimenty a výsledky s naším modelom rozšíreným na viacero skrytých vrstiev. Pre problém parity predstavujeme podrobné preskúmanie možností zapojenia pre dve skryté vrstvy. Na záver ukážeme predbežné výsledky pre problém dvoch špirál s troma skrytými vrstvami.

3.1 Problém parity

Problém parity je základný nelineárny problém, ktorý sa používa na základné overovanie nových modelov neurónových sietí. Problém spočíva v tom, že na vstupe model dostáva binárny reťazec dĺžky n a musí určiť, či je počet jednotiek párny (výstup = 0) alebo nepárny (výstup = 1). Úspešnosť modelu v probléme parity sa štandardne vyhodnocuje v zmysle konvergencie, čiže počtu sietí, ktoré skonvergujú k riešeniu. Aby sieť konvergovala musí určitý počet epoch po sebe určiť všetky vstupy na 100% správne. Tento nastaviteľný parameter budeme nazývať okno úspechu (po angl. success window). V predošlej práci (Malinovský, Ľ. and Malinovská, K., 2022) sme ukázali, že aj základný model s jednou klasickou skrytou vrstvou a jedným produktovým neurónom na výstupe, dosahuje oveľa lepšie výsledky ako klasický MLP pri parite stupňa 2 (XOR) až 7.

V prezentovanom experimente vyhodnocujeme efektívnosť zapojenia produktovej vrstvy pri parite 7. stupňa. Na testovanie použijeme model s 2 skrytými vrstvami a vyskúšame všetky kombinácie zapojenia. Ak architektúra používa produktovú vrstvu, želané výstupy budú preškálované z $\langle 0, 1 \rangle$ na $\langle -1, 1 \rangle$ kvôli väčšej efektívnosti pri násobení. Architektúry sú popísané skratkami aktivačných funkcií daných vrstiev. T značí Hyperbolický tangens, Q je naša pseudo-exponenciálna funkcia (quasi), čiže produktová vrstva, S je Sigmoida.

Pre každú architektúru sme menili počet neurónov na skrytej vrstve. Testované veľkosti skrytých vrstiev (h) boli nasledovné: [2, 2], [3, 5], [5, 3], [5, 15], [10, 15], [15, 10], [15, 15], [15, 10], [15, 5], [10, 5], [20, 20], [25, 25], [30, 30], [40, 20], [20, 40], [40, 40]. V Tab 1 sú popísané nastavenia ostatných hyperparametrov.

Výsledky experimentu môžeme vidieť v Tab 1. Pre MLP s aktivačnou funkciou tangens sa nám nepodarilo nájsť také nastavenie aby sieť konvergovala. Tak isto ani pre architektúru číslo 6. Architektúra číslo 4. je v tabuľke dvakrát. Toto zapojenie veľmi dobre konvergovalo pre viacero veľkostí skrytých vrstiev. Pre veľkosť [10,15] síce konvergovalo všetkých 20 sietí, ale

Názov parametra	Hodnota parametra
Počet inštancií siete	20
Rýchlosť učenia	0.5
Okno úspechu	5
Maximum epoch	400 / 1000

Tab. 1: Nastavenie fixných hyperparametrov v experimente parity. Počet 1000 maximálnych epoch bol použitý iba pre klasické MLP siete.

pri veľkosti [5,10] konvergujúce siete potrebovali najmenší počet epoch. Až 2 siete dokázali konvergovať len za 6 epoch, čiže pri odrátaní veľkosti okna úspechu (5), sa tieto inštancie naučili problém parity za 1 epochu.

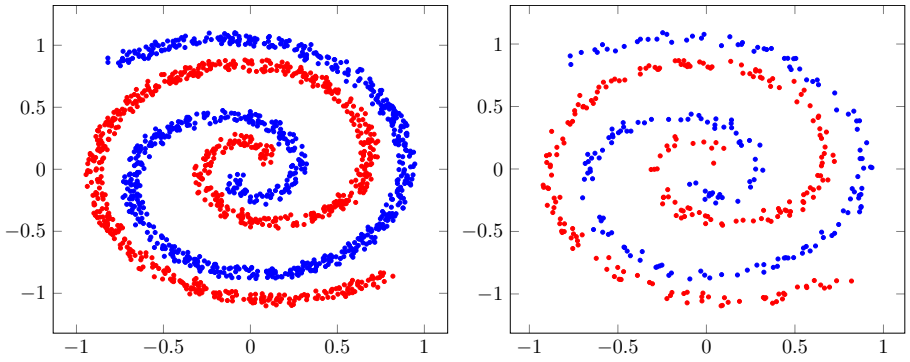
Arch.	Akt. f.	h	konverg.	epochy
1	[T, Q, T]	[15, 5]	18	100.7
2	[T, T, Q]	[20,20]	11	332.5
3	[Q, T, T]	[5,10]	14	148.5
4.1	[Q, Q, T]	[10,15]	20	82.5
4.2	[Q, Q, T]	[5,10]	19	14
5	[Q, T, Q]	[5,5]	20	22.15
6	[T, Q, Q]	[-,-]	0	-
7	[Q, Q, Q]	[2,2]	20	20.55
MLP.1	[S, S, S]	[100,100]	20	710
MLP.2	[T, T, T]	[-,-]	0	-

Tab. 2: Tabuľka popisuje koľko z 20 sietí konvergovalo pre danú architektúru, pri najlepšej nami nájdenej veľkosti skrytých vrstiev.

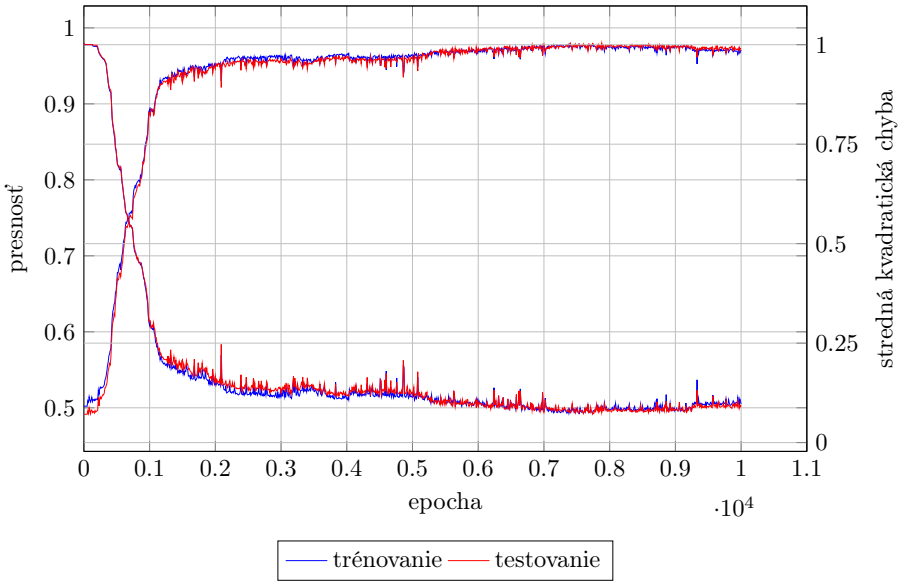
3.2 Problém 2 špirál

Nelineárny problém dvoch do seba zavinutých špirál sa taktiež používa na testovanie nových modelov. Na Obr. 1 je zobrazený dataset ktorý pozostáva z 2000 bodov. Želané výstupy boli taktiež preškálované na interval $\langle -1, 1 \rangle$.

Pre tento problém sme použili nasledujúcu architektúru: 2 vstupné neuróny, 10 tanh, 80 quasi, 5 tanh, 1 qausi výstupný neurón. Rýchlosť učenia bola nastavená na 0.01, okno úspechu bolo veľké 10 epoch a maximálny počet epoch bol 10000. V experimente sme spustili 10 sietí, z ktorých 4 konvergovali. Priemer dosiahnutej presnosti modelov bol 98.275%.



Obr. 1: Zobrazenie datasetu 2 špirál ktorý bol rozdelený v pomere 80:20 - tréning (vľavo) : test (vpravo).



Obr. 2: Priemer priebehu trénovania 10 sietí z experimentu 2 špirál.

4 Záver

V našej práci sme popísali experimenty s naším novým modelom neurónovej siete s produktovými neurónmi, v ktorých sme skúmali jeho rôzne zapojenia vo viacvrstvovej neurónovej sieti, a to na probléme parity 7. stupňa a dvoch špirál. Výsledky potvrdili, že náš model má značnú výhodu oproti klasickému viacvrstvovému perceptrónu pri problémoch takéhoto typu.

Poďakovanie

Tento príspevok vznikol v Centre pre kognitívnu vedu na KAI FMFI UK v Bratislave, s podporou grantu VEGA 1/0373/23 a a KEGA 022UK-4/2023. Za podporu tiež ďakujeme Slovenskej spoločnosti pre kognitívnu vedu SSKV[1].

[1] https://cogsci.fmph.uniba.sk/sskv/

Literatúra

Delalleau, O. a Bengio, Y. (2011). Shallow vs. deep sum-product networks. *Advances in Neural Information Processing Systems*, 24.

Diba, A., Sharma, V. a Van Gool, L. (2017). Deep temporal linear encoding networks. V *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, str. 2329–2338.

Ghosh, J. a Shin, Y. (1992). Efficient higher-order neural networks for classification and function approximation. *International Journal of Neural Systems*, 3(04):323–350.

Malinovský, Ľ. and Malinovská, K. (2022). Neurónová sieť s násobiacou vrstvou. Šejnová, G., Vavrečka a M., Hvorecký, J. (zost.), V *Kognice a umělý život XX*, str. 79–83. České vysoké učení technické v Praze.

Nilsson, N. J. (1965). *Learning machines.* McGrawHill New York.

Poon, H. a Domingos, P. (2011). Sum-product networks: A new deep architecture. V *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, str. 689–690. IEEE.

Rumelhart, D. E., Hinton, G. E. a McClelland, J. L. (1986). *A general framework for parallel distributed processing*, vol. 1, str. 26. Cambridge, MA: MIT Press.

Schmitt, M. (2002). On the complexity of computing and learning with multiplicative neural networks. *Neural Computation*, 14(2):241–301.

Zhu, J., Zeng, H., Du, Y., Lei, Z., Zheng, L. a Cai, C. (2018). Joint feature and similarity deep learning for vehicle re-identification. *IEEE Access*, 6:43724–43731.

Reweighting of binaural localization cues in virtual environment

Lucia Hucková[1] and Norbert Kopčo[1]

[1]Institute of Computer Science, P.J.Šafárik University, Šrobárová 2, 041 80 Košice, Slovakia
lucia.huckova@student.upjs.sk, norbert.kopco@upjs.sk

Abstract

The auditory system combines the binaural cues of interaural time difference (ITD) and interaural level difference (ILD) to determine the sound source location. The combining is frequency dependent. ITD dominates for low-frequency (LF) sounds and ILD for high-frequency (HF) sounds. We can experimentally measure the relative ITD/ILD weight by a localization task. A previous study (Spišák et al., 2021) showed that visually guided training on HF vs LF components in real environment induces reweighting in the binaural localization cues such that the ILD weight increases independent of the training type. We performed a follow-up experiment in real reverberant and virtual anechoic environments without training to examine the cause of the observed ILD weight increase. No reweighting was observed in this experiment, suggesting that the reweighting observed by Spišák et al. was not caused by mere exposure to different environments and that active localization is necessary to induce the binaural adaptation.

1 Introduction

Spatial auditory perception allows us to localize sounds and to separate sounds, e.g., when listening to speech in complex environments. Sound localization includes judgements about the direction (left/right and up/down) and distance of a sound source (Moore,2013)[1]. When we are presented with a sound from any position in azimuth, we can localize it due to two possible cues of the sound source an interaural time difference (ITD) and an interaural level difference (ILD). The way ITD or ILD contributes to the localization of the sound depends on the frequency content of the sound. At lower frequencies, ITDs are dominant. For higher frequencies, ILDs are dominant (Klingel et al. 2021). A previous study showed that, when listeners are trained to change the spectral weighting of components for localization in real environment, the spectral reweighting always results in an increase in the ILD weight (Spisak et al., 2021). Here, we examined whether performance of localization test in a real environment is sufficient to induce this binaural reweighting. Additionally, the study provides a control condition measurement for the Spisak et al. results.

[1] Moore (2013) stated for headphones term "lateralization" to describe the apparent location of the sound source within the head.

2 Methods

An experiment was performed, consisting of 2 parts, pretest and posttest, performed in virtual environment (VE) and real environment (RE), using methods similar to Spisak et al. (2021).

2.1 Setup and stimuli

In real environment, 11 loudspeakers were placed in a semicircle around the subject with 11.25° spacing (range ±56.25°) in a dark reverberant room. Position of the head was recorded using a headtracker. 5 types of stimuli were created using 0.5-octave noise bands at different frequencies. Low frequency stimuli were 0.35 kHz and 0.7 kHz, mid-frequency stimuli were 2.8 kHz and high frequency stimuli were 5.6 kHz and 11.2 kHz. 2 types of stimuli that were presented are: 1, 2-channel stimulus: 1 HF and 1 LF channel from locations separated by 1 or 2 speakers; 2, 4-channel stimulus: 2 HF and 2 LF channels from locations 1-2 speakers apart. Subject's task was to rotate their head towards the perceived sound location. Position of the head was guided with visual feedback. The experiment in virtual environment took place in a double-walled soundproof booth. We used ITD/ILD combination corresponding to one of 40 possible positions in horizontal plane in range from ±70.2° with spacing of 3.6°. The stimuli were 1-octave noises with center frequency 2.8 kHz. Subject's task was the same as in RE.

2.2 Experimental design

14 subjects were divided in two groups, one performed experiment in real and virtual environment (7 subjects, OR group) and one performed only virtual part (7 subjects. O group). The sequence of tests for the OR group was: day 1 (VE pretraining, VE pretest, RE pretest), day 2-4 (no training), day 5 (RE posttest, VE posttest). The sequence for the O group was identical, except that no RE tests were performed. Thus, if the

RE posttest performed immediately before the VE posttest was the main cause of the increased ILD weight in Spisak et al., then it was expected that it would be observed in the OR but not in the O group.

3 Results

Linear regression was used to estimate w_{HL} (the spectral weight of HF vs. LF components) and binaural weight w_{LT} (weight of ILD vs ITD cues). Weight of 1 means that subject oriented only according to HF/ILD component and 0 that subject oriented only according to LF/ITD component. Weights were averaged across azimuth as training effects were similar across azimuth. Fig. 1 show weights in RE for our data with no training (NoT group) and Spišák's data (LF and HF group, trained for given component). From pretest to posttest NoT weights showed no significant increase. This was expected because we omitted training. On the other hand, LF and HF group changed its spectral weighting in desired direction (results confirmed by ANOVA).

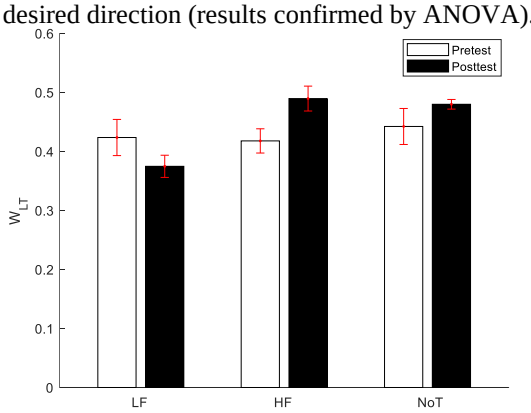


Fig. 1: Mean (±SEM) weights w_{HL} for different groups in pretest and posttest, averaged across locations.

On Figs. 2 and 3 we see results from VE. Fig. 2 is showing no significant change in weights of OR and O group, meaning that RE posttest done before VE posttest has no effect. Fig. 3 is a comparison of trained vs. not trained groups. Spišák observed the increase of ILD weight independent of the training group, the change was significant in the same direction for all groups. For our no training group, change in pretest to posttest was not significant and change in reweighting was not dependent on a group.

4 Summary and discussion

Results from RE show that change in spectral weighting does not occur, as we expected because no training of HF and LF components was present. Based on previous results we hypothesized that performing a posttest in real room immediately before the VE posttest may have caused the increased ILD weight. However, the results from virtual environment do not show the expected effect as no significant re-weighting occurred from pretest to posttest. Thus, a change in weighting cannot occur by changing from an anechoic environment to an echoic one, nor does it occur by getting used to an echoic room, but only by visual training in real environment. However, what aspect of the training is important is still unknown.

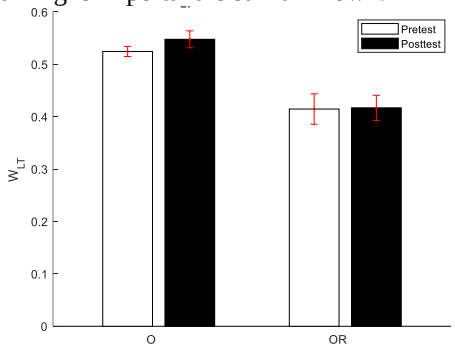


Fig. 2: Mean (±SEM) weights w_{LT} for O and OR group in pretest and posttest, averaged across locations.

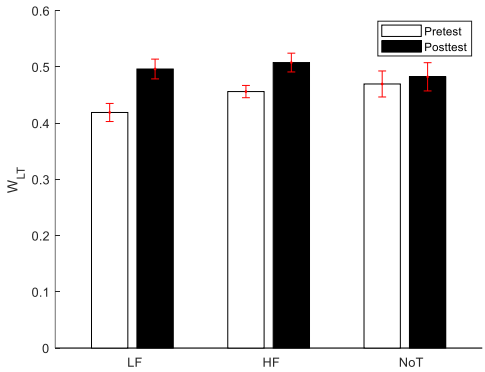


Fig. 3: Mean (±SEM) weights w_{LT} for different groups in pretest and posttest, averaged across locations.

Acknowledgment

This work was supported by VEGA 1/0350/22.

Literature

Spisak, O., Klingel, M., Lokša, P., Šebeňa, R., Laback, B., & Kopčo, N. (2021). "Reweighting the contributions of spectral regions to sound localization and its impact on binaural-cue reweighting", Proceedings of DAGA 2021, Vienna.

Klingel, M.,Kopco, N., Laback, B. (2021). Reweighting of Binaural Localization Cues Induced by Lateralization Training. Journal of the Association for Research in Otolaryngology. 22.

Moore BCJ.(2013). An Introduction to the Psychology of Hearing, Sixth Edition. Leiden, Boston, p. 245-281

Arousal a agrese: Vliv soutěživosti v násilných videohrách

Filip Kyslík[1*], Vojtěch Juřík[1], Oto Janoušek[2]

[1] Psychologický ústav, Filozofická fakulta, Masarykova univerzita, Arna Nováka 1, Brno
[2] Ústav biomedicínského inženýrství, Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně, Antonínská 548/1, Brno

[*]filip.kyslik@mail.muni.cz

Abstrakt

Téma vlivu násilí v počítačových hrách na agresivní chování je v oblasti psychologickém výzkumu již zkoumáno dlouho. V odborných studiích se však často vyskytují metodologické nedostatky, jako například volba odlišných herních titulů (omezující se např. jen na násilné/nenásilné) pro vytvoření různých experimentálních podmínek. Tyto se však často liší nejen v (ne)přítomnosti násilí, ale také v dalších aspektech, například v úrovni soutěživosti. Navzdory tomu, že soutěživost byla již zkoumána nejsou výsledky zcela konzistentní a studie neošetřují klíčové limity. Náš návrh experimentu navazuje na předchozí výzkumy, zohledňuje tato omezení a kontroluje mediující faktory. V tomto příspěvku však představujeme zejména proces vytváření jednotlivých úrovní nezávislé proměnné, protože jejich úspěšné stanovení je pro vyvozování závěrů naprosto klíčové. I přesto bývá takovému postupu v odborné literatuře věnováno relativně málo prostoru, což pak komplikuje snahu vyvozovat obecné závěry o vlivu násilných her.

1 Úvod

Jak tvrdí Adachi a Willoughbyová (2011b), násilné hry jsou většinou kompetitivní, ale ne všechny kompetitivní hry jsou přitom násilné. V roce 2017 se Americká psychologická asociace snažila shrnout výzkumy zaměřené na násilné digitální hry a agrese, dospěla však k závěru, že faktor kompetice nebyl dostatečně prozkoumán (Calvert et al., 2017). Od tohoto momentu již uplynulo určité časové období, avšak zdá se, že mnoho výzkumů na toto téma stále k dispozici není. Elosn a jeho kolegové upozorňují na to, že většina studií zaměřujících se na vliv násilných her pracuje s úplně odlišnými hrami, což je velmi problematické (Elson et al., 2014). Stejně tak kritizuje tuto praxi i Ferguson (2015), protože to mohou být právě tyto opomíjené aspekty hry, které mohou ke zkoumané agresivitě nemalou mírou přispívat, a přitom jsou často přecházeny. Z tohoto důvodu se jako problematické jeví porovnávání násilných a nenásilných her, protože tyto hry se často liší v mnoha dalších ohledech, než je pouhá (ne)přítomnost násilí. To přirozeně znesnadňuje tvoření závěrů o skutečném vlivu násilných her, resp. přímo samotného násilí.

Pokud jsou použity stejné násilné hry pro všechny skupiny účastníků, experimentální zásah se obvykle omezuje na výměnu násilných prvků za prvky nenásilné, ale již bez kontroly kompetice a jiných proměnných. Například, hráči místo plamenometu drží „generátor duhy“ (Kneer et al., 2016). Dalším problémem je, že pro měření agresivity (nebo agresivního chování) se někdy používá nevhodná metoda, která sama o sobě může být soutěživá – například tzv. Taylor Competitive Reaction Time Test. V té participant domněle soutěží s neexistujícím protihráčem a za jeho neúspěchy ho trestá (Epstein & Taylor, 1967). Skutečnost, že je takový nástroj sám o sobě soutěživé povahy přirozeně ohrožuje jeho validitu (Adachi & Willoughby, 2011; Parrott & Giancola, 2007; Ritter & Eslea, 2005). Tyto nedostatky jsme vzali v potaz a v námi prezentovaném příspěvku se s nimi snažíme vypořádat. Náš výzkum vychází převážně z tzv. General Aggression Model (Carnagey & Anderson, 2004), což je určitá forma metateorie, která popisuje vznik agrese pomocí kombinace více přístupů do jednoho uceleného schématu a zohledňuje mj. i důležitost arousalu, který přispívá k vyššímu agresivnímu chování.

2 Cíl výzkumu

Cílem výzkumu je přispět do diskuse na téma „násilných videoher a agresivity“ zkoumáním vlivu soutěživého faktoru – často přítomného v násilných hrách. Dalším přínosem pak bylo zohlednění limitů předchozích studií, které se soutěživostí často nepracují, nebo pracují nevhodným způsobem. Hlavní otázka, kterou si v našem výzkumu klademe je: Jaký vliv má na hráče hraní vysoce kompetitivní hry (bez ohledu na přítomnost násilí)? Jak faktor soutěživosti vstupuje do (případného) následného agresivního chování? Které „vnitřní proměnné“ k takovému chování mohou přispívat? Předpokládáme totiž, že v důsledku soutěživosti vzroste arousal hráčů, což se pak projeví i na vyšší míře agresivního chování.

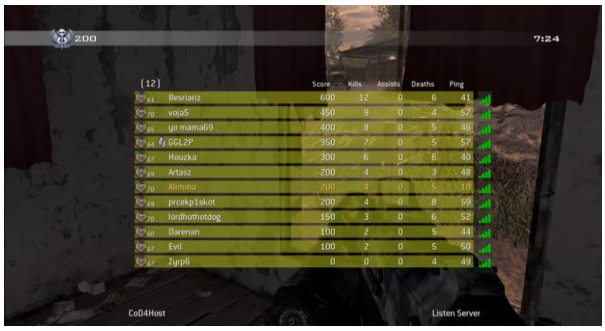
Zmíněný **arousal**, neboli „aktivace“ či mobilizace energie (Duffy, 1957; Nolen-Hoeksema et al., 2012) je stavem „fyziologické aktivace nebo kortikální citlivosti, spojený se smyslovou stimulací a aktivací vláken z retikulárního aktivačního systému“ (VandenBos, 2015). V průběhu arousalu (aktivace) dochází ke konkrétním fyziologickým změnám (mj. ke zrychlenému dýchání, zvýšení srdečního tlaku a zrychlení srdečního tepu) (Nolen-Hoeksema et al., 2012).
Již diskutovaná **agrese** může být charakterizována jako cílené chování s úmyslem způsobit negativní následky jiné osobě (Výrost et al., 2019). Agresivní chování pak představuje jakoukoli formu chování, která má za cíl zranit jinou živou bytost, přičemž osoba, vůči které je toto chování směřováno je přitom motivována k tomu, aby se takovému zacházení vyhnula (Baron & Richardson, 1994).

3 Příprava herních podmínek

V následujících podkapitolách budou detailně představeny různé herní varianty, které byly použity v této studii. Pro experimentální hraní byla vybrána hra Call of Duty: Modern Warfare 2, vyvinutá společností Infinity Ward. Jedná se o akční střílečku z pohledu první osoby, která byla uvedena na trh v roce 2009 společností Activision. Děj hry se odehrává v současném období (resp. v roce 2016) a obsahuje jak režim pro jednoho hráče s příběhem, tak online multiplayerový režim, který umožňuje hráčům soupeřit mezi sebou navzájem. V našem případě jsme použili právě tuto hru pro více hráčů.
Pro zajištění experimentálních podmínek s potřebnou manipulací a kontrolou všech nezbytných proměnných byla využita multiplayerová modifikace této hry s názvem iw4x_bot_warfare ve verzi 2.0.1, postavená na platformě IW4x (xLabs, 2022). Tato platforma je založena na původním herním enginu IW4 od Infinity Ward a je volně dostupná na serveru GitHub spolu s danou modifikací. Skupina xLabs upravila původní hru a umožňuje tak přidávat různé grafické balíčky a rozšíření, jako je například iw4x_bot_warfare.

3.1 Soutěživá varianta hry

Ve skupině, která hrála soutěživou variantu hry, měli účastníci za úkol soupeřit s ostatními herními postavami. Místo skutečných hráčů-protivníků byli využiti počítačově řízení boti, což garantovalo, že se účastníci nepřipojí do hry s extrémně silnými nebo slabými hráči. Toto jim ale nebylo předem sděleno, takže někteří skutečně mohli mít po celou dobu studie pocit, že hrají proti opravdovým hráčům. Pro zvýšení dojmu reálné hry byli boti pojmenováni skutečnými herními přezdívkami a psali zprávy do herního chatu v typickém videoherním slangu, včetně nadávek a posměšků a které se adekvátně týkaly nastalé herní situace. Tak například právě zabitý si stěžoval na podvádění u prvního hráče v žebříčku, při souboji s jiným hráčem se objevovala chvála nebo nadávky a podobně. Herní mód, ve kterém účastníci soutěžili, byl "všichni proti všem", což znamenalo, že na herní mapě nebyli žádní spojenci, a každý hráč se snažil dosáhnout co nejvyššího osobního skóre zabíjením ostatních. Když hráč zabil herního protivníka, na jeho obrazovce se zobrazil text s informací o navýšení jeho herního skóre, které si mohl během hry sledovat v žebříčku, který se průběžně měnil. Příklad žebříčku je na obrázku číslo 1.



Obr. 1: Tabulka herního žebříčku.

Herní kolo bylo nastaveno na 19 minut a všichni účastníci hráli za stejných podmínek. Údaj o zbývajícím čase byl vidět na obrazovce po celou dobu hry a minutu před koncem začala hrát dramatická hudba s tikajícím zvukem, čímž symbolizovala blížící se konec herní doby. Hráči, kteří v běžném životě hrají na svých vlastních herních účtech, během hraní získávají herní body a zkušenosti, což jim například umožňuje odemknout nové funkce ve hře. Proto je obecným cílem hrát co nejlépe. V laboratorních podmínkách je však obtížné tento efekt vytvořit, protože účastníci nehrají na svých účtech a v případě prohry tudíž ani neriskují snížení svého dlouhodobějšího, herního hodnocení. Pro kompenzaci těchto laboratorních omezení byla soutěživost podpořena nabídkou finanční odměny pro takové účastníky, kteří se v žebříčku po skončení hry umístili mezi prvními třemi pozicemi. Celkově hrál 1 člověk a 11 počítačem řízených botů. Aktuální skóre bylo zobrazeno na obrazovce vlevo dole vedle zbývajícího času a mohlo být sledováno v dynamicky se měnícím žebříčku (viz obrázek č. 1).
Pro vhodné stanovení videoherní obtížnosti byla provedena dvoufázová pilotní studie o celkovém počtu osmi participantů. Na jejím základě byla v samotném experimentu obtížnost upravena pseudoadaptivně.
Účastníci, kteří deklarovali, že střílečky hrají pravidelně, měli obtížnost botů nastavenou na „hard“ (těžkou), zatímco účastníci, kteří pravidelně nehráli, měli zachovánu původní obtížnost nastavenou na „medium“ (čili střední). Tento přístup byl již dříve použit jinými výzkumníky (Dowsett & Jackson, 2019).

Obr. 2: Screenshot soutěživé, násilné varianty hry.

3.2 Nesoutěživá varianta hry

Za účelem vytvoření nesoutěživého prostředí v experimentu nebyli herní protivníci agresivní a neútočili na participanty. Záměrem bylo eliminovat i vyrovnané nebo snadné souboje, neboť i ty mohou být vnímány jako forma soutěžení. Tento přístup byl již dříve použit v jiném výzkumu (Scarf et al., 2020). Participantům bylo zadáno, aby v průběhu hry zabíjeli jakékoli herní postavy, které spatří. Hráčům nebyla zobrazována tabulka s herním skóre a nezískávali žádné herní body za zabití protivníků. Herní kolo trvalo 30 minut a nebyl kladen časový tlak, přičemž informace o časovém limitu a aktuálním skóre byly skryty pomocí příkazového řádku hry. Po uplynutí 19 minut hraní ukončil experimentátor a v této variantě nebylo možné získat jakoukoliv finanční odměnu. Příklad nesoutěživé herní varianty je na obrázku číslo 3.

Obr. 3: Screenshot nesoutěživé, násilné varianty hry.

4 Měření arousalu

V našem případě jsme tedy srdeční aktivitu (jakožto fyziologický marker arousalu) monitorovali pomocí hrudního pásu Vernier Go Wireless® Excercise Heart Rate (kód produktu GW-EHR). Pro záznam, ovládání a export dat byla použita zdarma dostupná mobilní aplikace Elite HRV pro Android ve verzi 5.5.5, běžící na tabletu. Během hraní měli účastníci na sobě hrudní pás, který snímal časové intervaly (R-R, měřeno v milisekundách) mezi jednotlivými srdečními stahy. Na začátku experimentu byla u všech účastníků změřena činnost srdce v klidovém stavu, aby bylo možné činnost srdce při hraní lépe interpretovat. Všechny záznamy srdeční aktivity byly následně očištěny od artefaktů (byly vyřazeny všechny hodnoty, které byly větší a menší než 2,5 směrodatných odchylek od každého měření) a hrubé skóry byly převedeny na standardní Z-skóre, aby bylo možné porovnat tepovou frekvenci během hraní hry s klidovým stavem účastníka. To umožňuje určit, zda byla videohra pro účastníka výrazněji stimulující nebo jestli při hraní jeho srdce naopak tlouklo pomaleji. Činnost srdce se u každého jedince mírně liší, takže standardní rozsahy tepových frekvencí neplatí pro všechny jedince stejně bez výjimky. Standardní Z-skóre bylo vypočítáno jako rozdíl mezi průměrným časovým úsekem mezi srdečními stahy během hraní hry (H) a průměrným časovým úsekem v klidovém stavu (K), což bylo vyděleno výběrovou směrodatnou odchylkou získanou z intervalů mezi srdečními stahy při klidovém stavu (S_k).

$$Z - HR = \left(\frac{\overline{K} - \overline{H}}{S_k}\right)(-1)$$

Čím více se jeho hodnota blížila nule, tím menší byl rozdíl ve fyziologickém vybuzení mezi hrou a klidovým stavem. Čím dále se od nuly (která je v případě z-skóru průměrem) vzdalovala (u z-skórů je směrodatná odchylka standardně 1), tím odlišnější obě měření byla. V případě záporných hodnot tlouklo participantovi srdce rychleji při navozování klidového stavu, kladné hodnoty naopak značí vyšší srdeční aktivitu při hraní hry. Tento převod umožňuje velmi snadné statistické vyhodnocení i v rámci jiných typů analýz, protože obě měření o tisících parametrech transformuje do jediné, redukované hodnoty. Ilustrativní, zprůměrované záznamy srdeční činnosti v závislosti na hře jsou zobrazeny na obrázku č. 4.

Obr. 4: Předběžné, grafické srovnání srdeční činnosti v průměrných úderech za minutu v závislosti na herní podmínce. Osa X = herní doba, osa Y = srdeční tep.

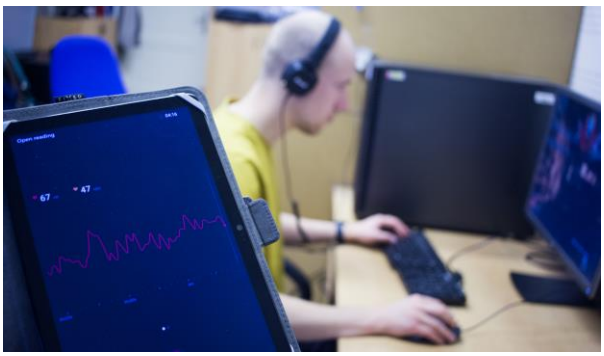
5 Měření agresivního chování

Pro měření míry agresivního chování bylo užito tzv. Hot sauce paradigma, což je typická metoda pro laboratorní měření agresivního chování a která bývá – i

navzdory některým svým nedostatkům – používána v kvalitních studiích (C. A. Anderson et al., 2010). Participant v této situaci věří, že odměřuje druhému účastníkovi pálivou omáčku, kterému bude posléze přidána do jídla. Původní participant zároveň dopředu ví, že účastník, který má jídlo ochutnávat pálivá jídla nesnáší. Ve skutečnosti však žádný ochutnavač neexistuje a množství omáčky, které je odměřeno se chápe jako indikátor míry agresivního chování. Pro více informací o této metodě a jejích limitech, viz práce Beierové (2013).

6 Průběh experimentu

Samotný experiment je navržen na několik částí. V úvodní fázi je účastníkům administrován falešný dotazník na jejich vlastní chuťové preference a připevněn hrudní pás. Od následné relaxační fáze, během které je monitorována srdeční aktivita participanta v klidovém stavu pak participant plynule přechází k videohernímu zácviku. Když se participant seznámí se všemi potřebnými ovládacími prvky, je náhodně přiřazen buď do soutěživé, nebo nesoutěživé podmínky a následuje fáze hraní. Po uplynutí příslušného herního času je účastník vyzván k odměřování pálivé omáčky dalšímu (ve skutečnosti neexitujícímu) participantovi, u kterého věří, že jídlo skutečně bude jíst. Pro podpoření tohoto dojmu je celý výzkum koncipován jako studie vlivu hraní na chuťové preference. Když je omáčka odměřena, participant je požádán o vyplnění několika závěrečných dotazníků, po kterých následuje debriefing a experimentátor se s účastníkem rozloučí. Průběh samotné fáze hraní za monitoringu srdeční aktivity je zachycen na následujícím obrázku č. 5.



Obr. 5: Fotografie z průběhu experimentu.

7 Závěr

Cílem tohoto příspěvku je upozornit na důležitost precizního stanovení nezávislé proměnné, což však bývá některými autory nedostatečně reportováno. Z tohoto důvodu obsah tohoto textu zaměřený právě na důkladný popis herních podmínek. Domníváme se, že přínos této studie spočívá zejména v tom, že je použita identická hra pro všechny experimentální podmínky; herní manipulace je v souladu s literaturou, kde zároveň participanti hrají po dobu, která se více blíží reálné době hraní, nikoliv jen pár minut a konečně ve které je užíváno méně problematické Hot Sauce paradigma (oproti CRTT). Výzvou pro budoucí výzkumníky, kteří by se chtěli podobným tématem zabývat bude pravděpodobně ještě jemnější nastavení vhodné videoherní obtížnosti, protože obtížnost má patrně přímý vliv na herní frustraci. Posledním přínosem může být zařazení fyziologického měření srdečního tepu, které slouží jako objektivní měření arousalu. Právě zaměření pozornosti na fyziologické nabuzení dle upozornění (Elson et al., 2015) se tak při hraní jeví jako hraní klíčová proměnná, která by měla být brána v potaz.

Literatura

Adachi, P. J. C., & Willoughby, T. (2011). The effect of violent video games on aggression: Is it more than just the violence? *Aggression and Violent Behavior*, *16*(1), 55–62. https://doi.org/10.1016/j.avb.2010.12.002

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., Rothstein, H. R., & Saleem, M. (2010). Violent Video Game Effects on Aggression, Empathy, and Prosocial Behavior in Eastern and Western Countries: A Meta-Analytic Review. *Psychological Bulletin*, *136*(2), 151–173. https://doi.org/10.1037/a0018251

Baron, R. A., & Richardson, D. R. (1994). *Human Aggression*. Springer Science & Business Media.

Beier, S. (2013). *Choose a juice! The effect of choice options, demand and harmful intentions on aggression in a modified Hot Sauce Paradigm.* Ruprecht-Karls-Universität Heidelberg.

Calvert, S. L., Appelbaum, M., Dodge, K. A., Graham, S., Hall, G. C. N., Hamby, S., Fasig-Caldwell, L. G., Citkowicz, M., Galloway, D. P., & Hedges, L. V. (2017). The American Psychological Association Task Force Assessment of Violent Video Games: Science in the Service of Public Interest. *American Psychologist*, *72*(2), 126–143. https://doi.org/10.1037/a0040413

Carnagey, N. L., & Anderson, C. A. (2004). Violent video game exposure and aggression: A literature review. *Minerva Psichiatrica*, *1*(45), 1–18.

Dowsett, A., & Jackson, M. (2019). The effect of violence and competition within video games on

aggression. *Computers in Human Behavior*, *99*, 22–27. https://doi.org/10.1016/j.chb.2019.05.002

Duffy, E. (1957). The psychological significance of the concept of "arousal" or "activation." *Psychological Review*, *64*(5), 265–275. https://doi.org/10.1037/h0048837

Elson, M., Breuer, J., Looy, J. V., Kneer, J., & Quandt, T. (2015). Comparing Apples and Oranges? Evidence for Pace of Action as a Confound in Research on Digital Games and Aggression. *Psychology of Popular Media Culture*, *4*(2), 112–125. https://doi.org/10.1037/ppm0000010

Elson, M., Breuer, J., & Quandt, T. (2014). *Handbook of Digital Games*. 362–387. https://doi.org/10.1002/9781118796443.ch13

Epstein, S., & Taylor, S. P. (1967). Instigation to aggression as a function of degree of defeat and perceived aggressive intent of the opponent1. *Journal of Personality*, *35*(2), 265–289. https://doi.org/10.1111/j.1467-6494.1967.tb01428.x

Ferguson, C. J. (2015). Do Angry Birds Make for Angry Children? A Meta-Analysis of Video Game Influences on Children's and Adolescents' Aggression, Mental Health, Prosocial Behavior, and Academic Performance. *Perspectives on Psychological Science*, *10*(5), 646–666. https://doi.org/10.1177/1745691615592234

Kneer, J., Elson, M., & Knapp, F. (2016). Fight fire with rainbows: The effects of displayed violence, difficulty, and performance in digital games on affect, aggression, and physiological arousal. *Computers in Human Behavior*, *54*, 142–148. https://doi.org/10.1016/j.chb.2015.07.034

Nolen-Hoeksema, S., Fredrickson, B. L., Loftus, G. R., & Wagcnaar, W. A. (2012). *Psychologie Atkinsonové a Hilgarda*. Portál.

Parrott, D. J., & Giancola, P. R. (2007). Addressing "The criterion problem" in the assessment of aggressive behavior: Development of a new taxonomic system. *Aggression and Violent Behavior*, *12*(3), 280–299. https://doi.org/10.1016/j.avb.2006.08.002

Ritter, D., & Eslea, M. (2005). Hot Sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior*, *31*(5), 407–419. https://doi.org/10.1002/ab.20066

Scarf, D., Zimmerman, H., & Jao, C.-W. (2020). *Failure to demonstrate competitive video game play increases aggressive affect, cognition, or behavior*. https://doi.org/10.31234/osf.io/mwrtu

VandenBos, G. R. (2015). *APA dictionary of psychology (2nd ed.)*. American Psychological Association. https://doi.org/10.1037/14646-000

Výrost, J., Slaměník, I., & Sollárová, E. (2019). *Sociální psychologie: teorie, metody, aplikace*. Grada.

xLabs. (2022). *IW4x*. xLabs. https://github.com/XLabsProject/iw4x-client

Asociovanie videného obrazu a motorickej akcie na jeden pokus

Andrej Lúčny

Katedra aplikovanej informatiky, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského
KAI FMFI UK, Mlynská dolina, 842 48 Bratislava
lucny@fmph.uniba.sk

Abstrakt

Pojednávame o možnostiach učenia na jeden pokus, ktoré skúšame na reakcii robota na videný obraz. Vysvetľujeme, prečo je priame asociovanie obrazu s akciou neúčinné a prečo pomocou kóderov a dekodérov získaných hlbokým učením je možné túto úlohu úspešne vyriešiť. Pritom pre použitie asociácii využívame mechanizmus kľúčov a hodnôt, ktorý je jadrom tzv. transformátorov. Priestor príznakových vektorov – či už kódujúcich obraz alebo motorickú akciu – je totiž na rozdiel od priestoru vstupných obrazov a výstupných motorických akcií plynulý a každý jeho prvok zodpovedá ako tak rozumnej inštancii situácie, v rámci ktorej robot koná. Prínosom nášho príspevku je vhodný spôsob reprezentácie percepcie a akcie ako aj ich asociovania.

1 Úvod

V prírode pozorujeme rôzne podoby učenia sa. V Dennet 2008 sa rozlišuje medzi tzv. skinnerovským a popperovským typom mysle. Kým v prvom prípade prebieha zdokonaľovanie schopností postupným učením, na ktoré je potrebné podstúpiť veľa opakovaných pokusov, v druhom je možné aby proces učenia prebehol náhle, na základe jedinej skúsenosti s predmetnou situáciou. Hoci súčasné výdobytky umelej inteligencie nás vedia ohúriť svojimi schopnosťami, získavajú ich skinnerovským, nie popperovským spôsobom: v procese trénovania, pri ktorom je im každý zo vzorov správania predložený mnoho krát. Bolo by ale možné z nejakou takto vopred pripravenou sadou schopností rozbehnúť zdokonaľovanie popperovským spôsobom?

Táto otázka je naliehavá hlavne v mobilnej robotike, kde na robotovi máme už dnes k dispozícii technické možnosti spúšťania modelov hlbokého učenia, ale vykonať trénovanie alebo doladenie týchto modelov je kapacitne problematické.

V tomto príspevku sa zameriavame na zjednodušenú situáciu, kedy má takýto robot získavať schopnosť zvoliť správnu akciu na základe určitej percepcie. Konkrétne asociujeme motorickú akciu vedúcu k zaujatiu určitej pózy tela humanoidného robota (používame iCubSim z Vernon 2007) na základe videného obrazu pri tzv. imitačnej hre (Boucenna 2014) (kapitola 2). Pritom robot je vždy určitej situácii fyzicky vystavený, takže zapamätať si asociáciu medzi videným obrazom a motorickou akciou nie je problém. Problém je vedieť tieto asociácie neskôr použiť. Háčik spočíva v tom, že robot sa už nikdy viac nedostane do presne rovnakej situácie, než v akej si asociáciu zapamätal. Musíme teda navrhnúť nejaký šikovný mechanizmus pomocou ktorého robot reaguje na obraz, ktorý ešte nikdy nevidel a na ktorý sa to, čo pozná, len podobá. Toto riešime pomocou tzv. attention mechanizmu z Vaswani 2017, ktorý je bežnou súčasťou transformátorových hlbokých neurónových sietí, pričom my ho používame pomerne netradičným spôsobom (kapitola 3).

Avšak ani takýto mechanizmus nám reálne neumožní implementovať imitačnú hru, ak by sme asociovali priamo videný obraz a motorickú akciu. Ich dátové priestory sú príliš veľké, deravé a málo plynulé. Pokiaľ napríklad videnú situáciu reprezentujeme ako bod v mnohorozmernom priestore, tak:

- počet dimenzií tohto priestoru je daný súčinom počtu pixelov a farebných kanálov, čo sú rádovo státisíce
- v priestore máme všetky možné obrazové vstupy, pričom väčšinu z nich robot nikdy nemôže reálne uvidieť
- pri drobnej zmene situácie môže v takom dátovom priestore dôjsť k dramatickej zmene polohy reprezentujúceho bodu

Čo sa týka motorickej akcie:

- počet dimenzií bude rovný počtu stupňov voľnosti robota, t. j. počtu kĺbov, ktorých uhlami robota riadime (uhlové rýchlosti zanedbávame), čo síce nebude až tak veľa (ide o jednotky až desiatky), avšak:
- v priestore máme všetky možné variácie kĺbových uhlov, z ktorých väčšinu tvoria pózy, ktoré nie je vhodné zaujať (na rozdiel od človeka robot necíti, ktorá póza je mu príjemná a ktorá nekomfortná)
- pri drobnej zmene nastavenia stupňov voľnosti, väčšinou prichádza k drobnej zmene pózy, avšak vplyv zmeny rôznych stupňov voľnosti je veľmi rozdielny: hierarchicky vyššie stupne majú rádovo väčšie dopady na výslednú pózu

Prekonať zlé vlastnosti priestorov obrazov a póz nám však umožňuje hlboké učenie. Práve túto schopnosť jeho modelov považujeme za kľúčovú (kapitola 4).

Keď obraz a pózu spracujeme modelmi hlbokého učenia na zodpovedajúce príznaky, tieto dva príznakové vektory budeme vedieť nielen asociovať, ale túto asociáciu aj účinne použiť. Ich dátové priestory sú totiž:

- oveľa menšie (v prípade obrazu stovky – maximálne tisíce – dimenzií, v prípade pózy ide o jednotky)
- bez dier: každý príznakový vektor zodpovedá nejakému možnému videnému obrazu, respektíve rozumnej póze, ponajviac nejakej prechodnej forme medzi dvoma takými obrazmi či pózami
- plynulé: príznakové vektory zodpovedajúce postupnej zmene videnej situácie či zaujatej pózy predstavujú v priestore príznakov trajektóriu podobnú prechodu medzi počiatočným a koncovým stavom.

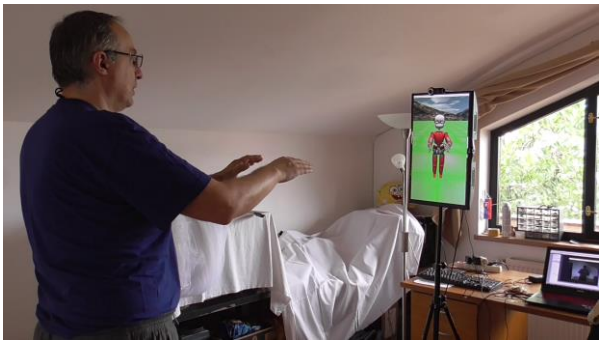
Vďaka týmto vlastnostiam príznakov sme schopní imitačnú hru implementovať (kapitola 5), čím implementujeme ukážku učenia sa robota na jeden pokus.

2 Imitačná hra

Cieľom imitačnej hry, na ktorej testujeme náš prístup, je naučiť robota imitovať človeka na základe toho, že človek imituje robota (obr. 1). Hra prebieha v dvoch fázach.

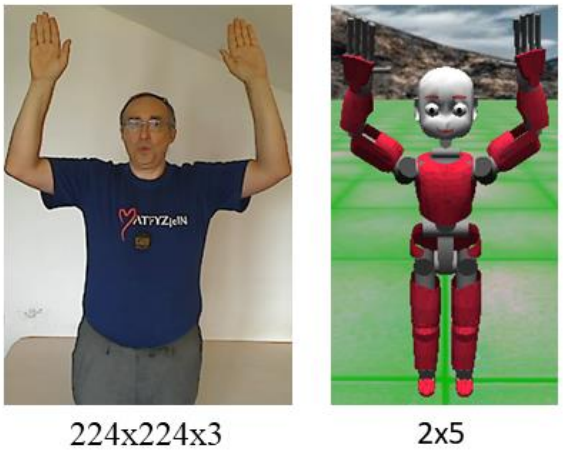
V prvej fáze robot vyzýva človeka, aby ho napodobňoval. Vytvára napríklad rôzne polohy rúk a človek ich napodobňuje svojím telom pred kamerou robota. To dáva šancu robotovi zapamätať si asociácie medzi pózami svojho tela a videným obrazom.

V druhej fáze robot napodobňuje človeka pomocou asociácií získaných v prvej fáze.



Obr. 1: Imitačná hra

Asociácie získané robotom predstavujú zoznam reprezentácii obrazu, ktorý robot vidí a pózy, ktorú robot zaujal. Obraz je vo svojej pôvodnej forme trojrozmernými poľom s rozmermi: výška, šírka a kanál, ktorého prvky sú intenzity od 0 do 255. V našej implementácii s farebným obrazom (3 kanály) používame rozlíšenie obrazu 224x224 pixelov. I pri tomto skromnom rozlíšení je počet dimenzií tohto dátového priestoru úctyhodných 224x224x3 = 150528. Pri póze uvažujeme len polohy rúk robota, čo je 2x5 = 10 stupňov voľnosti. V pôvodnej forme teda ide o vektor desiatich čísel zodpovedajúcim stupňom, ktoré normalizujeme do intervalu $<-1,1>$ (čo je vhodnejší rozsah pre spracovanie neurónovou sieťou). (Obr. 2) Spôsob reprezentácie oboch si môžeme zvoliť. Prínosom tohto príspevku je, že sme našli vhodný taký spôsob.



Obr. 2: Asociované dáta: obraz a póza

Technicky je náročné zabezpečiť, aby robot vo prvej fáze správne vystihol moment, kedy človek zaujme jeho pózu. To sme si zjednodušili tak, že to človek robotovi naznačí. Keďže jeho ruky sú zaneprázdnené zaujatím správnej pózy, využili sme na to detektor zapískania. Ten funguje na báze Fourierovej transformácie zvuku a analýzy amplitúdového spektra.

3 Mechanizmus asociovania

Nami používaný mechanizmus asociovania, známy pod – pre nás nie celkom vhodným – menom attention (pozornosť), pracuje s množinou l párov kľúč - hodnota. Keď máme na vstupe dotaz q, namiešame dotaz z kľúčov K a výstup vytvoríme ako analogickú zmes zo zodpovedajúcich hodnôt V, kde:

$$K = \begin{pmatrix} k_1 \\ k_2 \\ \dots \\ k_l \end{pmatrix} \qquad V = \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_l \end{pmatrix}$$

Všetky dotazy a kľúče sú vektory dimenzie n, takže K je matica $l \times n$. Hodnoty a výstupy sú vektory dimenzie m, takže V je matica $l \times m$. Najprv nájdeme také $c_i \in <0,1>$, že $\Sigma\, c_i\, k_i = pr_K(q)$, $\Sigma\, c_i = 1$ a $i=1,2,...l$, kde

$pr_K(q)$ je vektor podobný projekcii q do podpriestoru generovaného kľúčmi K. Pritom chceme, aby c_i vyjadrovalo podobnosť medzi kľúčom k_i a dotazom q, takže ho môžeme odvodiť od skalárneho súčinu $q\ k_i$, úmernému uhlu, ktorý q a k_i zvierajú. Tieto podobnosti – pozitívne pre zhodné, nulové pre navzájom kolmé a záporné pre opačné vektory – však musíme dostať do $<0,1>$, čo nám dokáže zariadiť funkcia $softmax(x)_i = exp(x_i) / \Sigma_k\ exp(x_k)$. Koeficienty pomocou ktorých namiešame z kľúčov k_i niečo podobné dotazu q, zvolíme preto ako:

$$c = softmax\left(\frac{qK^T}{d}\right)$$

kde d je škálovací faktor, ktorým určujeme koľko namiešame z podobných kľúčov a koľko z odlišných. Čim menšia táto konštanta je, tým sa viac sa koeficienty blížia k tzv. one-hot kódu (jedna jednotka a ostatné nuly). Pre $d = 1/n$, kde n je dimenzia kľúčov sa už prakticky vždy prikloníme k prevahe jedného kľúča, zatiaľ čo obľúbená hodnota $d = \sqrt{n}$ zabezpečuje, že vždy trochu miešame aj z ostatných kľúčov. To môže byť prínosom pre schopnosť asociačného mechanizmu vynájsť správnu odozvu aj pre také dotazy, ku ktorým podobný kľúč zapamätaný nemá, avšak dajú sa považovať za prechodnú formu medzi dvomi, či viacerými zapamätanými kľúčmi.

Keď už máme koeficienty zmesi c, ktorými sme približne vyjadrili dotaz podľa kľúčov, môžeme analogickým spôsobom zmiešať hodnoty V na výstup $o = cV$. Takže úplná odpoveď asociačného mechanizmu **A** na dotaz q je:

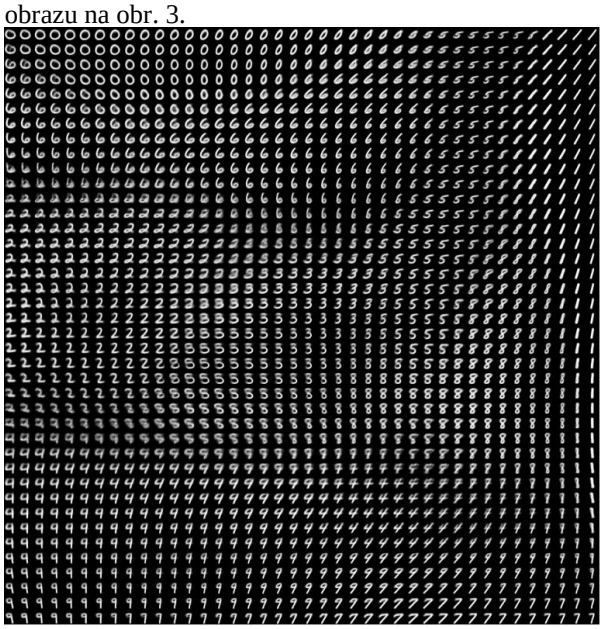
$$A(q, K, V) = softmax\left(\frac{qK^T}{d}\right)V$$

Sám o sebe však tento mechanizmus našu úlohu nevyrieši. Keby sme asociovali priamo obraz s akciou (čo je pri dimenziách $n=150528$ a $m=10$ technicky možné), dostávali by sme v konečnom dôsledku skoro vždy rovnakú odozvu, v dôsledku čoho by sa imitujúci robot sotva pohol.

4 Príprava modelov hlbokého učenia

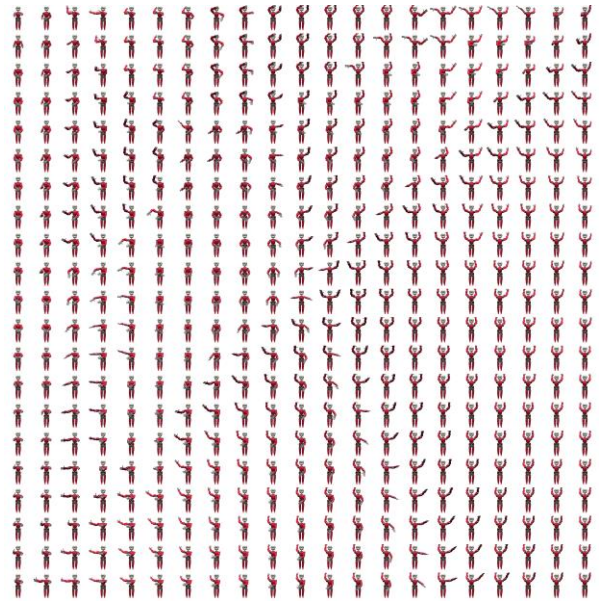
Na to, aby asociačný mechanizmus fungoval, potrebujeme, aby priestory kľúčov a hodnôt boli plynulé a neobsahovali žiadne diery. To sa dnes dá našťastie ľahko zariadiť, lebo práve tieto vlastnosti sú podstatou fungovania modelov tzv. hlbokého učenia. Náš robot do hry vstupuje s dvoma hotovými modelmi: kóderom obrazu a dekóderom pózy. Oba sa dajú získať bez potreby anotovania dát.

Ako kóder obrazu sme využili predtrénovaný vizuálny transformátor trénovaný samoučením (metódou DINO podľa Caron 2021, ktorej podstatou je predkladanie podobných a rozdielnych obrazov dvom kópiám tej istej siete, pričom v prvom prípade požadujeme rovnakú odozvu a v druhom rozdielnu), ktorý obraz kóduje do 384 príznakov. Urobiť si určitú predstavu o tom čo robí, je možné z oveľa jednoduchšieho príkladu analogického spracovania obrazu na obr. 3.



Obr. 3: Mapovanie ručne písaných číslic s rozlíšením 28x28 do latentného priestoru dimenzie 2. Podobné mapovanie robí DINO pre obraz pred kamerou, avšak v tomto prípade má latentný priestor 384 dimenzií a je ťažko si ho predstaviť

Dekodér pózy sme získali natrénovaním variačného autokódera (Kingma 2019) (vybrali sme lepší výsledok z viacerých trénovacích pokusov) z dátových vzoriek pohybov rúk robota do bodov v okolí robota. Tieto pohyby sme získali spätnou kinematikou (obr 4).



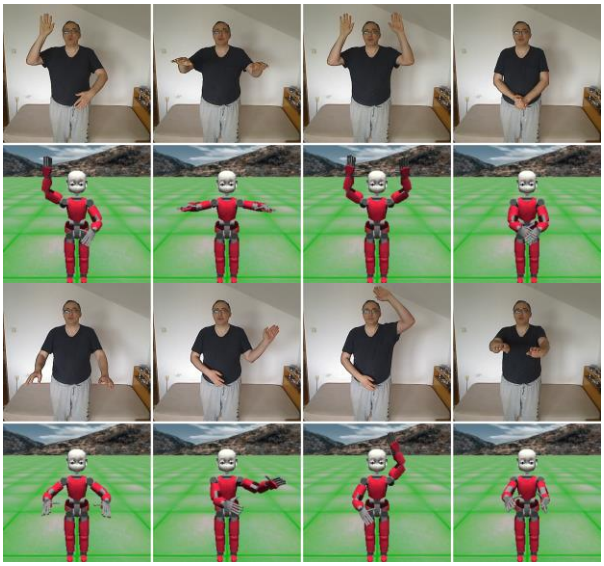
Obr. 4: Mapovanie póz robota 2x5 stupňov voľnosti do latentného priestoru dimenzie 2 pomocou konvolučného variačného autokódera.

Učenie robota imitácii bude spočívať v tom, že pomocou asociačného mechanizmu namapujeme jeden latentný priestor na druhý.

5 Implementácia imitačnej hry

V prvej fáze imitačnej hry robot z niekoľko vybraných príznakových vektorov pózy dekóduje pózu, zaujme ju, počká na signál od človeka, že aj on zaujal správnu pózu, zakóduje videný obraz do príznakov a uloží si oba príznakové vektory do zoznamu asociácii (príznaky obrazu budú kľúčom a príznaky pózy hodnotou). Na naučenie sa určitej sady polôh, stačí toľko asociácii, koľko je polôh, prípadne o niečo málo menej, keďže niektoré polohy je možné zložiť z ostatných.

V druhej fáze robot zakóduje obraz do príznakov, vypočíta z asociácii zodpovedajúci príznakový vektor pózy, dekóduje a zaujme zodpovedajúcu pózu (obr. 5). Pritom je schopný sa nielen prikloniť k niektorej zapamätanej póze, ale tieto aj vhodne skombinovať (táto schopnosť však silne závisí škálovacieho faktoru asociačného mechanizmu)[1].



Obr. 5: Priebeh imitačnej hry (druhá fáza)

Zaujímavou vlastnosťou tohto riešenia bolo, že ak človek robota v prvej fáze oklamal a miesto správnej pózy urobil niečo iné – napríklad mu ukázal nejaký objekt – robot sa túto neadekvátnu reakciu naučil[2].

6 Záver

Učenie asociovaním predpripravených modelov je prístupom, ktorý môže byť zaujímavý, ak potrebujeme, aby učenie prebehlo náhle a rýchlo. Užitočné môže byť hlavne pre mobilné roboty, ktoré na palubnom počítači vedia spúšťať modely hlbokého učenia, avšak nedisponujú kapacitou na ich trénovanie či doladenie. Vzbudzuje viacero otázok, ako dosiahnuť pri tomto spôsobe učenia čo najlepšiu kvalitu. Rezervy máme jednak v povahe kóderov, ktoré síce mapujú plynule, ale zaostávajú v uniformite, jednak v samotnom asociačnom mechanizme, ktorého rôzne obmeny sa chystáme skúmať.

Kódy a modely zdieľame na Github-e:
https://github.com/andylucny/cvae.git
https://github.com/andylucny/learningImitation.git

[1] viď video https://youtu.be/-3BVbU9BeRE
[2] viď video https://youtu.be/_CBnCOnWRdY

Poďakovanie

Tento príspevok vznikol s podporou grantovej agentúry VEGA v rámci projektu 1/0373/23 a EU projektu TERAIS, č. 101079338.

Literatúra

Boucenna, S., Anzalone, S., Tilmont, E., Cohen, D., Chetouani, M.: Learning of social signatures through imitation game between a robot and a human partner. *IEEE Transactions on Autonomous Mental Development 6(3)*, 213–225 (2014). https://doi.org/10.1109/TAMD.2014.2319861

Caron, M., Touvron, H., Misra, I., J´egou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: *Proceedings of the International Conference on Computer Vision. ICCV* (2021)

Dennett, D.C.: *Kinds of minds: towards an understanding of consciousness.* (1996) Weidenfeld & Nicolson, London

Kingma, D.P., Welling, M.: An introduction to variational autoencoders. (2019) *Foundations and Trends in Machine Learning 12(4)*, 307–392

Lucny, A.: Towards one-shot learning via attention. (2022) In: *CEUR Workshop Proceedings, ITAT 2022*. pp. 4–11. 3226

Šejnová, G., Štepánová, K.: Feedback-driven incremental imitation learning using sequential VAE. In: *2022 IEEE International Conference on Development and Learning (ICDL)*. pp. 238–243. IEEE (2022)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Lukasz Kaiser, Polosukhin, I.: Attention is all you need. (2017) In: *31st International Conference on Neural Information Processing Systems. ACM, Long Beach*

Vernon, D., Metta, G., Sandini, G.: The icub cognitive architecture: Interactive development in a humanoid robot. In: *2007 IEEE 6th International Conference on Development and Learning.* pp. 122–127 (2007). https://doi.org/10.1109/DEVLRN.2007.4354038

Unraveling the hidden Influence of Ernst Mach on the Foundations of Cognitive Science – Interdisciplinary approach

Ján Pastorek

Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava
pastorek20@uniba.sk

Isabella Sarto-Jackson

The Konrad Lorenz Institute for Evolution and Cognition Research (KLI)
Martinstraße, 3400 Klosterneuburg
isabella.sarto-jackson@kli.ac.at

Abstract

Existing narratives often overlook the significant impact of Ernst Mach and the Vienna Circle on the foundations of cognitive science. In this study, we delve into the underexplored influence of Mach's theories on the emergence of cognitive science, employing a unique interdisciplinary approach that blends rigorous argumentation with cutting-edge computational methods in network science and natural language processing. Our findings reveal multiple, previously unexplored pathways of influence from Mach to pivotal figures in cognitive science, thereby showcasing the efficacy of our combined approach in illuminating the intricate web of intellectual connections. This innovative method offers valuable insights into tracing the potential influences of key thinkers, addressing a longstanding challenge in the history of science arising from the ever-growing corpus of academic literature. To our knowledge, this is one of the first papers to use both citation networks and natural language processing for the investigations of the history of cognitive science.

1 Introduction

The foundations of cognitive science are often associated with research conducted in the United States and the United Kingdom (Riedl, 2022; Thagard, 2020). However, we know that before 1940s, Vienna was one of the world leading intellectual hubs with dominant figures such as Mach, Boltzmann, Schrödinger, and later it became the center of logical positivism regularly hosting Carnap, Schlick, Gödel, Neurath, Tarski, Popper, etc. Today we call this center, Vienna Circle. Furthermore, the Vienna Circle formulated novel philosophical doctrines that were deeply grounded in logic and science. This reconceptualization of philosophy contributed significantly to the emergence of the discipline of philosophy of science and several members and forerunners of the Vienna Circle were particularly interested in epistemology, the philosophical theory of knowledge. Therefore, the question arises: how did Austrian philosophers and scientists of the late 19th and early 20th century influence the birth and development of cognitive science? We focus on Mach's influences.

2 Methods

We used citation networks and natural language processing along with data mining techniques to find and validate these hypotheses. The approach is roughly outlined in Fig. 1, and summarized in several steps here:

1. Firstly, the citation network referring to Mach's publication *The analysis of sensations* in both English and German versions were extracted from Google Scholar. It contained 15,101 nodes and 16,707 links, and was extracted to a maximum depth of two citations away from the target papers. To reduce the scope of the extraction, only the first 200 most relevant papers (sorted by Google Scholar relevance algorithm) at a distance of one citation and a maximum of 100 papers for each of those 200 at a distance of two citations from the Mach's publication were extracted for further analysis.
2. We compiled a list of people that a) might have been influenced and b) were key to the development of cognitive science, and c) we extracted a subgraph with all the connections to these authors if indeed they were in the citation network. We call these connections *possible paths of influence*.
3. We downloaded the publications that satisfied a-c.
4. To begin with, we created a keyword frequencies plots for each publication, and searched for the overlapping keywords.
5. Then, we created UMAP representation of sentences from each publication for visual indications.

6. Furthermore, the classification model was trained using gradient boosting to classify sentences to publications using the H2o Python library and a set of default hyperparameters: $ntrees = 200$, $max\_depth = 4$, $col\_sample\_rate = 0.5$, $min\_rows = 10$, $nfolds = 5$, $learn\_rate = 0.1$, $learn\_rate\_annealing = 0.99$.
7. Hierarchical clustering was done on TF-IDF vectors extracted from texts using multiple linkages.
8. Lastly, we have read the relevant parts of publications to understand the possible paths of influence.

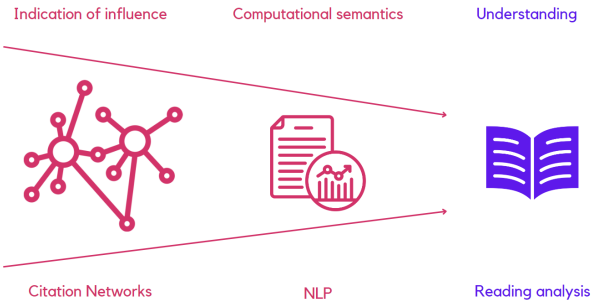


Fig. 1: The image shows a workflow of interdisciplinary approach for history of sciences. Firstly, citation networks are analyzed to identify potential indications of paths of influence. Then, NLP methods are used on the texts of publications to identify potential content of influences. Lastly, classical reading analysis is performed to confirm and elaborate on the findings.

3 Results

1. In the citation network, we found 22 connections to notable figures in Cognitive Science within two citations away from Mach, including Piaget and Skinner that directly cited Mach.
2. Furthermore, the word *Gestalt* was used in Mach's original publication in German, *Beiträge zur Analyse der Empfindungen*.
3. The classification model achieved $\approx 70\%$ accuracy with $0.99\ r^2$ and nice convergence.
4. Hierarchical clustering differentiated between publications as predicted.

4 Conclusion and discussion

Given that some publications were short, and many publications were of the same authors on similar topics, it was easier for the model to misclassify these publications. For instance, Skinner had 5 out of those 22 connections, some of which had only several sentences. In view of these limitations, we think that the model has nice performance which is supported by high r^2. However, for bigger publications, we hypothesize that misclassifications might signify similarities between authors and provide a supporting argument. For this reason, we constructed hierarchical clustering based on TF-IDF, and found another supporting data-driven evidence that some of these works are clustered according to the possible influences as expected.

Ernst Mach's ideas have had a significant impact on the development of Cognitive Science. He advocated for the unification of the physical and psychological, which is a fundamental characteristic of modern Cognitive Science (Pléh & Gurova, 2013). He interpreted cognition in an evolutionary selectionist framework, where hypotheses and trials are a key aspect of both science and everyday cognition (Mach, 1959). This view was further developed by Piaget (2005). Moreover, both Mach and Piaget share similar ideas about the adaptation of thoughts to sensations and the construction of reality and cognition (Riegler, 2012). The connection between Mach and Gestalt psychology lies primarily in Mach's similar use of the Gestalt concept in his publication *Contributions to the Analysis of Sensations*, which he used well before the Gestalt school even arose. Furthermore, his concept of economy of thought and the idea that the mind simplifies and abbreviates information aligns well with the Gestalt focus on global patterns and holistic processing. Additionally, Mach's idea that all is made out of sensations aligns with the Gestalt emphasis on the active role of perception in constructing the world. Several other paths of influence were identified without completing the last step of reading analysis.

Data availability

Data, analysis, and the code used in this study are available on GitHub at https://github.com/JanPastorek/mach_influence.

References

Mach, E. (1959). *The Analysis of Sensations*. Dover Publications.

Piaget, J. (2005). *Language and Thought of the Child: Selected Works vol 5*. Routledge.

Pléh, C., & Gurova, L. (2013). Existing and would-be accounts of the history of cognitive science: An introduction. In *New Perspectives on the History of Cognitive Science.*

Riedl, A. (2022). Cognitive Science Map. Retrieved 01/10/2023, from https://www.riedlanna.com/cognitivesciencemap.html

Riegler, A. (2012). Constructivism. In *Paradigms in theory construction* (pp. 235–255). Springer.

Thagard, P. (2020). Cognitive Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020). https://plato.stanford.edu/archives/win2020/entries/cognitive-science/

Explorácia pomocou internej motivácie a samokontrolovaného učenia

Matej Pecháč a Igor Farkaš

Fakulta matematiky, fyziky a informatiky
Univerzia Komenského v Bratislave
Email: {matej.pechac,igor.farkas}@fmph.uniba.sk

Učenie posilňovaním (reinforcement learning, RL) predstavuje významnú kategóriu metód strojového učenia, ktoré boli úspešne použité na riešenie rôznych sekvenčných úloh, kde agent interaguje s prostredím, napríklad v robotike alebo pri hraní počítačových hier. Jedným z kľúčových konceptov RL je vnútorná motivácia, ktorá umožňuje agentovi zlepšiť svoje správanie a dosahovať tak vyššie odmeny (kritérium pri RL).

Vnútorná motivácia (intrinsic motivation, IM) je definovaná (Ryan a Deci, 2000) ako vykonávanie činnosti pre prirodzené uspokojenie, a nie pre nejaký oddeliteľný následok (alebo inštrumentálnu hodnotu). IM bola operačne definovaná rôznymi spôsobmi, podloženými psychologickými teóriami, ktoré však poukazujú na určitú neistotu v tom, čo IM presne znamená. V kontexte RL, ak je motivácia generovaná v rámci štruktúr, ktoré sú súčasťou agenta, znamená to, že ide o internú motiváciu. Na to môže agent využívať hneď niekoľko prístupov, ktoré sa delia na dve veľké kategórie (Oudeyer a Kaplan, 2009): IM *založená na znalosti* a IM *založená na kompetencii*. Prvá vedie agenta k preskúmaniu každého možného stavu prostredia, druhá vedie zasa k nadobudnutiu všetkých možných zručností (napr. cieľom podmienené stratégie), ktoré v danom prostredí existujú. Podľa toho, akým mechanizmom dochádza ku generovaniu IM v prípade metód založených na znalosti, ich rozdelujeme na 3 prelínajúce sa kategórie: IM založená na predikčnej chybe, IM založená na detekcii novosti, IM založená na informačných konceptoch. Modely prvej kategórie obsahujú nejaký predikčný modul a jeho chyba slúži ako zdroj motivačného signálu. Modely druhej kategórie často využívajú koncept počítania návštevnosti stavu a z toho odvodzujú motiváciu agenta. Modely tretej kategórie sa snažia maximalizovať získanú informáciu z prostredia a napr. miera poklesu neistoty vo vzťahu k prostrediu alebo k agentovi je mierou motivácie.

V našej práci sme navrhli a otestovali novú triedu motivačných modelov *Self-supervised Network Distillation* (SND) založených na algoritmoch samokontrolovaného učenia (self-supervised learning) a využívaní destilačnej chyby ako detekcie novosti. Preto je ich možné zaradiť do rodiny metód IM založených na znalosti a pohybujú sa kdesi na rozhraní kategórie IM založenej na predikčnej chybe a detekcii novosti. Prvým takýmto modelom bol *Random Network Distillation* (RND) model (Burda a spol., 2018), ktorý sa stal základom našich modelov a v tomto kontexte sa javí ako ich špeciálny prípad. Naša metóda používa dva modely, *cieľový model*, ktorý poskytuje cieľové reprezentácie (vektorov príznakov) a *učiaci sa model*, ktorý sa ho snaží napodobniť (destilovať znalosti). Oba modely používajú ako svoj vstup reprezentáciu stavu v každom kroku diskrétneho času. Rozdiel vo výstupoch medzi oboma modelmi slúži ako signál pre vnútornú motiváciu. Predpokladali sme, že destilácia náhodného cieľového modelu poskytuje dostatočný signál iba na začiatku učenia. Preto sme navrhli tri metódy (Pecháč a spol., 2023) samokontrolovaného učenia pre trénovanie cieľového modelu: metódu SND-V založenú na kontrastívnom učení (Chopra a spol., 2005), metódu SND-STD založenú na SpatioTemporal DeepInfomax algoritme (Anand a spol., 2019) a metódu SND-VIC založenú na VICReg algoritme (Bardes a spol., 2022). Tieto metódy vytvorili priestor reprezentácií vhodných pre destiláciu. Spoločnou črtou metód samokontrolovaného učenia (v kontexte reprezentácií v RL) je predpoklad, že stavy, ktoré nasledujú po sebe, by mali mať veľmi podobné reprezentácie, zatiaľ čo stavy, ktoré sa v trajektórii nachádzajú ďalej od seba, by mali mať čo najrozdielnejšie reprezentácie. Tento cieľ je možné dosiahnuť optimalizáciou rôznych chybových funkcií spadajúcich do kategórie samokontrolovaného učenia.

Celkovo sme testovali naše metódy na 6 prostrediach Atari (Montezuma's Revenge, Gravitar, Venture, Private eye, Pitfall, Solaris) a 4 prostrediach Procgen (Coinrun, Caveflyer, Jumper and Climber), ktoré sa považujú za zložité na prehľadávanie prostredia (explorácia), pretože majú veľmi riedku externú odmenu. Z testovaných prostredí iba hra Pitfall predstavovala príliš zložitý problém, nakoľko žiaden zo všetkých testovaných algoritmov vôbec nezafungoval (t.j. nezískal ani jeden bod odmeny). V ostatných 9 prostrediach najlepšie výsledky dosiahli modely založené na SND metódach, pričom v 8 prípadoch to bolo s výrazným náskokom pred existujúcimi algoritmami (v prostredí Venture boli výsledky takmer rovnaké ako pri modeli RND). Pri porovnaní v skóre dosiahli modely SND najvyššie skóre v 5 prostrediach Atari hier a v 3 prípadoch (Montezuma's Revenge, Gravitar, Private Eye) bolo skóre výrazne vyššie ako porovnávané mo-

dely. Dosiahnuté skóre sa používa aj na porovnávanie s modelmi od iných tvorcov. V celosvetovom rebríčku [1] sa naše modely umiestnili zväčšna na 2. mieste. Okrem toho, pokiaľ nám je známe, sme prví, kto úspešne natrénoval agentov v Procgen prostrediach ťažkých na exploráciu.

V analýze sme sa zamerali na preskúmanie priestoru reprezentácií *cieľového modelu* pre RND aj SND algoritmy. Zostrojili sme viacero metód pre jeho analýzu a skúmali sme spojitosť medzi meranými parametrami a výsledkami daného modelu. Hoci ide o nelineárny vysokorozmerný priestor, rozhodli sme sa použiť nástroje lineárnej algebry a aspoň zhruba získať predstavu o niektorých jeho základných vlastnostiach.

Pomocou natrénovaného agenta sme zozbierali 10000 vzoriek pre jednotlivé prostredia. Najskôr sme pomocou QR dekompozície získali vektorovú bázu daného priestoru reprezentácií a zisťovali sme, či sú všetky dimenzie lineárne nezávislé, alebo či je reálna dimenzionalita daného priestoru nižšia. Z tejto analýzy vyšlo, že priestory vytvorené každou z metód majú všetky dimenzie lineárne nezávislé.

V druhom kroku sme vizualizovali pomocou dištančnej matice vzájomnú vzdialenosť vstupných stavov a následne vzájomnú vzdialenosť reprezentácií vytvorených jednotlivými metódami. Tam sme objavili zaujímavé výsledky. Zatiaľ čo RND vytváral v priestore reprezentácií podobnú štruktúru aká bola na vstupe v priestore stavov, SND metódy túto štruktúru do značnej miery potlačili. Posun v priestore stavov vyvolával veľký posun v priestore reprezentácií a bolo jedno "ktorým smerom" sme sa pohli. Inými slovami, malá aj veľká zmena vstupného stavu sa prejavila ako podobná zmena v reprezentácii, čo značne sťažovalo úlohu *učeného modelu*, ktorý sa snažil replikovať tieto reprezentácie. Naopak, pri RND modeli stačilo, ak sa natrénoval *učený model* niekoľko podobných stavov a zvyšné sa už od nich v priestore reprezentácií nelíšili, a preto ich ani nedetekoval ako nové, hoci mohli byť podstatné pre ďalšie napredovanie.

Ďalšou analýzou bol odhad natiahnutia jednotlivých dimenzií priestoru reprezentácií *cieľového modelu* (dalo by sa obrazne povedať, že išlo o odhad tvaru). Pomocou PCA metódy sme odhadli vlastné čísla lineárneho obalu a získali ich distribúciu. Z výsledkov bolo vidieť, že naše modely SND-VIC a SND-V majú oba tendenciu využívať rovnomerne všetky dimenzie, a obaly týchto priestorov majú tvar hyperelipsoidov, ktoré sa blížia k hyperguli. To koreluje s predchádzajúcimi zisteniami, že ľubovolná zmena v stavovom priestore (malá, či veľká) vedie k rovnakej zmene v rámci priestoru reprezentácií. Zdá sa, akoby neexistoval nejaký preferovaný smer, ale reprezentácie sú v ňom homogénne rozložené a každá reprezentácie je od každej približne rovnako vzdialená.

Poslednou analýzou SND modelov je pochopenie ich časového vývoja a schopnosti poskytnúť veľký signál internej odmeny pre predtým nevidené stavy. Pre účely explorácie je najdôležitejšia schopnosť odhaliť stavy v blízkej budúcnosti, ktoré sú veľmi podobné už videným stavom. Zozbierali sme súbor 2700 stavov od nášho najlepšieho agenta pre prostredie Montezuma's Revenge. Počas experimentu sme *cieľové a učené moduly* trénovali iba na vzorkách z minulosti a testovali sme ich citlivosť na vzorkách z budúcnosti, ktoré ešte nevideli. Pri všetkých troch SND metódach je vnútorná motivácia oveľa vyššia pre nevidené stavy a nekonverguje k nule ako pri RND.

Naše zistenia ukazujú, že pri trénovaní cieľového modelu je dôležité zabezpečiť *dekoreláciu reprezentácií a rovnaké využitie všetkých dimenzií príznakov*. Takýto model je pomerne robustný a dostatočne citlivý na novosť vďaka tomu, že jeho reprezentácie reagujú veľkou zmenou aj na malú zmenu v stave. Zároveň samokontrolovaná regularizácia zabraňuje kolapsu motivačného signálu na nulu. Na základe našich výsledkov môžeme konštatovať, že metódy samokontrolovaného učenia sú určite perspektívne pri vytváraní detektorov novosti, ktoré možno úspešne použiť z hľadiska vnútornej motivácie a zlepšiť tak skúmanie prostredia.

Poďakovanie: Tento výskum bol podporený projektmi VEGA 1/0373/23 a KEGA 022UK-4/2023.

[1] https://paperswithcode.com/task/atari-games

Literatúra

Anand, A., Racah, E., Ozair, S., Bengio, Y., Côté, M. a Hjelm, R. D. (2019). Unsupervised state representation learning in Atari. *CoRR, abs/1906.08226*.

Bardes, A., Ponce, J. a LeCun, Y. (2022). VICReg: Variance-invariance-covariance regularization for self-supervised learning. V *International Conference on Learning Representations*.

Burda, Y., Edwards, H., Storkey, A. a Klimov, O. (2018). Exploration by random network distillation. *arXiv:1810.12894*.

Chopra, S., Hadsell, R. a LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. V *International Conference on Pattern Recognition*.

Oudeyer, P.-Y. a Kaplan, F. (2009). What is intrinsic motivation? a typology of computational approaches. *Frontiers in Neurorobotics*, 1:6.

Pecháč, M., Chovanec, M. a Farkaš, I. (2023). Exploration by self-supervised exploitation. *arXiv:2302.11563*.

Ryan, R. a Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1):54–67.

Co na srdci, to na jazyku?

Jan Pešán

Speech@FIT - Fakulta informačních technologií, Vysoké učení technické v Brně
Email: ipesan@fit.vutbr.cz

Vojtěch Juřík

Psychologický ústav, Filozofická fakulta Masarykovy univerzity, Brno
Email: jurik.vojtech@mail.muni.cz

Abstrakt

Identifikace kognitivního nebo fyzického přetížení je zásadní v řadě oblastí, kde lidské rozhodování představuje klíčový faktor týkající se ochrany zdraví a bezpečnosti osob či majetku. Specificky se jedná o oblasti, ve kterých se pohybují lidé jako piloti stíhacích letounů, řidiči autobusů, chirurgové či operátoři jaderných elektráren. Okamžitá stresová reakce na fyzické nebo kognitivní podněty může zásadně narušit fungování člověka a ztížit tak proces rozhodování v kritických situacích. Význam paralingvistického automatického zpracování řeči zde tudíž nelze opomíjet. Intenzita, úroveň vyčerpání, intonace a rytmus řeči jsou příklady paralingvistických vlastností mluveného projevu, které mohou nejen změnit význam komunikovaného obsahu, ale mohou také posloužit jako příznaky vhodné pro detekci rizikových situací. V rámci výzkumu je třeba věnovat těmto paralingvistickým vlastnostem pozornost a snažit se vyvinout nástroje pro jejich efektivní rozpoznávání. Nicméně, problém pro celou oblast stále představuje nedostatek vysoce kvalitních referenčních dat pro trénování detekčních systémů. Jako příspěvek k řešení této otázky jsme vyvinuli nástroj BESST pro generování potřebných řečových dat ve stresovém kontextu a s jeho pomocí jsme shromáždili reálná data mapující stresové projevy v lidské řeči. Tento příspěvek diskutuje možnosti a omezení vyplývající z navrženého nástroje, k čemuž využívá analýzy originálních naměřených dat.

1 Stres a řeč

Účinky lidského stresu na produkci řeči mohou být významné. Stres může způsobit změny v hlasu, jako je chrapot, třes a změny výšky a hlasitosti (Laukka a spol., 2008), (Zsiga, 2013). Tyto změny jsou způsobeny fyziologickou reakcí těla na stres, která může zahrnovat zvýšené napětí ve svalech používaných pro řeč. Když je tělo ve stresu, aktivuje se sympatický nervový systém (SNS), což může způsobit uvolnění adrenalinu a kortizolu (Dickerson a Kemeny, 2004). Adrenalin může zvýšit svalové napětí, zatímco kortizol může ovlivnit hlas tím, že způsobí vysychání hlasivek. To může vést ke změnám výšky a hlasitosti, stejně jako ke zvýšení hlasového úsilí, což může dále zhoršit napětí ve svalech používaných pro tvorbu řeči. Kromě změn v hlase může stres ovlivnit také kognitivní procesy, které se podílejí na tvorbě řeči. Pro tělo ve stresu, může být obtížné soustředit se, což může ovlivnit schopnost plánovat a organizovat řeč. To může mít za následek neorganizovanou nebo obtížně sledovatelnou řeč (Laures-Gore a spol., 2019). Stres může navíc způsobit změny v rychlosti a rytmu řeči. Lidé ve stresu často mluví rychleji nebo koktají, což může ostatním ztěžovat porozumění. To může vést k chybné komunikaci a dalšímu zvýšení úrovně stresu, což lze rovněž považovat za další faktory, na základě kterých je možné identifikovat stres u člověka.

1.1 BESST - Brno Extended Stress and Speech Test

Brno Extended Stress and Speech Test (BESST) je experimentální metodologie určená pro sběr dat, která jsou vhodná pro trénování strojového učení (ML) pro detekci stresu v řeči. Metoda je navržena tak, aby maximalizovala sběr řečových výstupů od účastníků umístěných v náročných stresových situacích. K tomuto účelu bylo navrženo experimentální prostředí, které tvoří záznam z několika kamer a mikrofonů v situacích, kdy účastníci plní řadu úkolů, které mají stresový charakter. V prvním úkolu je fyziologický stres vyvoláván ponořením nedominantní ruky účastníka do ledové vody, zatímco je zaznamenáván jeho volný projev řeči. Druhý stresový kontext je založen na duálním úkolu, který využívá tzv. Reading Span Task Daneman a Carpenter (1980), který zvyšuje kognitivní zátěž, zatímco účastník řeší a čte textový rébus.

Kromě audio a video záznamů jsou také pořizovány elektrokardiogram (EKG) pomocí zařízení Faros 180[1] a měření hladiny galvanického odporu kůže (GSR) pomocí náramku Empatica E4[2]. Data z obou senzorů nám pomáhají hodnotit objektivní úrovně fyziolo-

[1] https://www.bittium.com/medical/bittium-faros

[2] https://www.empatica.com/research/e4/

gického stresu. Subjektivní stavy stresu jsou účastníky hlášeny prostřednictvím Perceived Stress Scale 14 (PSS14; Cohen a spol., 1983), State Trait Anxiety Inventory Y1 (STAI-Y1; Hedberg, 1972) a NASA Task Load Experience (NASA TLX; Hart a Staveland, 1988), a jsou také hodnoceny na základě vlastních odhadů zátěže po každém úkolu. Celá metodologie je navržena tak, aby byla funkční a škálovatelná pro sběr klíčových dat nezbytných pro detekci stresu pomocí strojového učení a mohla tak být přenesena do jiných výzkumných prostředí nebo kulturních kontextů, čímž nabízí také možnost sběru dat na různých populacích.

1.2 Dosavadní závěry

V rámci studie jsme se zaměřili na vztah mezi srdeční činností a subjektivním/objektivním stresem. Zjistili jsme, že srdce reaguje relativně rychle na stresový podnět, ale jakmile tělo vyhodnotí míru ohrožení, adaptuje se velice rychle zpět (mezi 20-30 údery srdce), což komplikuje využití frekvence srdečního rytmu jako objektivní indikátor stresu. Tato komplikace má vliv na analýzu řeči - je potřeba zachytit akutní stresovou reakci zároveň s řečí (ale 20-30 úderů je pouze 15-30 sekund, což představuje pro potřebnou analýzu malý překryv).

Další výzvou se ukazuje správné zarovnání stresové reakce a řečových projevů v čase. I malý posun v časování může být problematický a vést k nesprávné interpretaci dat. Kromě toho je třeba zvážit, jaká granularita dat je nejvhodnější pro jejich anotaci - zda se jedná o celé experimenty, části experimentů nebo dokonce mikrosegmenty.

1.2.1 Synchronizace datových zdrojů

V experimentu BESST jsme použili několik nezávislých datových zdrojů; audiovizuální záznamy, EKG, GSR. Abychom mohli tato data správně analyzovat a interpretovat, je důležité zajistit, aby byla synchronizována s přesným časováním. Existuje několik přístupů, jak uvádíme v textu dále.

1.2.2 Explicitní synchronizace

Jednou z možností, jak zajistit synchronizaci, je použití tzv. explicitní synchronizace. To znamená, že v určitém okamžiku vytvoříme viditelnou a slyšitelnou zvukovou událost (např. tlesknutí), na základě kterého jsme schopni data synchronizovat, protože slouží jako referenční bod pro synchronizaci dat. Nicméně, u některých datových zdrojů, jako je EKG, je obtížné použít tuto metodu.

1.2.3 Implicitní synchronizace pomocí časových značek

Další možností je použití implicitní synchronizace pomocí časových značek. Každý datový zdroj může být označen časovou známkou, která určuje přesný čas, kdy byl zdroj zaznamenán. Tyto časové značky se následně používají k synchronizaci dat. Nicméně, časové značky mohou driftovat s ohledem na použitá zařízení, takže je důležité dbát na to, aby byly zařízení správně kalibrovány a aby nedošlo k rozdílnému driftu u různých zařízení.

1.2.4 Lab Streaming Layer

Další možností je použití nástroje Lab Streaming Layer (LSL)[3], což je open-source nástroj pro synchronizaci datových zdrojů. LSL umožňuje snadnou integraci různých datových zdrojů a poskytuje přesné časování a synchronizaci. Nicméně, použití LSL vyžaduje určitou konfiguraci a způsob použití, takže jeho nasazení není v některých kontextech vhodné.

2 Závěr

Výzkum stresu a jeho detekce pomocí řečových projevů a strojového učení přináší řadu výzev, a to jak na úrovni metodologické, tak na úrovni analýzy dat. Zároveň nabízí značný potenciál pro celou diskutovanou oblast, kdy se hledání cest k dosažení validních dat vhodných pro strojové učení stává středobodem celé problematiky. V tomto ohledu dále rozvíjíme metodologii BESST a shromažďujeme data o lidském řečovém projevu naměřená v rámci stresového kontextu. Aktuálně plánujeme uveřejnění prvního datastetu pro účely strojového učení a hledáme adekvátní způsoby jejich analýzy. Součástí plánované publikace dat je podrobný popis metodologie sběru dat, jejich podrobná charakteristika, navržení možného zpracování, popis limitů celého výzkumu a doporučení ve formě nabytých zkušeností.

Souhrnem, navrhovaná metodologie BESST poskytuje funkční a škálovatelnou základnu pro sběr klíčových dat nezbytných pro detekci stresu pomocí hlubokého učení, a to nejen v českém prostředí, ale i v jiných kulturních a jazykových kontextech.

Reference

Cohen, S., Kamarck, T. a Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24(4):385.

Daneman, M. a Carpenter, P. A. (1980). Individual di-

[3] https://labstreaminglayer.org/

fferences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4):450–466.

Dickerson, S. S. a Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130(3):355–391.

Hart, S. G. a Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. V *Advances in Psychology*, str. 139–183. Elsevier.

Hedberg, A. G. (1972). Review of state-trait anxiety inventory. *Professional Psychology*, 3(4):389–390.

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, O., Faria, V., Michelgard Palmquist, A., Appel, L., Fredrikson, M. a Furmark, T. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior*, 32:195–214.

Laures-Gore, J., Cahana-Amitay, D. a Buchanan, T. W. (2019). Diurnal cortisol dynamics, perceived stress, and language production in aphasia. *Journal of Speech, Language, and Hearing Research*, 62(5):1416–1426.

Zsiga, E. (2013). *The sounds of language : an introduction to phonetics and phonology*. Wiley-Blackwell, Chichester.

Vplyv subjektívnej vizuálnej senzitivity na percepciu času

Alexandra Ružičková (1), Vojtěch Juřík (1), Lenka Jurkovičová (2, 3) a Jan Páleník (1)

1 Psychologický ústav, Filozofická fakulta, Masarykova univerzita, Arna Nováka 1, Brno
2 CEITEC Masarykova Univerzita, Kamenice 5, Brno
3 1. neurologická klinika, Lékařská fakulta, Masarykova univerzita, Pekařská 53, Brno

alexandra.ruzickova@mail.muni.cz, jurik.vojtech@mail.muni.cz, jurkovicova.lenka@mail.muni.cz, jan.palenik@mail.muni.cz

Abstrakt

Subjektívne vnímaný čas môže byť skreslený vizuálnymi vlastnosťami podnetov, ktoré sledujeme. Predošlé výskumy dokázali, že nepríjemné podnety a podnety vyvolávajúce väčšiu neurálnu odpoveď spôsobujú expanziu subjektívne vnímaného času. Nejasné zostáva, či môže byť veľkosť tohto efektu rôzna s ohľadom na individuálnu senzorickú citlivosť pri vnímaní vizuálnych podnetov. Pre objasnenie navrhujeme inovatívny výskumný design dvoch experimentov vo virtuálnej realite (VR). Metodologické a inštrumentálne aspekty diskutované v tomto príspevku majú potenciál viesť k tvorbe nového behaviorálneho nástroja na testovanie vizuálnej senzitivity. Výstup výskumu môže pomôcť skvalitniť dizajnovanie virtuálnych a reálnych prostredí.

1 Úvod

Ľudský mozog nedisponuje špecializovaným centrom pre vnímanie času tak ako je to v prípade spracovania zmyslových vnemov (napr. Gruber a spol., 2018). Za jadrový komponent pri vnímaní časovosti sa však pokladá insulárny kortex (Mella a spol., 2019), rovnako zodpovedný za integráciu informácií o telesnom diskomforte a bolesti (Lu a spol., 2016). Pravdepodobne preto spôsobujú nepríjemné a bolestivé podnety expanziu vnímaného času (Rey a spol., 2017). Okrem toho môže byť subjektívne plynutie času na úrovni sekúnd skreslené pozornosťou, pamäťou alebo tiež zmyslovým vnímaním (Matthews a Meck, 2016). Fyzikálne vlastnosti podnetov, ako napríklad zvýšená hlasitosť a jas, podporujú význačnosť (z angl. *salience*) a tým predlžujú ich vnímané trvanie. Manipulovaním relatívnej význačnosti podnetov pozorujeme všeobecné skreslenia vo vnímaní času, no na individuálne rozdiely pri spracovaní význačnosti samotnej sa výskum dosiaľ nezameriaval.

2 Vnímanie času a vizuálna senzitivita

Relativita subjektívneho času býva v psychológii a neurovedách najčastejšie vysvetľovaná variovaním stupňa nabudenia (angl. *arousal*) a veľkosťou neurálnej odozvy príslušného mozgového centra na prezentovaný podnet či situáciu. Predmetom diskusií je, ktorý z týchto procesov má na skreslenie subjektívneho času významnejší dopad. Modernejšie pohľady smerujú k efektivite neurálneho kódovania informačného vstupu (Eagleman a Pariyadath, 2009). Zmyslové podnety aktivujúce obsiahlejšiu oblasť nervových sietí sú na úrovni subsekúnd vnímané ako dlhšie trvajúce (Cai a spol., 2015; Kruijne a spol., 2021; Matthews a Gheorghiu, 2016). Samotná veľkosť neurálnej odpovede na zmyslové podnety sa však medzi jednotlivcami líši (Ward, 2018).

Pri interakcii s okolitým svetom rozoznávame spektrum intenzít spracovania podnetov senzorickými (t. j. zmyslovými) systémami, súhrnne označované ako senzorická senzitivita (Aron a Aron, 1997; Robertson a Simmons, 2013). Pól hypersenzitivity sa vyznačuje nadmerne intenzívnymi reakciami napríklad na jasné svetlá, hlasné zvuky alebo výrazné pachy. Opakom je pocit nedostatočnej zmyslovej stimulácie u hyposenzitivity. Vyhranenosť na tomto spektre sa objavuje súčasne v oboch smeroch a naprieč viacerými zmyslami. Podmienka, pre ktorú je zvýšená senzorická senzitivita ústredná je autizmus. Autistickí jedinci vykazujú väčšiu neurálnu odozvu na senzorickú stimuláciu (Schwarzkopf a spol., 2014), ako aj distorzie vnímania plynutia času (Allman a Meck, 2012).

Rozsah aktivácie senzorických centier mozgu aký má určitý podnet potenciál vyvolať, tzv. kortikálna excitabilita, reflektuje senzorickú senzitivitu jedinca (Jurkovičová a spol., 2023; Ward, 2018; Wilkins a spol., 1984). Excitabilitu vizuálneho kortexu a vizuálnu senzorickú senzitivitu je v laboratórnych podmienkach možné študovať pomocou *Pattern Glare Testu* (PGT; Braithwaite a spol., 2013a; Braithwaite a spol., 2013b; Wilkins a spol., 1984). Podnetmi PGT sú rôzne priestorové frekvencie (*f*) čierno-bielych pruhovaných vzorov. Kritickým faktorom pre aktiváciu vizuálneho kortexu je potom miera ich deviácie (1/*f*) od štruktúr

vyskytujúcich sa v prírode (Penacchio a Wilkins, 2015).

Zrakový systém je optimalizovaný pre vnímanie prírodných prostredí, čo sa manifestuje neefektívnym neurálnym spracovaním neprirodzených štruktúr urbanistickej architektúry (Le a spol., 2017). Sledovanie urbanistických budov je oproti vystaveniu prírodným motívom vystriedané obsiahlejšou aktiváciou vizuálneho kortexu a výraznejším vizuálnym diskomfortom. Odchýlka od prírodnej štruktúry spôsobujúca takýto efekt vrcholí medzi 1 – 4 cyklami za stupeň (cpd; z angl. *cycles per degree;* Huang a spol., 2011; Juricevic a spol., 2010; O'Hare a Hibbard, 2011). V prípade PGT to je práve priestorová frekvencia 2 – 3 cpd, ktorá je považovaná za najviac averzívnu a sprevádzanú najmenej efektívnou neurálnou odpoveďou (Wilkins a spol., 1984).

Aaen-Stockdale a spol. (2011) skúmali pomocou rôznych priestorových frekvencií čierno-bieleho pruhovania vplyv novosti podnetu na skreslenie subjektívneho času. Ukázalo sa, že priestorová frekvencia 2 cpd spôsobuje expanziu subjektívneho času bez ohľadu na to či je prezentovaná v repetícii alebo ako ojedinelý podnet (obvykle vyvolávajúci väčšiu neurálnu odozvu). Iná štúdia dokladá, že zvýšená koncentrácia kyseliny gama-aminomaslovej (GABA) vo vizuálnom kortexe má za následok kontrakcie subjektívneho času (Terhune a spol., 2014). To je nosné v kombinácii so zisteniami Jurkovičovej a spol. (v revízii), že koncentrácia GABA v okcipitálnom kortexe negatívne koreluje s počtom reportovaných vizuálnych distorzií pri PGT. Vyššia miera subjektívnych zrakových distorzií odráža hyperexcitabilitu vizuálneho kortexu (Fong a spol., 2019).

3 Výskumné ciele

Prehľad literatúry podnecuje úvahy o existencii rozdielov vo vnímaní krátkych časových intervalov (v rozpätí milisekúnd až sekúnd) na základe miery citlivosti vizuálneho systému jednotlivcov. V reakcii na to navrhujeme výskumný podklad dvoma paralelnými experimentmi vo VR.

Experiment 1 bude založený na vizuálnej stimulácii modifikovaným PGT (Braithwaite a spol., 2015; Fong a spol., 2019; Wilkins a spol., 1984). Naše hypotézy sú:

(E1H1) Averzívna priestorová frekvencia 3 cpd bude spôsobovať väčšiu expanziu subjektívneho času ako vysoká (14 cpd) a nízka (0,5 cpd) priestorová frekvencia.

(E1H2) Subjektívne senzoricky senzitívnejší ľudia budú vnímať väčšiu dilatáciu trvania averzívnej priestorovej frekvencie ako tí menej senzitívni.

(E1H3) Reportovaná kortikálna hyperexcitabilita bude pozitívne korelovať s dĺžkou subjektívneho času pri pozorovaní averzívnej priestorovej frekvencie.

V Experimente 2 budú podnetmi komplexné virtuálne prostredia – statické urbanistické alebo prírodné exteriéry. Hypotézy sú:

(E2H1) Vizuálne diskomfortná urbanistická scéna predĺži subjektívny čas významnejšie ako vizuálne komfortná prírodná scéna.

(E2H2) Vyššia subjektívna senzorická senzitivita bude spojená s najväčšou expanziou subjektívneho času v prípade sledovania diskomfortnej urbanistickej scény.

(E2H3) Subjektívne trvanie vizuálne diskomfortnej urbanistickej scény bude u jedincov reportujúcich vyššiu hyperexcitabilitu kortexu najdlhšie.

Dáta z oboch experimentov budú spracované samostatne, ale výskum na jednej vzorke nám umožní ich medzi sebou porovnať a preskúmať efekt komplexnosti podnetu. Okrem toho budeme v oboch experimentoch kontrolovať vplyv ďalších premenných, ktoré sú v úzkom vzťahu s tými študovanými – autistické črty (Haigh, 2018), spánková deprivácia (Meisel a spol., 2015), cybersickness (O'Hare a spol., 2018), biologické pohlavie a deň menštruačného cyklu (Jurkovičová a spol., v revízii; Rudroff a spol., 2020).

4 Metódy

4.1 Participanti

Výskumná vzorka bude pozostávať z neurotypických mladých dospelých (18 – 30 rokov) s normálnym alebo korigovaným videním, bez migrén alebo epileptických záchvatov v anamnéze. Pri odhadovanej veľkosti vzťahu premenných $\rho = 0,4$, požadovanej sile testu 0,9 a hladine významnosti 0,01 potrebujeme na identifikáciu signifikantného trendu približne 75 participantov. Tých oslovíme prevažne prostredníctvom sociálnych sietí a ich účasť bude dobrovoľná. Participantov tiež požiadame, aby aspoň 4 hodiny pred účasťou nepili kofeínové nápoje. Výskum bol schválený Etickým panelom Psychologického ústavu Filozofickej fakulty Masarykovej univerzity.

4.2 Materiály a podnety

4.2.1 Dotazníky

Okrem otázok na biologické pohlavie, deň menštruačného cyklu a počet hodín spánku za poslednú noc budú participanti vypĺňať sériu validovaných dotazníkov zameraných na: senzorickú senzitivitu – *Glasgow Sensory Questionnaire* (GSQ; Robertson a Simmons, 2013), autistické črty – *Autism-Spectrum Quotient Short* (AQ-Short; Hoekstra a spol., 2011), kortikálnu hyperexcitabilitu – *Cortical Hyperexcitability index II* (Chi – II; Fong a spol., 2019) a cybersickness – *Simulator Sickness Questionnaire* (SSQ; Kennedy a spol., 2009) administrovaný štyrikrát.

4.2.2 Temporálna bisekcia

Úloha temporálnej bisekcie (angl. *Temporal Bisection Task*; TBT; Allan a Gibbon, 1991; Church a Deluty, 1977; Wearden, 1991; Wearden a Ferrara, 1995) je jedným z mnohých prístupov k štúdovaniu vnímania krátkych časových intervalov. V tréningovej fáze sú participanti požiadaní, aby sa naučili rozlišovať dve referenčné trvania – „Krátke“ (400 ms) a „Dlhé“ (1600 ms). V Experimente 1 budú po tieto doby prezentované čierne elipsy a v Experimente 2 tréningové virtuálne prostredie. Každý z dvoch intervalov prezentovaný 10 krát (20 prezentácií celkovo) sa objaví v pseudo-náhodnom poradí. Prvým desiatim prezentáciám bude predchádzať informácia o tom, ktoré z dvoch referenčných trvaní to bude. Ďalších desať prezentácií budú participanti kategorizovať sami. Po každom zatriedení intervalu obdržia spätnú väzbu. Odpoveď bude následovať tzv. medzistimulový interval (z ang. *inter-stimulus interval*; ISI). Zrakový fixačný bod v podobe bieleho štvorca bude miznúť iba pri zobrazení odpoveďového rozhrania.

Počas testovacej fázy sa k referenčným trvaniam pridá päť ďalších medziľahlých intervalov (400, 600, 800, 1000, 1200, 1400, 1600 ms). Ako podnety sa v Experimente 1 použijú pruhované disky modifikovaného PGT (Braithwaite et al., 2015; Fong et al., 2019; Wilkins et al., 1984). Priestorové frekvencie achromatických pruhov vypĺňajúcich elipsoidy zodpovedajú 0,5 cpd (nízka frekvencia pruhovania), 3 cpd (kritická, averzívna priestorová frekvencia) a 14 cpd (vysoká frekvencia), keď je pozorovacia vzdialenosť 80 cm. Tie naše budú proporčne korešpondovať so vzdialenosťou 2,5 metra od virtuálnej steny, teda so štandardnou vzdialenosťou pre celý experimentálny beh. V Experimente 2 budú podnetmi štyri virtuálne exteriéry – 2 urbanistické a 2 prírodné – vybrané participantmi prípravnej štúdie. Z oboch kategórií zahrnieme vždy tie hodnotené ako vizuálne najviac a najmenej komfortné.

Podnety testovacej fázy budú prezentované každý osemkrát v pseudo-náhodnom poradí, čo je celkovo 392 prezentácií (E1: 7 trvaní x 3 podnety x 8 prezentácií; E2: 7 trvaní x 4 podnety x 8 prezentácií). Participanti budú podnety kategorizovať ako skôr “Krátky” alebo skôr “Dlhý”, tentokrát bez údaju o správnosti. Odpoveď bude vyžadovaná okamžite po zmiznutí podnetu a ISI bude opäť 1000 ms.

4.3 Procedúra

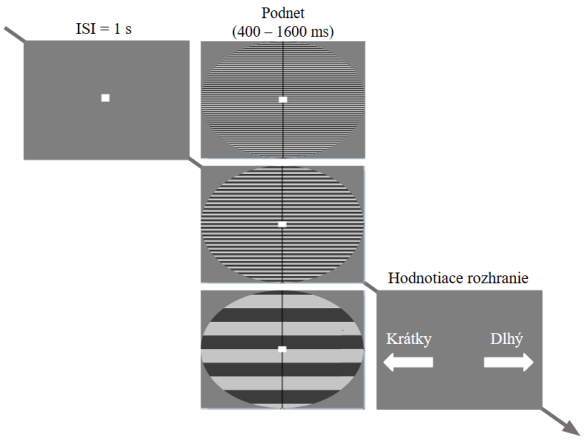
Úvodom participantom zhrnieme priebeh experimentálneho behu a upozorníme na riziko cybersickness spojené s využívaním náhlavných dispejov (z angl. *head-mounted displays*; HMDs) pre naše výskumné účely (O’Hare a spol., 2018). Zdôrazníme dobrovoľnosť účasti a možnosť odstúpenia. Následne participanti podpíšu informovaný súhlas.

Expozícii vo VR bude predchádzať zodpovedanie otázok na biologické pohlavie, deň menštruačného cyklu, počet hodín spánku, počiatočnej úrovne cybersickness (SSQ1) a test FrACT (Bach, 1996; 2006) ohľadom normality vnímania kontrastu. Počas expozície budú participanti sedieť na stoličke za stolom s klávesnicou, pomocou ktorej budú značiť svoje odpovede. Medzi oboma experimentami bude prestávka určená na zmiernenie možnej cybersickness. Tento čas využijeme na administrovanie dotazníkov v poradí: SSQ2, AQ-Short, Chi – II, SSQ3. Následne participanti podstúpia druhý experiment vo VR. Na konci oboch experimentov vyzveme participantov k jednorázovému ohodnoteniu každého zo stimulov na škále vizuálneho komfortu: -5 („Veľmi diskomfortný“), 0 („Ani komfortný ani diskomfortný“), 5 („Veľmi komfortný“).

Na záver prebehne administrácia dotazníkov SSQ4 a GSQ a debriefing. V rámci debriefingu nás bude tiež zaujímať vnímanie paobrazov a ich možný prekryv s ďalšími podnetmi. Celá účasť zaberie zhruba hodinu a pol.

4.3.1 Experiment 1

Participanti sa budú na prostredie VR adaptovať vo virtuálnej obývacej izbe. V žiadnej fáze experimentálneho behu nebudú môcť s VR interagovať. Experiment 1 bude prebiehať v prázdnej, tmavej miestnosti so sivými stenami, v ktorej bude participant sedieť. Inštrukcie a podnety (viz Obr. 1) mu odprezentujeme na virtuálnej stene vzdialenej zhruba 2,5 metra pred ním.



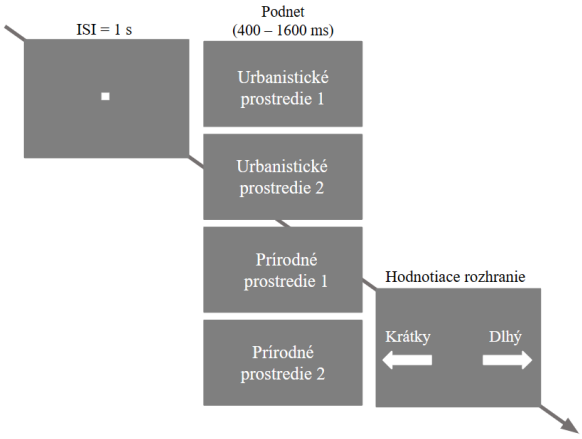
Obr. 1: Postupnosť priebehu jedného testovacieho úseku Experimentu 1.

4.3.2 Experiment 2

Pre účely Experimentu 2 zrealizujeme prípravnú štúdiu. Jej cieľom je výber predlôh pre vymodelovanie piatich virtuálnych prostredí (1 tréningové a 4 testovacie). Fotografie všeobecne neznámych urbanistických (N=25) a prírodných (N=25) prostredí budú pochádzať

z rôznych častí sveta. Pôjde o náhľady miest Google Maps. Kritériami selekcie stanovujeme najväčší vizuálny komfort prírodnej a urbanistickej scény a najväčší vizuálny diskomfort scén oboch kategórií. Tréningové prostredie získa indiferentné hodnotenie.

Adaptácia na prostredie VR bude v Experimente 2 umožnená zasadením do tréningového prostredia. Toto prostredie ďalej poslúži na memorovanie referenčných intervalov. ISI a všetky hodnotenia budú v rámci druhého experimentu prebiehať v rovnakej virtuálnej miestnosti a rovnakým spôsobom ako v Experimente 1. Priebeh jedného experimentálneho úseku je zhrnutý Obrázkom 2.



Obr. 2: Postupnosť priebehu jedného testovacieho úseku Experimentu 2.

4.4 Analýza dát

Priama reprezentácia subjektívneho času sa v TBT určuje lokalizáciou bodu temporálnej indiferencie (z angl. *temporal indifference point*; TIP). Je to trvanie podnetu, ktoré je s rovnakou pravdepodobnosťou kategorizované ako „Krátke“ alebo „Dlhé“. TIP sa počíta individuálne pre každého participanta a pre každý druh podnetu. Všetky odpovede sú prevedené na proporcie dlhých intervalov a zakreslené do grafu ako psychometrická funkcia o siedmich bodoch. Posunutie psychometrickej funkcie oproti objektívnym dátam, tzn. skutočným trvaniam podnetov indikuje skreslenie vnímania času. Expanziu subjektívneho času reprezentuje ľavostranné posunutie psychometrickej funkcie.

Okrem výpočtu TIP dokážeme pomocou Weberovej frakcie určiť mieru citlivosti s akou participanti čas pri TBT vnímali. Weberova frakcia je smerodajná odchýlka vydelená hodnotou TIP. Rozdiely medzi proporciami odpovedí „Dlhý“ u menej a viac vizuálne komfortných podnetov sú zachytávané indexom zvaným *d*-skóre. Pozitívna hodnota *d* vypovedá o predĺžení vnímaného trvania vizuálne diskomfortného podnetu, kým negatívne *d*-skóre svedčí o kontrakcii subjektívneho času.

Analýzou proporcií odpovedí „Dlhý“ pre trvanie každého typu podnetu porovnáme ako sa líšilo vnímané trvanie týchto podnetov navzájom. Následne budeme skúmať vzťahy medzi TIP, WF, *d*-skóre a skóre z dotazníkov na senzorickú senzitivitu a kortikálnu hyperexcitabilitu. Tieto dotazníky by mali korelovať s vizuálnym komfortom reportovaným v závere oboch experimentov pre každý typ podnetu. Vzťah očakávame za predpokladu kontroly vplyvu biologického pohlavia a počtu hodín spánku. Skreslenie vnímania času zvážime aj z hľadiska možného efektu cybersickness. Dotazník na autistické črty a test FrACT poslúžia na vyradenie jedincov sporne spĺňajúcich kritériá zahrnutia do datasetu.

Finálne výsledky osobitných analýz dát z oboch experimentov budú môcť byť porovnané medzi sebou. Predbežne predpokladáme ich ekvivalenciu. Dáta budeme analyzovať s využitím najnovšej verzie programu R.

5 Záver

Špecifiká jednotlivca môžu mať ďalekosiahly dopad na jeho fungovanie vo svete. Senzorická senzitivita spôsobuje alterácie zmyslového vnímania premosťujúce do behaviorálnych prejavov (Robertson a Simmons, 2013). Tie dokresľujú naše predstavy o takých podmienkach akou je napríklad autistické spektrum (Haigh, 2018; Robertson a Simmons, 2013). Samotné prežívanie je však vždy podkladom akéhokoľvek správania. Výskum navrhnutý v tomto príspevku je preto zameraný na rozšírenie poznatkovej základne o aspektoch prežívania sveta senzoricky senzitívnymi ľuďmi.

Popisovaný výskumný dizajn poskytne dátové výstupy s potenciálom objasniť ako môže senzorická senzitivita súvisieť s vnímaním času. Dva experimenty túto otázku ďalej rozlíšia podľa úrovne komplexnosti zmyslového podnetu. V prípade že by sa po zrealizovaní výskumu naše hypotézy potvrdili, je možné jeho implikácie ďalej rozvíjať. Ak sa ukáže, že individuálne rozdiely v subjektívnom čase odkazujú k výške senzorickej senzitivity, zamýšľame v budúcnosti túto logiku otočiť a testovať mieru vizuálnej senzorickej senzitivity rýdzo behaviorálnym testom. Nasledovalo by teda štúdium jemnejších nuancií vizuálneho komfortu skresľujúceho subjektívny čas.

Inou oblasťou na ktorú by malo potvrdenie hypotéz dopad je dizajnovanie virtuálnych a skutočných prostredí. Vizuálny komfort priestorových usporiadaní by totiž dostal nový rozmer – časový. Nakoniec je vizuálny diskomfort aj najväčším limitom tohto výskumného návrhu. Hrozí, že spôsobí nadmernú experimentálnu mortalitu a tým zabráni dosiahnutiu stanovených výskumných cieľov. Napriek tomu sa však domnievame, že nami navrhnuté postupy sú realizovateľné a vedúce k experimentálnemu uchopeniu psychologickej relativity času.

Poďakovanie

Špeciálne poďakovanie za rady ohľadom analýzy dát patrí Edite Chvojkovej. Za pripomienky súvisiace s realizáciou experimentov vo VR ďakujeme Pavlovi Ugwitzovi.

Literatúra

Aaen-Stockdale, C., Hotchkiss, J., Heron, J., & Whitaker, D. (2011). Perceived time is spatial frequency dependent. *Vision Research*, *51*(11), 1232–1238.

Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. Learning and motivation, 22(1-2), 39-58.

Allman, M. J., & Meck, W. H. (2012). Pathophysiological distortions in time perception and timed performance. *Brain*, *135*(3), 656–677.

Aron, E. N., & Aron, A. (1997). Sensory-Processing Sensitivity and Its Relation to Introversion and Emotionality. *Journal of Personality and Social Psychology*, *73*(2), 345–368.

Bach, M. (1996). The Freiburg Visual Acuity Test—Automatic Measurement of Visual Acuity. *Optometry and Vision Science*, *73*(1), 49.

Bach, M. (2006). The Freiburg Visual Acuity Test-Variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *245*(7), 965–971.

Braithwaite, J. J., Broglia, E., Bagshaw, A. P., & Wilkins, A. J. (2013a). Evidence for elevated cortical hyperexcitability and its association with out-of-body experiences in the non-clinical population: New findings from a pattern-glare task. *Cortex*, *49*(3), 793–805.

Braithwaite, J. J., Broglia, E., Brincat, O., Stapley, L., Wilkins, A. J., & Takahashi, C. (2013b). Signs of increased cortical hyperexcitability selectively associated with spontaneous anomalous bodily experiences in a nonclinical population. *Cognitive neuropsychiatry*, 18(6), 549-573.

Braithwaite, J. J., Mevorach, C., & Takahashi, C. (2015). Stimulating the aberrant brain: Evidence for increased cortical hyperexcitability from a transcranial direct current stimulation (tDCS) study of individuals predisposed to anomalous perceptions. *Cortex*, *69*, 1–13.

Cai, M. B., Eagleman, D. M., & Ma, W. J. (2015). Perceived duration is reduced by repetition but not by high-level expectation. *Journal of Vision*, *15*(13), 19.

Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. Journal of Experimental Psychology: Animal Behavior Processes, 3(3), 216–228.

Eagleman, D. M., & Pariyadath, V. (2009). Is subjective duration a signature of coding efficiency? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1841–1851.

Fong, C. Y., Takahashi, C., & Braithwaite, J. J. (2019). Evidence for distinct clusters of diverse anomalous experiences and their selective association with signs of elevated cortical hyperexcitability. *Consciousness and Cognition*, *71*, 1–17.

Gruber, R. P., Smith, R. P., & Block, R. A. (2018). The Illusory Flow and Passage of Time within Consciousness: A Multidisciplinary Analysis. *Timing & Time Perception*, *6*(2), 125–153.

Haigh, S. M. (2018). Variable sensory perception in autism. *European Journal of Neuroscience*, *47*(6), 602–609.

Hoekstra, R. A., Vinkhuyzen, A. A. E., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron-Cohen, S., Posthuma, D., & van der Sluis, S. (2011). The Construction and Validation of an Abridged Version of the Autism-Spectrum Quotient (AQ-Short). *Journal of Autism and Developmental Disorders*, *41*(5), 589–596.

Huang, J., Zong, X., Wilkins, A., Jenkins, B., Bozoki, A., & Cao, Y. (2011). FMRI evidence that precision ophthalmic tints reduce cortical hyperactivation in migraine. *Cephalalgia*, *31*(8), 925–936.

Juricevic, I., Land, L., Wilkins, A., & Webster, M. A. (2010). Visual Discomfort and Natural Image Statistics. *Perception*, *39*(7), 884–899.

Jurkovičová, L., Sakálošová, L., Kudlička, P., Páleník, J., Ružičková, A., Juřík, V., Mareček, R., Roman, R., Braithwaite, J. J., Sandberg, K., Near, J. P., Brázdil, M. (v revízii). *Resting GABA and glutamate concentrations in primary visual cortex and right anterior insula predict subjective visual sensitivity* [Manuskript odoslaný na publikovanie].

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (2009). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The international journal of aviation psychology*, 3(3), 203-220.

Kruijne, W., Olivers, C. N. L., & Rijn, H. van. (2021).

Neural Repetition Suppression Modulates Time Perception: Evidence From Electrophysiology and Pupillometry. *Journal of Cognitive Neuroscience*, *33*(7), 1230–1252.

Le, A. T. D., Payne, J., Clarke, C., Kelly, M. A., Prudenziati, F., Armsby, E., Penacchio, O., & Wilkins, A. J. (2017). Discomfort from urban scenes: Metabolic consequences. *Landscape and Urban Planning*, *160*, 61–68.

Lu, C., Yang, T., Zhao, H., Zhang, M., Meng, F., Fu, H., Xie, Y., & Xu, H. (2016). Insular Cortex is Critical for the Perception, Modulation, and Chronification of Pain. *Neuroscience Bulletin*, *32*(2), 191–201.

Matthews, W. J., & Gheorghiu, A. I. (2016). Repetition, expectation, and the perception of time. *Current Opinion in Behavioral Sciences*, *8*, 110–116.

Matthews, W. J., & Meck, W. H. (2016). Temporal cognition: Connecting subjective time to perception, attention, and memory. *Psychological Bulletin*, *142*(8), 865.

Meisel, C., Schulze-Bonhage, A., Freestone, D., Cook, M. J., Achermann, P., & Plenz, D. (2015). Intrinsic excitability measures track antiepileptic drug action and uncover increasing/decreasing excitability over the wake/sleep cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), 14694–14699.

Mella, N., Bourgeois, A., Perren, F., Viaccoz, A., Kliegel, M., & Picard, F. (2019). Does the insula contribute to emotion-related distortion of time? A neuropsychological approach. *Human Brain Mapping*, *40*(5), 1470–1479.

O'Hare, L., & Hibbard, P. B. (2011). Spatial frequency and visual discomfort. *Vision Research*, *51*(15), 1767–1777.

O'Hare, L., Sharp, A., Dickinson, P., Richardson, G., & Shearer, J. (2018). Investigating Head Movements Induced by 'Riloid' Patterns in Migraine and Control Groups Using a Virtual Reality Display. *Multisensory Research*, *31*(8), 753–777.

Penacchio, O., & Wilkins, A. J. (2015). Visual discomfort and the spatial distribution of Fourier energy. *Vision Research*, *108*, 1–7.

Rey, A. E., Michael, G. A., Dondas, C., Thar, M., Garcia-Larrea, L., & Mazza, S. (2017). Pain dilates time perception. *Scientific Reports*, *7*(1), 1–6.

Robertson, A. E., & Simmons, D. R. (2013). The Relationship between Sensory Sensitivity and Autistic Traits in the General Population. *Journal of Autism and Developmental Disorders*, *43*(4), 775–784.

Rudroff, T., Workman, C. D., Fietsam, A. C., & Kamholz, J. (2020). Response Variability in Transcranial Direct Current Stimulation: Why Sex Matters. *Frontiers in Psychiatry*, *11*, 585.

Schwarzkopf, D. S., Anderson, E. J., de Haas, B., White, S. J., & Rees, G. (2014). Larger Extrastriate Population Receptive Fields in Autism Spectrum Disorders. *Journal of Neuroscience*, *34*(7), 2713–2724.

Terhune, D. B., Russo, S., Near, J., Stagg, C. J., & Kadosh, R. C. (2014). GABA Predicts Time Perception. *Journal of Neuroscience*, *34*(12), 4364–4370.

Ward, J. (2018). Individual differences in sensory sensitivity: A synthesizing framework and evidence from normal variation and developmental conditions. *Cognitive neuroscience*, 10(3), 139-157.

Wearden, J. H. (1991). Human performance on an analogue of an interval bisection task. The Quarterly Journal of Experimental Psychology Section B, 43(1b), 59-81.

Wearden, J. H., & Ferrara, A. (1995). Stimulus spacing effects in temporal bisection by humans. The Quarterly Journal of Experimental Psychology, 48(4), 289-310.

Wilkins, A., Nimmo-smith, I., Tait, A., Mcmanus, C., Sala, S. D., Tilley, A., Arnold, K., Barrie, M., & Scott, S. (1984). A neurological basis for visual discomfort. *Brain*, *107*(4), 989–1017.

Využitie samoorganizácie v čiastočne riadenom učení hlbokých neurónových sietí

Sabína Samporová a Kristína Malinovská
Katedra aplikovanej informatiky, FMFI,
Univerzita Komenského v Bratislave
Mlynská dolina, 84248 Bratislava
Email: {samporova1, rebrova1}@uniba.sk

Abstrakt

Kvalitný a rozsiahly súbor dát je pri trénovaní hlbokých neurónových sietí veľmi dôležitý. Vytvorenie takéhoto súboru dát s označeniami je ale veľmi náročné, preto sa hľadajú modely na využitie neoznačených dát, popri trénovaní na označených dátach. Čiastočne riadené učenie sa zaoberá skúmaním týchto modelov. Zamerali sme sa na model Mean Teacher a jemu podobné, založené na dvoch hlbokých sieťach, ktoré dostanú rovnaký vstup s rôznymi augmentáciami. Cieľom Mean Teacher modelov je trénovanie na označených dátach a zároveň snaha o konzistenciu označovania medzi odpoveďami na označené aj neoznačené dáta. Spravidla sa na určenie chyby konzistencie používa štvorec chýb označení. Náš prístup je využiť samoorganizáciu neurálnych reprezentácií na poslednej konvolučnej vrstve, čím by sa mohla zlepšiť úspešnosť modelu, ako aj vhľad do modelu v zmysle vysvetliteľnej umelej inteligencie.

1 Úvod

Hlboké neurónové siete sú v súčasnosti pravdepodobne najpoužívanejšími a najskúmanejšími modelmi v strojovom učení s aplikáciami v mnohých rôznych oblastiach. Trénovanie takýchto modelov si vyžaduje veľké množstvo adekvátne označených trénovacích dát, no zvyčajne je dobre označených dát z reálneho sveta málo. Paradigma čiastočne riadeného učenia (semi-supervised learning) rieši tento problém prostredníctvom rôznych techník. Mnohé z nich používajú pri učení odchýlku medzi výstupmi siete pre dve rôzne augmentácie toho istého vstupu, čo sa nazýva regularizácia na základe konzistencie (consistency regularization). Jedným zo známych predstaviteľov tohto druhu čiastočne riadeného učenia je model Mean Teacher model (MT), ktorý navrhli Tarvainen a Valpola (2017). V tomto príspevku predstavujeme návrh na adaptáciu tohto modelu s použitím samoorganizácie.

2 Mean teacher model

Mean Teacher model (Tarvainen a Valpola, 2017) pozostáva z dvoch hlbokých sietí s rovnakou architektúrou, ale samostatnými trénovateľnými parametrami, teda váhami. Prvá z nich je nazývaná študent a označujeme ju ako θ a druhá učiteľ θ'. MT model je vhodný na riešenie problému klasifikácie. Na trénovanie model využíva nie len označené, ale aj neoznačené dáta, teda také, ktoré nemajú priradenú príslušnosť do niektorej z tried.

Dôležitým komponentom modelu je použitie takzvaných augmentácii vstupných obrázkov ako je napríklad náhodná translácia či rotácia a rôzne druhy slabého šumu (Gaussovský, zaušumenie farieb, atď.). Technika augmentácie dát je bežne používaná a spravidla vylepšuje aj klasické učenie s učiteľom (Krizhevsky a spol., 2012). Pri čiastočne riadenom učení zohráva dôležitú úlohu, ale na samotnú kompenzáciu absencie označení nestačí.

MT na trénovanie študentského modelu používa 2 druhy stratových funkcií (loss functions), Supervised loss $S(\theta)$ definovanú v (1) a takzvanú Consistency loss $J(\theta)$ definovanú v (2). $S(\theta)$ môžeme určiť len pre dáta, ktoré majú označenie a je to krížová entropia (cross entropy) predikcie siete pre vstup x_j, na ktorý aplikujeme augmentáciu η a príslušného označenia y_j.

$$S(\theta) = \frac{1}{m} \sum_{j}^{m} \left[-\log \; P_f(y_j | x_j; \theta, \eta) \right], \qquad (1)$$

Ako $J(\theta)$, čiže chyba konzistencie, je použitá stredná kvadratická chyba (MSE) predikcie študentského a učiteľského modelu, ktoré sa trénujú súčasne pre rovnaký vstup, na ktorý aplikujeme rôzne augmentácie η a η'. Tento komponent chybovej funkcie nevyžaduje pre svoj výpočet žiadne označenia y čiže hovoríme o učení bez učiteľa a môžeme ho aplikovať aj na dáta bez označení.

$$J(\theta) = \frac{1}{n} \sum_{i}^{n} \left\| f(x_i, \theta', \eta') - f(x_i, \theta, \eta) \right\|^2, \qquad (2)$$

Celková hodnota stratovej funkcie, ktorá sa použije pri učení je potom vyjadrená ako vážený súčet $S(\theta)$ a $J(\theta)$

$$Loss(\theta) = S(\theta) + w_t J(\theta), \tag{3}$$

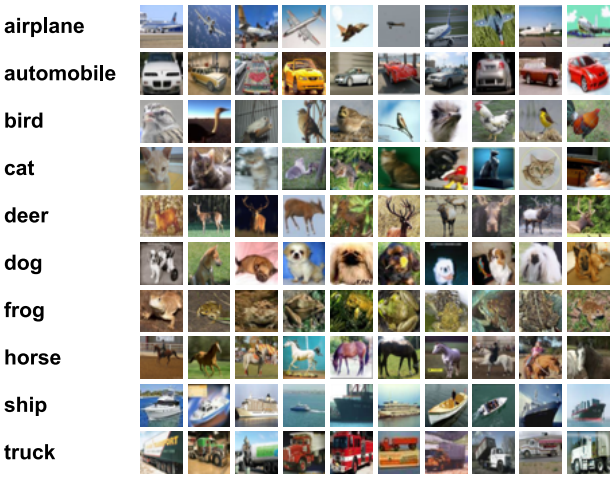
kde w_t je váha chyby konzistencie, ktorá je na začiatku pomerne malá ale v priebehu trénovania rastie. Na základe stratovej funkcie je študentský model θ je trénovaný spätným šírením chyby metódou najmenšieho gradientu.

Učiteľský model nemá rovnaký spôsob trénovania, ale funguje ako exponenciálny kĺzavý priemer (exponential moving average, EMA) študentského modelu. Táto stratégia sa nazýva Temporal Ensembling (Laine a Aila, 2016) čiže skladanie modelu v čase a predstavuje akýsi ďalší druh regularizácie. Úprava váh modelu učiteľa je vykonaná pomocou pravidla (4), kde θ'_t je označenie pre váhy učiteľského modelu v čase t, a θ_t sú váhy študenta v čase t a α je hyperparameter - rýchlosť učenia.

$$\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t \tag{4}$$

3 Dataset a baseline

Ako dataset sme zvolili štandardný CIFAR10 dataset (Krizhevsky a spol., 2009), ktorý pozostáva zo 60 tisíc farebných obrázkov s rozmermi 32×32 pixelov. Obrázky sú označené a sú z 10 rôznych tried (lietadlá, autá, vtáky, mačky, jelene, psy, žaby, kone, lode, nákladné autá) ilustrovaných na Obr. 1.



Obr. 1: CIFAR10: desať náhodne zvolených obrázkov[1].

Pri vývoji modelu sme vychádzali priamo zo zverejneného modelu a kódu od autorov článku, Tarvainena a Vapolu[2], kde sú zverejnené aj optimálne parametre pre klasifikáciu datasetu CIFAR10, ktorý dosahuje presnosť $93.72\% \pm 0.15$. Tento výsledok budeme brať ako základ pre porovnanie úspešnosti nášho modelu a so zapojením princípu samoorganizácie očakávame zlepšenie úspešnosti v zmysle presnosti klasifikácie.

4 Využitie samoorganizácie: náš model

Pri tvorbe nášho modelu sme uvažovali, ako formulovať chybu konzistencie, ktorá by zahŕňala viac informácie ako v doterajšom modeli. Keďže chyba konzistencie v MT je vyjadrená vzdialenosť vektorov, s rozmerom rovným počtu tried, ide o pomerne málo informácie. Tuna a spol.(2021) vo svojom modeli Binary Mean Teacher modelovali binárnu klasifikáciu prítomnosti objektu na obrázku. V svojom modeli teda nemohli vyjadriť chybu konzistencie pri odozve z intervalu $(0, 1)$ pre dve triedy (oproti 10 triedam v CIFAR10) a teda vyjadrili $J(\theta)$ ako MSE na poslednej konvolučnej vrstve použitej architektúry, čiže ešte pred plne prepojenou časťou siete.[3]

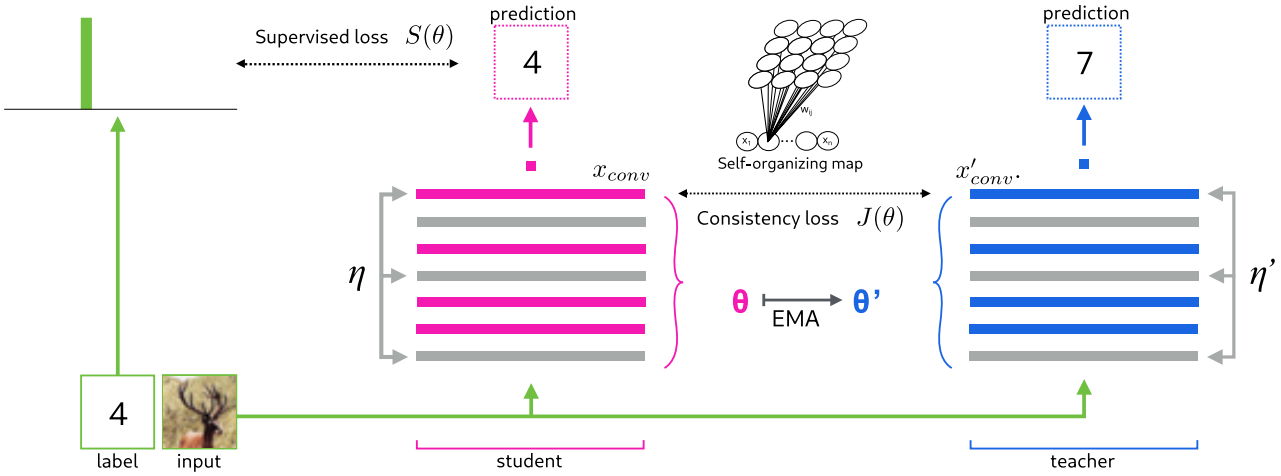
Túto reprezentáciu využijeme aj my a nazývame x_{conv}. Predpokladáme, že reprezentácia vstupu po prechode všetkými konvolučnými vrstvami, bude vhodnejšia pre zachytenie rozdielov a podobností medzi obsahom obrazu, než jeho samotná klasifikácia. Náš navrhovaný model ilustrujeme na Obr. 2. Ďalej navrhujeme, že pridanie konceptu samoorganizácie môže zlepšiť trénovanie tým, že bude lepšie odzrkadľovať vzdialenosť dát, než pôvodný spôsob počítania $J(\theta)$. Samoorganizáciu zapojíme do počítania $J(\theta)$ tak, že využijeme samoorganizujúcu sa mapu (SOM) (Kohonen, 1990), ktorú natrénujeme na rozpoznávanie reprezentácií x_{conv} zo študentského aj učiteľského modelu počas danej epochy, keď sa model učí. Potom túto SOM použijeme na počítanie Consistency loss $J(\theta)$.

Samoorganizujúce sa mapy majú vstupnú vrstvu veľkosti vstupného vektora plne prepojenú s nasledujúcou vrstvou, ktorú tvorí mapa neurónov konkrétneho tvaru, napríklad mriežka. V každom neuróne máme parameter siete, vektor veľkosti vstupu, ktorý reprezentuje niečo ako prototyp. Vyjadrením euklidovskej vzdialenosti medzi vstupným neurónov a váhami neurónov vyjadríme víťaza v zmysle najbližšieho neurónu, ktorého váhu potom pri učení posilníme daným vstupom a zároveň posilníme aj jeho okolie. Výsledkom učenia sa SOM je topografická mapa vstupného priestoru, ktorá zachytáva štruktúru a podobnosti v dátach nelineárnym spôsobom.

Označme reprezentácie z poslednej konvolučnej vrstvy študentského a učiteľského modelu ako x_{conv} a x'_{conv}. V SOM nájdeme víťazný neurón (winner neuron, best matching unit) pre x_{conv} a x'_{conv}, označme c a $c\prime$. Keďže neuróny SOM sú definované ich váhami a

[1] https://www.cs.toronto.edu/~kriz/cifar.html
[2] github.com/CuriousAI/mean-teacher/

[3] Podobná stratégia sa bežne používa pri transfer learning (Weiss a spol., 2016), kde sa použije model, ktorý už je dobre natrénovaný na všeobecnej úlohe a jeho plne prepojená vrstva alebo vrstvy sa nahradia novou architektúrou, ktorá modeluje inú úlohu.



Obr. 2: MT model

tie vlastne reprezentujú polohu tohto neurónu v mnohorozmernom priestore. Potom našim cieľom je, aby c a $c\prime$, predstavujúce vstupné reprezentácie, ktoré vznikli z rovnakého obrázku, boli v tomto priestore čo najbližšie. Preto našu novú chybu konzistencie definujeme ako súčet vzdialenosti víťazov a tiež vzdialenosti víťaza od konvolučnej reprezentácie vstupu. Vzdialenosť dvoch bodov a, b v tomto priestore označíme $d(a, b)$ a teda chybovú funkciu $J(\theta)$ vieme zapísať ako:

$$J(\theta) = d(c, c\prime) + d(x_{conv}, c) + d(x'_{conv}, c\prime) \qquad (5)$$

Našu intuíciu sme overili jednoduchým experimentom. Z nášho baseline MT modelu natrénovaného na CIFAR10 sme vybrali x_{conv} reprezentácie vstupných obrázkov z validačnej sady a použili ich ako trénovacie dáta pre SOM. Na Obr. 3 zobrazujeme rozloženie tried na natrénovanej mape, kde vidno klastre pre jednotlivé kategórie aj bez použitia SOM, ako chybovej funkcie pri učení. Očakávame, že ak zapojíme nami navrhovanú chybu konzistencie bude táto organizácia lepšia a zároveň očakávame aj zlepšenie samotnej úspešnosti modelu v klasifikačnej úlohe.

Potenciál nášho modelu môže byť aj lepší vhľad do vnútorných reprezentácií neurónových sietí, s možným využitím v doméne vysvetliteľnej umelej inteligencie, keďže nám umožní porovnať odpovede modelu pre neoznačené dáta s prototypmi naučených tried a umožní nám teda dáta klasifikovať a skúmať presnosť modelu v zmysle klasifikácie nových vstupov. Akýsi vhľad môžeme pozorovať už v našom experimente na Obr. 3, kde vidíme prekryv medzi dvoma triedami, ktoré ale znamenajú jelene a kone, čiže to, že sú na mape zaznačené na rovnakom mieste môžeme chápať ako prirodzenú vlastnosť podobnosti týchto tried.

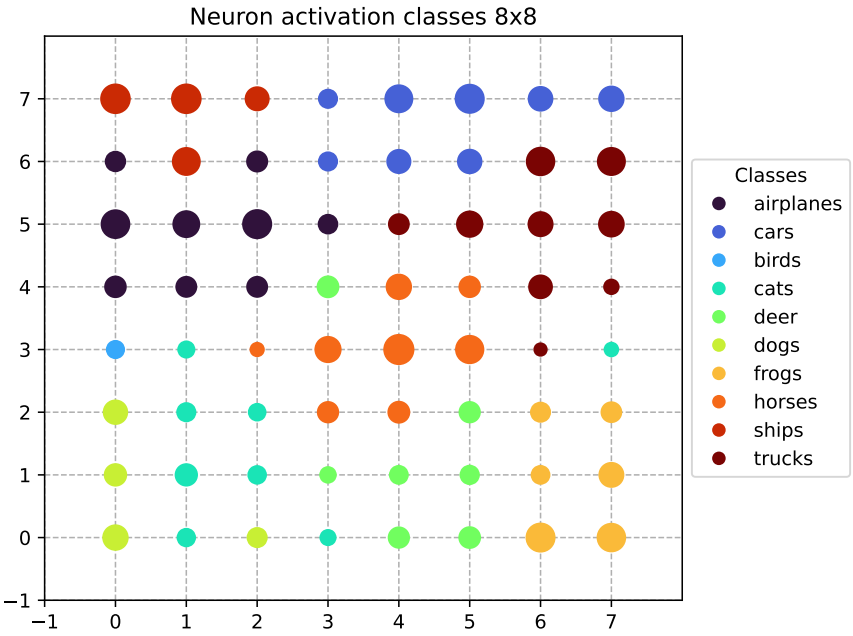
5 Implementácia a predbežné výsledky experimentu

V našej implementácii sme vychádzali z pôvodného kódu autorov modelu MT a prebrali architektúru, ktorú natrénovali na najlepšiu presnosť a zapojili sme do nej samoorganizujúcu sa mapu. Trénovanie prebiehalo tak, že sa v rámci jednej epochy pri doprednom prechode zapamätali reprezentácie x_{conv} z oboch modelov, pomocou spätného šírenia chyby sa upravil študentský model a pomocou EMA sa upravil učiteľský model a následne po úprave váh modelov sa zapamätané x_{conv} použili na natrénovanie SOM. SOM loss sme začali používať na úpravu váh až od druhej epochy, keďže v prvej epoche nebola SOM ešte natrénovaná.

Keďže architektúra Tarvainen a Valpola (2017) bola veľmi veľká, približne 30 miliónov parametrov, nemali sme dostatočné zdroje na trénovanie takéhoto modelu. Rozhodli sme sa teda pre účely experimentovania ďalej pracovať s o niečo menšou architektúrou modelu MT. Vybrali sme si architektúru, ktorú používa Muhammad Sarmad[4]. Tá obsahuje o niečo viac než 3 milióny parametrov, ale je možné ju s použitím našich zdrojov natrénovať.

Aktuálne ešte stále prebieha optimalizácia hyperparametrov, s dôrazom na ladenie nami navrhnutej stratovej funkcie. V budúcnosti plánujeme experimentovať aj s hyperparametrami samoorganizujúcej mapy ako aj jej topológiou. Pri učení SOM plánujeme zapojiť mechanizmus na kompenzáciu toho, že feature vektory, ktoré dostáva SOM na vstupe sa neustále vyvíjajú. Napríklad, že pri výbere víťaza uprednostníme vzdialenejší neurón a/alebo taký, ktorý je málo saturovaný (teda málo krát zvíťazil), s cieľom zapojiť čo najviac neurónov SOM a získať čo najviac vyrovnanú organizáciu na mape.

[4] github.com/iSarmad/MeanTeacher-SNTG-HybridNet



Obr. 3: Distribúcia tried na mape: pre každý neurón siete zobrazujeme, koľko krát bol neurón víťazom pre najpočetnejšie zastúpenú triedu označenú farebne, táto početnosť je zobrazená veľkosťou značky.

6 Záver

Paradigma čiastočne riadeného učenia je zaujímavý nástroj na vysporiadanie sa s nedostatkom adekvátne označených dát a využitím neoznačených dát na zlepšovanie úspešnosti hlbokého modelu. Ukázalo sa, že trieda modelov MT je účinná na takéto situácie. Keďže pôvodná chyba konzistencie, ktorú model využíva, je málo informovaná, navrhli sme ju nahradiť samoorganizujúcou mapou, pretože očakávame, že by mohla vernejšie zachytiť štruktúru a podobnosť dát a tak lepšie určiť ich konzistenciu, ako aj zlepšiť výkon modelu MT a jemu podobných. Takto navrhnutá chybová funkcia by mohla mať využitie aj v iných čiastočne riadených modeloch a obohatiť tieto modely o známy a biologicky relevantný spôsob učenia bez učiteľa.

Poďakovanie

Tento príspevok vznikol v Centre pre kognitívnu vedu na KAI FMFI UK v Bratislave, s podporou grantu VEGA 1/0373/23 a KEGA 022UK-4/2023. Za podporu tiež ďakujeme Slovenskej spoločnosti pre kognitívnu vedu SSKV[5].

[5] https://cogsci.fmph.uniba.sk/sskv/

Literatúra

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480.

Krizhevsky, A., Hinton, G. a spol. (2009). Learning multiple layers of features from tiny images. Technická správa, University of Toronto, Toronto.

Krizhevsky, A., Sutskever, I. a Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. V *Advances in Neural Information Processing Systems*, vol. 25, str. 1097–1105.

Laine, S. a Aila, T. (2016). Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*.

Tarvainen, A. a Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. a Garnett, R. (zost.), V *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc.

Tuna, M., Malinovská, K., Farkas, I., Kraus, S. a Krsek, P. (2021). Semi-supervised learning in camera surveillance image classification. V *2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP)*, str. 155–162.

Weiss, K., Khoshgoftaar, T. M. a Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1):9.

Demografické a kognitívne prediktory klimatického skepticizmu

Beáta Sobotová, Jakub Šrol & Magdalena Adamus
Ústav experimentálnej psychológie, Centrum spoločenských a psychologických vied, Slovenská akadémia vied
Dúbravská cesta 5819/9, 841 04, Bratislava
beata.sobotova@savba.sk, jakub.srol@savba.sk, magdalena.adamus@savba.sk

Abstrakt

Napriek prevažnému vedeckému konsenzu týkajúceho sa hlavných predpokladov človekom zapríčinenej klimatickej zmeny stále zostáva nezanedbateľná časť populácie, ktorá tieto predpoklady z rôznych dôvodov odmieta. V tomto príspevku sme preskúmali vybrané demografické a kognitívne (konšpiračné presvedčenia, nedôvera vo vedu) prediktory klimatického skepticizmu na veľkej reprezentatívnej vzorke slovenskej populácie (N = 1233). Kým demografické prediktory, s výnimkou politickej orientácie, vykázali iba slabý či zanedbateľný súvis s klimatickým skepticizmom, nedôvera vo vedu a konšpiračné presvedčenia silno pozitívne korelovali s touto premennou. Navyše, konšpiračné presvedčenia pôsobili ako mediátor vzťahu medzi konzervatívnou orientáciou a klimatickým skepticizmom.

1 Úvod

Klimatický skepticizmus predstavuje spochybňovanie alebo popieranie fyzikálnych a vedeckých aspektov klimatickej zmeny, jej antropogénneho komponentu, škodlivosti, alebo závažnosti klimatickej zmeny (Capstick & Pidgeon, 2014). Dôvodov tohto skepticizmu je mnoho, avšak za hlavné sa považujú ohrozenie hodnôt a ideológií (Campbell & Kay, 2014), nedostatok dôkazov ľudského zapríčinenia klimatickej zmeny (Whitmarsh, 2008) , ako aj neistá a kontroverzná prezentácia klimatickej zmeny v médiách (Boykoff & Boykoff, 2004). Hornsey a kol. (2016) vo svojej meta-analýze identifikovali sedem demografických prediktorov viery v klimatickú zmenu (teda opak klimatického skepticizmu): pohlavie, vek, príjem, vzdelanie, rasa, politická afiliácia a politická ideológia.

Avšak, ako ukázala neskoršia štúdia (Hornsey et al., 2018), v prediktoroch klimatického skepticizmu existuje značná variabilita naprieč rôznymi krajinami. Preto sme sa v našom príspevku rozhodli preskúmať vybrané prediktory klimatického skepticizmu na reprezentatívnej vzorke slovenskej populácie. Skúmali sme *demografické prediktory* – vek, pohlavie, vzdelanie, politickú orientáciu, subjektívny socio-ekonomický status a *kognitívne p*rediktory – vieru v konšpiračné teórie a nedôvera vo vedu. Nadväzujúc na výskum Hornseyho a kol. (2018), sme tiež preskúmali, či konšpiračné presvedčenia mediujú vzťah medzi konzervativizmom a klimatickým skepticizmom v slovenskej vzorke.

2 Metódy

Výskumu sa zúčastnilo 1233 participantov (571 mužov) vo veku od 18 do 86 rokov (M = 45.8, SD = 16.08), regrutovaných prostredníctvom externej agentúry. Išlo o reprezentatívnu vzorku slovenskej populácie z hľadiska veku a pohlavia. Tento výskum bol súčasťou väčšej série štúdií (viď https://osf.io/5qy2b/). Pre účely tohto výskumu participanti vypĺňali Škálu klimatického skepticizmu (Whitmarsh, 2011), dotazník konšpiračných presvedčení (Šrol et al., 2022) a nedôvery vo vedu (Hartman et al., 2017) a tiež viacero demografických otázok (ich konkrétne znenia uvedené v materiáloch dostupných na https://osf.io/5qy2b/) zameraných na: vek, pohlavie, vzdelanie, politickú orientáciu a subjektívny socio-ekonomický status.

3 Výsledky

3.1 Demografické prediktory

Pohlavie. Na analýzu rozdielov v klimatickom skepticizme medzi pohlaviami sme požili analýzu ANOVA, ktorá ukázala signifikantný rozdiel $F(1,1231) = 4.00$, $p = .046$ naznačujúci vyššiu mieru klimatického skepticizmu u mužov (M = 2.54, SE = 0.04) ako u žien (M = 2.43, SE = 0.04). Avšak, tento rozdiel bol z hľadiska veľkosti efektu zanedbateľný (d = .11).

Vek. Korelácia medzi vekom a klimatickým skepticizmom bola zanedbateľne slabá (r = .09, p = .003). Avšak, keď sme rozčlenili participantov do vekových skupín 18 – 30 (N = 313) , 31 – 50 (N = 437), 51 – 65 (N = 291) a >75 (N = 192) analýza ANOVA ukázala štatisticky signifikantný rozdiel $F(3,1229) = 6.64$, p <.001 medzi jednotlivými skupinami, pričom v najmladšej vekovej skupine bola miera klimatického skepticizmu významne nižšia oproti všetkým ostatným vekovým skupinám (d = .23 - .33, p < .01).

Vzdelanie. Vyššie vzdelanie bolo negatívne, avšak slabo, spojené s nižšou mierou klimatického skepticizmu, r = -.18, p < .001.

Subjektívny socio-ekonomický status. Pri analýze sme sa pozreli na korelácie medzi touto hodnotou a mierou klimatického skepticizmu a našli sme slabú negatívnu koreláciu $r = -.15$, $p < .001$.

Konzervatívno-liberálna politická orientácia. Medzi klimatickým skepticizmom a liberálnou orientáciou sme zistili strednú negatívnu koreláciu $r = -.35$, $p < .001$. Teda, čím boli participanti konzervatívnejší, tým mali vyššiu mieru klimatického skepticizmu.

3.2 Kognitívne prediktory

V rámci kognitívnych prediktorov sme analyzovali vzťah medzi mierou klimatického skepticizmu a konšpiračnými presvedčeniami a nedôverou vo vedu.

Nedôvera vo vedu. Korelácia, ktorú sme našli medzi klimatickým skepticizmom a nedôverou vo vedu je silná pozitívna korelácia $r = .61$, $p < .001$.

Konšpiračné presvedčenia. Klimatický skepticizmus silno koreloval s konšpiračnými presvedčeniami ($r = .68$, $p < .001$). Navyše, konšpiračné presvedčenia mediovali vzťah medzi konzervativizmom a klimatickým skepticizmom. Kým konzervativizmus v mediácii predikoval klimatický skepticizmus ($\beta = .10$, $p < .001$), 72% vzťahu medzi týmito premennými sa dalo vysvetliť nepriamym efektom ($\beta = .25$, $p < .001$) mediovaným konšpiračnými presvedčeniami (konzervativizmus > konšpiračné presvedčenia: $\beta = .39$, $p < .001$; konšpiračné presvedčenia > klimatický skepticizmus: $\beta = .64$, $p < .001$).

4 Diskusia a záver

V zhode so štúdiou Hornseyho a kol. (2016) sme zistili, že kým väčšina demografických prediktorov má slabý či zanedbateľný súvis s klimatickým skepticizmom, konzervativizmus bol stredne silno, a konšpiračné presvedčenia a nedôvera vo vedu silno korelované s klimatickým skepticizmom. Predchádzajúci výskum naznačuje, že vzťahy klimatického skepticizmu, konzervativizmu a konšpiračných presvedčení sú silnejšie v USA v porovnaní s inými krajinami pravdepodobne v dôsledku polarizovanej debaty o tejto téme (Hornsey et al., 2018). Zaujímavým výsledkom nášho výskumu je, že sila vzťahov medzi týmito premennými na Slovensku je konzistentná skôr s výsledkami USA ako iných krajín. Podobne, aj pozorovanie, že konšpiračné presvedčenia mediujú vzťah politickej orientácie a klimatického skepticizmu naznačuje, že diskurz o klimatickej zmene na Slovensku môže byť prepájaný s konšpiračnými teóriami. Dôsledkom toho môže byť podrývanie dôvery vo vedecký konsenzus o klimatickej zmene a neochota prijímať potrebné opatrenia na zabránenie zhoršenia environmentálnych problémov.

Poďakovanie

Tento príspevok vznikol s grantovou podporou projektu VEGA 2/0053/21 a APVV-20-0335.

Literatúra

Boykoff, M. T., & Boykoff, J. M. (2004). Balance as bias: global warming and the US prestige press. *Global Environmental Change*, *14*(2), 125–136. https://doi.org/10.1016/j.gloenvcha.2003.10.001

Campbell, T. H., & Kay, A. C. (2014). Solution aversion: On the relation between ideology and motivated disbelief. *Journal of Personality and Social Psychology*, *107*(5), 809–824. https://doi.org/10.1037/a0037963

Capstick, S. B., & Pidgeon, N. F. (2014). What is climate change scepticism? Examination of the concept using a mixed methods study of the UK public. *Global Environmental Change*, *24*, 389–401. https://doi.org/10.1016/j.gloenvcha.2013.08.012

Hartman, R. O., Dieckmann, N. F., Sprenger, A. M., Stastny, B. J., & DeMarree, K. G. (2017). Modeling Attitudes Toward Science: Development and Validation of the Credibility of Science Scale. *Basic and Applied Social Psychology*, *39*(6), 358–371. https://doi.org/10.1080/01973533.2017.1372284

Hornsey, M. J., Harris, E. A., Bain, P. G., & Fielding, K. S. (2016). Meta-analyses of the determinants and outcomes of belief in climate change. *Nature Climate Change*, *6*(6), 622–626. https://doi.org/10.1038/nclimate2943

Hornsey, M. J., Harris, E. A., & Fielding, K. S. (2018). Relationships among conspiratorial beliefs, conservatism and climate scepticism across nations. *Nature Climate Change*, *8*(7), 614–620. https://doi.org/10.1038/s41558-018-0157-2

Šrol, J., Čavojová, V., & Ballová Mikušková, E. (2022). Finding someone to blame: The link between COVID-19 conspiracy beliefs, prejudice, support for violence, and other negative social outcomes. *Frontiers in psychology*, *12*, 6390. https://doi.org/10.3389/fpsyg.2021.726076

Whitmarsh, L. (2008). Are flood victims more concerned about climate change than other people? The role of direct experience in risk perception and behavioural response. *Journal of Risk Research*, *11*(3), 351–374. https://doi.org/10.1080/13669870701552235

Whitmarsh, L. (2011). Scepticism and uncertainty about climate change: Dimensions, determinants and change over time. *Global Environmental Change*, *21*(2), 690–700. https://doi.org/10.1016/j.gloenvcha.2011.01.016

Poetická autopoiesis

Mgr. Aleš Svoboda

Univerzita Karlova, Fakulta humanitních studií
Pátkova 2137/5, Praha 8 – Libeň, 182 00
ales.svoboda@fhs.cuni.cz

Abstrakt

Počítačové generativní umění inspirující se umělým životem vychází pravidelně z různých vlastností tradičně spojovaných s představou života. Ústředním tématem by ovšem měla být vizualizace principu autentického vzniku cyklických přeměn, které postupně zvrátí chaotickou vizuální homogennost do odolné strukturovanosti. Taková odolnost by však neměla být rigidností vůči dalším možným sebetransformacím. Důsledkem by pak mohl v souhlasu s poetikou výtvarných děl současný nárůst druhu skladebných prvků a celkové komplexity vizuální reprezentace odpovídající druhovému vývoji a rámci ekosystému.

1 Atributy života

Zřejmě panuje shoda nad výčtem atributů života. Lze tvrdit, že každá reprezentace života zahrnuje (v nějakém stadiu či v úhrnu celého životního cyklu) všech těchto devět charakteristik: samoorganizaci, autonomii, emergenci, růst, adaptaci, citlivost, rozmnožování, vývoj a metabolismus (Boden 2016). I Bodenová ovšem pokládá za nejvýznamnější vlastnost samoorganizaci, která ostatní rysy v nějakém smyslu zahrnuje.

Lze také říci, že principiální rozdíl mezi neživými a živými entitami spočívá v rozdílné odolnosti proti vnějším vlivům. Trvání neživých entit je pasivní, jejich struktura odolává v mezích svých fyzikálních a chemických vlastností. Naopak živé entity se jeví jako aktivní, jejich trvání zajišťuje řada strategií doplňování, přeskupování a přizpůsobování vlastní struktury.

Základní princip zachování existence takové struktury je „sebeudržování“, ten ovšem přerůstá ve strategii vyššího stupně – ve vytvoření nové struktury se stejnými vlastnostmi, ve strategii „autoreprodukce“. Vztah sebeudržování a autoreprodukce v nás pak může vyvolat potřebu určit, co bylo prvotní, tradiční otázku paradoxu počátku – je klíčovým principem pro vznik života autoreprodukce, tedy rozmnožování, posléze opět rozvinuté do vývoje, nebo spíš schopnost adaptace a autonomie, podporované citlivostí a růstem? Naše pragmatická intuice naznačuje, že oba póly existují souběžně a jsou v jistém ustáleném stavu rovnováhy.

2 Autopoiéze

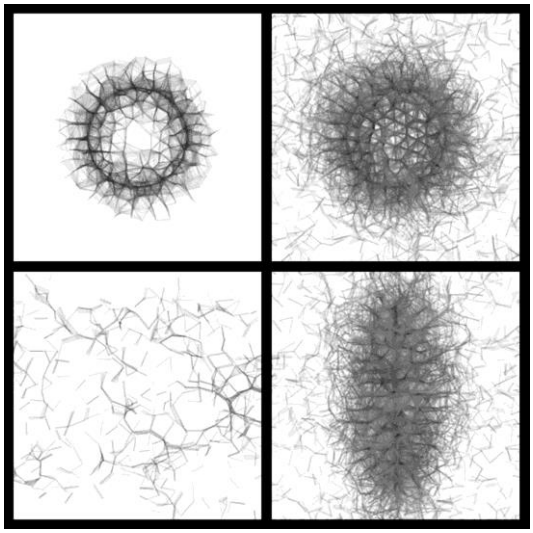
V podnětném, ale často i kritizovaném díle *Strom poznání* přicházejí autoři Maturana a Varela (2016) s termínem autopoiéze. Vtělují do něj „základní princip živých bytostí“. V širokém kosmologickém kontextu vytváří dispozici pro vznik života překonání homogenity abiogenním stvořením specifických organických molekul. Jejich zrod lze vlastně chápat jako náhodné dovršení náhodných předpokladů, které však ústí do pravidelných a zákonitých rámců.

Organické molekuly přinášejí nekonečnou diverzitu a plasticitu, stávají se platformou pro diverzitu molekulárních reakcí, které existenci živých bytostí umožňují. Představu o odlišnosti živých bytostí a priori sdílíme, je nesena základním kognitivním aktem, který postihuje jistou nutnou *organizaci* vztahů, které tuto třídu tvoří. Maturana a Varela docházejí k závěru, že třída, typ této jednotky je specifický tím, že stále utváří sebe sama. Příslušný termín skládají z řeckých slov *autos* (samo) a *poiein* (tvořit). Jako srozumitelný příklad toho, co rozumí autopoiézí, předkládají buňku. Buňka je trvalou nepřetržitou interakcí jednak svých částí navzájem, jednak sebe a vnějšku. Její nutné vymezení tvoří membrána, která má ambivalentní pozici – je zároveň hranicí i účastníkem interakcí. Nejosobitější charakteristika autopoietického systému je, „že se utváří jako systém odlišný od okolního prostředí pomocí své vlastní dynamiky.“(Maturana a Varela, 2016: 40)

3 Modelování atributů života generativním počítačovým uměním

Počítačové generativní umění adaptuje od 70. let v řadě případů techniky modelování umělého života. Staví na „hře života“ Johna Conwaye (Paul Brown), na modelování růstu (Yoichiro Kawaguchi), na modelování genetických přenosů započatých názornou metaforou „biomorf“ Richarda Dawkinse (William Latham a Stephen Todd, Karl Sims), na růstových programech L-systémů (Jon McCormack, Casey Reas), případně na produkci genetických algoritmů ve vývojových programech pro křížení samotných

vizuálních struktur (Penousal Machado, Scott Draves, Philip Galanter, Erwin Driessens a Maria Vestappen). Modelování jádra „autopoiéze“, spontánního přechodu nudné homogenity autoreferenčním nárůstem organizace ke strukturám zhmotňujícím svými složkami a vztahy právě zmnožující se organizaci v individuální jednotce, však v zásadě zůstává opomenuto.



Obr. 1: Casey Reas: Articulate, 2003. Čtyři sekvence z průběhu generativního počítačového díla

4 Poetika

Jedním z pilířů umělecké teorie je poetika, která má opět stejný etymologický původ ve slově „tvořit“. Rozumí se jí nauka o způsobu budování uměleckého díla a porozumění jeho struktuře, což je samozřejmě umožněno předpokladem, že umělecké dílo takovou racionálně postižitelnou strukturu má. Tato struktura uměleckého díla je v historické a společenské perspektivě přístupná zobecněním, která dovolují o jednotlivých uměleckých dílech uvažovat v kategoriích stylu. Ve výtvarném díle se především jedná o analýzu jeho součástí a o principy jejich spojení, o prvky a jejich vazby, což tradičně označuje pojem kompozice.

Formalistická estetika, fundovaná vznikem a rozvojem nezobrazivého, především abstraktně geometrického umění, uvažuje nad výtvarnými uměleckými díly v termínech prvků a vztahů, provázaných částí a diferencujících se celků. Představa díla jako jevu „organické“ povahy se nezdá být pouhou metaforou, ale jasným odrazem fungování vizuálních kvalit díla v procesu percepce a porozumění. Chybějící aspekt úplného zživotnění výtvarných uměleckých děl byl samotný reálný pohyb, který po nesystémových manifestacích z počátku 20. století a více méně mechanické podobě v kinetickém umění nalezl plnou podporu a rozvinutí právě v počítačovém generativním umění.

5 Vizuální struktury

Mechanismy vnímání, respektive jeho evoluční povahou se empirickým způsobem zabýval John D. Barrow (2000). Jeho naturel astronoma a přírodního vědce nestaví hypotézy o estetickém a uměleckém vnímání na iracionálních a idealistických východiscích, naopak pro estetické jevy hledá pragmatické a praktické důvody. Stručně řečeno, vnímat a klasifikovat vizuální struktury je naše základní adaptivní reakce. Přičemž „[s]chopnost rozpoznávat struktury poskytuje dostatek prostoru k tomu, aby jako vedlejší produkt rozkvétalo naše estetické cítění.“(Barrow, 2000: 140)

Základní principy systematické tvorby a popisu struktur lze odvodit od lineárních vlysů a existují pouze čtyři: translace (posunutí), zrcadlení, rotace a sestupová zrcadlení. Pro plošné struktury stoupne počet těchto základních postupů na sedmnáct. Jedním z elementárních postupů je posunutí a rotace při vzniku moiré. To se může stát základem rekurzivních struktur, které vlastně definičně odpovídají principům autopoietického systému.

6 Sebeprodukce poznání

Pravděpodobně nejkontroverznějším stanoviskem Maturanovy a Varelovy knihy je tvrzení, že poznání lze pochopit, zjednodušeně řečeno, spíše jako produkt životní praxe, než jako otisk skutečnosti. Ve vědomí je radno hledat samu příčinu podoby poznaného, nikoliv záznam pouhého důsledku. Tedy podobně, jako je život sebereprodukcí a sebeprodukcí, je poznání sdíleným konsensuálním „sebetvořením“ účinných struktur. K tomu právě výtvarné umění trvale přispívá – rodí nové struktury, které dynamizují naše vědomí, rozšiřují jeho kapacitu a připravují ho na nové účinnější koncepty. Průběhový, stále se proměňující mod „poetické autopoiéze“ by měl činit zadost jak modelování zřejmě základního principu života, tak i povaze estetického poznání.

Poděkování

Tento výstup vznikl v rámci programu Cooperatio, vědní oblasti Vědy o umění a kultuře.

Literatura

Barrow, J. D. (2000). *Vesmír plný umění*. Jota.

Boden, M. A. (2016). *AI, Its nature and future*. Oxford University Press.

Maturana, H. R. a Varela, J. F. (2016). *Strom poznání. Biologické základy lidského rozumu*. Portál.

Svoboda A. (2017). Modelování života a organismus uměleckého díla. *Kognícia a umelý život XVII,* 163–168

Imerzivní virtuální realita jako nástroj pro simulaci evakuačního chování

Čeněk Šašinka, Zdeněk Stachoň, Alžběta Šašinková, Kateřina Jochecová

Katedra informačních studií a knihovnictví,
Geografický ústav
Katedra informačních studií a knihovnictví
Katedra informačních studií a knihovnictví a Psychologický ústav
Masarykova univerzita
cenek.sasinka@mail.muni.cz, stachon@mail.muni.cz, asasinkova@mail.muni.cz, katerina.jochecova@mail.muni.cz

Abstrakt

Virtuální realita, a zvláště ta plně imerzivní představuje silný nástroj pro realizaci simulací, které by v reálném světě byly jen těžko proveditelné. Jednou z oblastí užití jsou simulace orientace a navigace v prostředí, například v kontextu únikového chování z budov při krizových situacích. Než lze ale výsledky takových simulací skutečně použít pro praxi je třeba odpovědět na otázku? V jak velké míře jsou simulace ve virtuální realitě spolehlivé a věrné skutečnosti a v jaké míře a v jakých parametrech se svého předobrazu z reálného světa liší. Cílem příspěvku je přiblížit postupy, které byly použity pro empirické srovnání chování lidí v reálné budově a při simulaci v jeho digitálním dvojčeti. Při dodržení definovaných standardů dosažení zjištění mohou vést k optimismu, nicméně zároveň poukazují na aspekty, kterým je třeba věnovat další pozornost.

1 Potenciál virtuální reality pro simulace

Virtuální realita jako médium nabízí příležitost převést část aktivit z reálného světa, především těch nebezpečných, nákladných, či obtížně uskutečnitelných, do světa umělého. Existují tak různé letecké simulátory, simulátory pozemních dopravních prostředků, probíhá nácvik prezentačních dovedností či komunikace příslušníků PČR při řešení zátěžových situací (Moravčík, 2021). Zásadní otázkou ale je, v jaké míře je chování a prožívání osob ve virtuálním světě shodné s tím reálným, a v jak velké míře je tedy možné z chování ve VR odvozovat na chování v reálném světě. V jak velké míře dochází rovněž k přenesení zkušenosti.

2 Výzkum v oblasti evakuace z budov

Jednou z oblastí, ve které má virtuální realita skutečně významný přínos, je simulace evakuačního chování z budov či větších komplexů, jako jsou např. nádraží, letiště či stadiony. Důvodů je několik. I pokud stavba již v dané době skutečně existuje, jsou simulace za běžného provozu velice komplikované a mnohdy nákladné. Dalším důvodem, proč využívat pro simulace virtuální realitu je především možnost, že lze testovat evakuační chování v budovách, které jsou teprve ve fázi návrhu. U budov navrhovaných v současnost je běžnou součástí dokumentace v podobě digitálního modelu (BIM - Building Information Modeling), který lze po úpravách využít pro implementaci do platformy umožňující simulace (Kvarda, 2021). V případě, že je model dostupný, lze ho testovat a případně upravovat před vlastní realizací. V rámci aplikovaného výzkumného projektu "Kognitivní psychologie a prostorová syntaxe ve virtuálním prostředí pro agentní modely - TL02000103" bylo hlavním cílem ověřit, v jaké míře odpovídá chování jedinců při evakuaci ve virtuální realitě jejich chování v reálné budově. Aby bylo možné tuto otázku zodpovědět, bylo nejdříve nutné navrhnout samotné řešení simulací ve virtuální realitě. Byl tak navržen optimalizovaný postup, který umožnil využít BIM a na základě definovaných kroků vytvořit 3D model (digitální dvojče), které bylo následně využito při simulaci v herním enginu Unity. Druhým nutným krokem byl tedy vývoj aplikace v Unity, která umožňovala realizovat nejen simulovanou evakuaci při adekvátní interakci jedince s objekty budovy, ale zároveň umožňovala trekovat jeho akce, pohyb a dokonce oční pohyby (Ugwitz et al., 2022). Výstupem první fáze tak je metodika a aplikace, která umožňuje vytvářet digitální dvojčata budov a realizovat evakuaci. Díky dalším knihovnám a možnostem Unity lze řešení doplnit např. i autonomní agentní modely (viz Jualiani et al., 2018). Neméně důležitým výstupem je návrh metod a analýzy chování jedinců při evakuaci ve VR a jeho srovnání s daty z reálného průchodu reálnou budovou. Toto řešení navrhly společně Geografický ústav a Katedra informačních věd a knihovnictví na MU. Výsledky srovnání prokázaly, že v mnoha zásadních parametrech (např. vzdálenost první registrace navigačního značení) bylo chování srovnatelné (Stachoň et al., preprint 2022). Přes náročné řešení nebylo možné reflektovat všechny parametry, resp. operovat s jejich úrovní.
Hardware

2.1 Hardware a software

Pro výzkum v reálném prostředí byly použity ET brýle 2 od SMI (frekvence snímání 60 Hz, rozlišení 960 x 720p, 30 FPS). Pohyb participantů byl nahráván na přenosnou kameru Forever SC-210 Plus action cam (1920 x 1080p, 30FPS). Pro pohyb ve virtuálním prostředí byl použit headset HTC VIVE Pro Eye (1440 x 1600, 90 Hz, integrovaný ET snímal data ve frekvenci 120 Hz) a pro pohyb byl využit Lighthouse Position Tracking system (gen 2), který je v headsetu již zabudovaný (Stachoň et al., 2022).
Původní model budovy byl převeden do Unity pomocí Revit 2020 a Tridify cloud service. Pořipostupné optimalizaci byly doplněny další menší objekty vytvořené v Blenderu 2.80. Jako podpora eye-trackeru byl využit Vive SRanipal SDK 1.3, pro logování interakcí uživatelů s prostředím byl použit Toggle Toolkit (Ugwitz et al., 2021).
Pro analýzu dat z reálného prostředí byl použit program BeGaze od SMI (Stachoň et al., 2022).

2.2 Výsledky studie

Výsledky analýzy eye-trackingových dat ukázaly, že z hlediska pozornosti na objekty není v prostředích rozdíl. Pravděpodobnost, že se člověk zaměří na určitou kategorii (evakuační značení, informační cedule, nábytek, okna) je přibližně stejná, bez statisticky významných rozdílů. I vzdálenost, ze které se na daný objekt poprvé člověk podívá je srovnatelná až na okna, které upoutaly pozornost osob ve VR dříve o 1.6 metru, což je statisticky významné (Stachoň et al., 2022).
Z analýzy trasy pohybu vyplývá, že překryv trajektorie pohybu napříč celou trasou je 70 %. Osoby v reálném prostředí ale měli tendence držet se blízko zábradlí (oproti VR, kde se lidé drželi uprostřed schodiště) a obcházet sloupy ve větší vzdálenosti oproti VE (Stachoň et al., 2022).

3 Limity studie a další možnosti vývoje

I přesto, že se jedná o jedinečný výzkum, je stále ještě mnoho aspektů na které je třeba se zaměřit. Ať už kvůli hlubšímu porozumění kognitivních procesů během evakuace, tak kvůli vytvoření širší oblasti využití těchto prostředí.

3.1 Single vs. multiplayer

V reálném světě se jen málokdy evakuují jedinci samostatně. Jako další možný krok je tedy výzkum skupinového chování. Skupinová dynamika v reálném prostředí se zdá být přirozeně selektivní (např. Fu, Cao, Song a Fang, 2019), ale je třeba ověřit, zda virtuální prostředí nemá vliv na tvorbu sociálních struktur.

3.2 Více individuální odlišnosti

Na evakuační strategii v reálném prostředí má vliv více faktorů například pohlaví (Lin et al., 2012) či věk (Yamamoto et al., 2019), vliv mají i zkušenost (Hemmer er al., 2015) nebo kognitivní styl (Bocchi et al., 2019). Je možné, že některé z těchto faktorů bude ovlivňovat evakuaci ve virtuálním prostředí jinak než v tom reálném.

3.3 Více zapojit stress

Rozhodování, které je nedílnou součástí reálné evakuace, je ovlivněno stresem (Lerner, Li, Valdesolo, & Kassam, 2015). Částečně to šlo pozorovat i v tomto výzkumu, i když stres navozován nebyl. Někteří participanti však byli i tak velmi vystresovaní a chovali se jinak oproti zbytku. V navazujících studiích lze navozovat stres (například nezvyklým alarmem nebo prezentací s nenadálým podnětem) a sledovat rozdíly v stresové reakce v závislosti na prostředí a následný vliv stresu na evakuační strategie.

3.4 Více změn prostředí a typů scénářů

Prostředí je možné modifikovat - například přidáním kouře nebo ohně pro autentičnost. Virtuální budovy jsou také ideální pro nenákladné zkoušení různých prototypů nového orientačního značení. Podle potřeb jednotlivých výzkumů je možné měnit i scénáře - použít participanty, kteří budovu již znají (narozdíl od participantů v aktuálním výzkumu) nebo změnit jejich cíl, aby došlo ke komplexnější orientaci - například hledat konkrétní místo oproti jakémukoliv východu.

3.5 Otázka limitů způsobů lokomoce

Existuje několik způsobů ovládání pohybu po virtuálním prostředí (Boletsis & Chasanidou, 2022). Pro každý scénář se však hodí pouze některá. Je potřeba brát v potaz blízkost k reálnému pohybu, vyvolávání nevolnosti (*motion sickness*) nebo rychlost přenosu do prostředí. Jako nejpřirozenější náhrada pohybu by v našem typu experimentu byl mnohosměrný pohyblivý pás. Ten je však zbytečně nákladný a jak bylo potvrzeno v tomto i v předchozích výzkumech, pro takový typ experimentu není nezbytný.

4 Oblasti využití

Jelikož byla ověřena ekvivalence vizuálního zpracování prostoru, lze vytvořené prostředí používat v rámci nácviků situací, které jsou na 100% podobnosti prostředí závislé. Protože se může jednat o nebezpečné situace (např. nácvik evakuačních postupů v komplexních budovách, vojenské výcviky) bylo nejprve potřeba ověřit ekvivalenci, aby nácviky byly efektivní. Pokud je prostředí jako intervenující

proměnná vyřazeno, lze virtuální realitu využívat k velmi detailnímu výzkumu kognice.

Poděkování

Tento výzkum vznikl za podpory Technologické agentury České republiky, grantové číslo TL02000103 (Kognitivní psychologie a prostorová syntaxe ve virtuálním prostředí pro agentní modely).

Literatura

Bocchi, A., Palmiero, M., Nori, R. et al. (2019). Does spatial cognitive style affect how navigational strategy is planned?. *Exp Brain Res,* 237, 2523–2533.

Boletsis, C., & Chasanidou, D. (2022). A Typology of Virtual Reality Locomotion Techniques. *Multimodal Technol. Interact.,* 6, 72.

Fu, L., Cao, S., Song, W., & Fang, J. (2019). The influence of emergency signage on building evacuation behavior: An experimental study. *Fire and Materials*, 43(1), 22-33.

Hemmer, I., Hemmer, M., Neidhardt, E., Obermaier, G., Uphues, R., & Wrenger, K. (2015) The influence of children's prior knowledge and previous experience on their spatial orientation skills in an urban environment, Education 3-13, 43:2, 184-196

Juliani, A., Berges, V. P., Teng, E., Cohen, A., Harper, J., Elion, C., ... & Lange, D. (2018). Unity: A general platform for intelligent agents. arXiv preprint arXiv:1809.02627.

Kvarda, O. (2021). Usability of building information modeling (BIM) for generating virtual geographic environments (VGEs). In International Cartographic Conference, Florence. doi:10.5194/ica-abs-3-168-2021.

Lerner, J., Li, V., Valdesolo, P., & Kassam, K. (2015). Emotion and Decision Making. *Annual Review of Psychology*, 66, 799-823.

Lin, C., Huang, T. Y., Lin, W. J., Chang, S. J., Lin, Y. H., Ko, W. I. et.al (2012). Gender differences in wayfinding in virtual environments with global or local landmarks. *Journal of Environmental Psychology*, 32, 89-96.

Moravčík, O. (2021). Virtuální realita ve výcviku policistů. PČR: https://www.policie.cz/clanek/virtualni-realita-ve-vycviku-policistu.aspx

Stachoň, Jochecová, Kvarda, Snopková, Ugwitz, Šašinková, Ježek, Kubíček, Juřík, Švedová, Šašinka. The Possibilities of Using Virtual Environments in Research on Wayfinding, 13 September 2022, PREPRINT (Version 1) available at Research Square.

Ugwitz, P., Šašinková, A., Šašinka, Č., Stachoň, Z., & Juřík, V. (2021). Toggle toolkit: A tool for conducting experiments in Unity virtual environments. *Behavior Research Methods*.

Yamamoto, N., Fox, M., Boys, E., & Ord, J. (2019). Effects of orientation change during environmental learning on age-related difference in spatial memory. *Behav. Brain Res*., 365, 125–132.

Ugwitz, P., Kvarda, O., Juříková, Z., Šašinka, Č., & Tamm, S. (2022). Eye-Tracking in Interactive Virtual Environments: Implementation and Evaluation. Applied Sciences, 12(3), 1027.

Paradoxy hodnotovej orientácie Sloveniek a Slovákov: demokratické hodnoty a pravicové autoritárstvo

Jakub Šrol & Vladimíra Čavojová

Ústav experimentálnej psychológie, Centrum spoločenských a psychologických vied, Slovenská akadémia vied
Dúbravská cesta 9, 841 04 Bratislava
jakub.srol@savba.sk; vladimira.cavojova@savba.sk

Abstrakt

Hodnotová orientácia participantov predstavuje dôležitý faktor v mnohých spoločenskovedných výskumoch – viaže sa napríklad s podliehaním konšpiračným teóriám, či akceptáciou spoločensky kontroverzných vedeckých poznatkov. Meranie hodnotovej orientácie však často komplikuje ich kultúrna špecifickosť. V našom výskume sme sa zamerali na určenie hodnotovej orientácie participantov (demokratické hodnoty, pravicové autoritárstvo) vo vzťahu k politickej orientácii (konzervativizmus – liberalizmus) a podliehaniu konšpiračným teóriám. Štúdie sa zúčastnilo 896 participantov kontaktovaných prostredníctvom agentúry, ktorá zabezpečila reprezentatívnosť vzorky z hľadiska veku, pohlavia a kraja bydliska. Výsledky ukázali, že nástroje na meranie demokratických hodnôt a pravicového autoritárstva vykazovali nízku mieru vnútornej konzistencie. Detailná analýza poukázala na nekonzistentnosť a prekvapivé smerovanie vzťahov niektorých položiek s inými premennými (napríklad pozitívne vzťahy položiek merajúcich pravicové autoritárstvo s politickým liberalizmom, či ich negatívny vzťah s konšpiračnými presvedčeniami). V diskusii sa venujeme hlbšej analýze nástrojov na meranie hodnotovej orientácie pre účely spoločenskovedného výskumu na Slovensku s návrhmi na riešenia problémov identifikovaných na základe našej štúdie.

1 Úvod

Hodnotová orientácia je kľúčovým faktorom v mnohých spoločenskovedných výskumoch – viaže sa napríklad s dôverou konšpiračným teóriám (van Mulukom et al., 2023), akceptáciou spoločensky kontroverzných vedeckých poznatkov (Drummond & Fishhoff, 2017), či správaním v oblasti ochrany životného prostredia (Adamus et al., 2023), alebo sociálne žiadúceho správania počas pandémie COVID-19 (Čavojová et al., 2022). Avšak, slovenské výskumy týkajúce sa hodnotovej orientácie opakovane narážajú na problémy s meraním tohto konceptu – napr. nízka reliabilita, nekonzistentná faktorová štruktúra metód (napr., Adamus et al., n.d.; Kanovský & Kocičová, 2018; Kostovičová et al., 2017) – čo môže byť dôsledkom vysokej kultúrnej špecifickosti hodnotovej orientácie a neprispôsobenia meracích nástrojov na slovenské podmienky.

Napríklad, na Slovensku zriedka funguje delenie sa na politickú pravicu a ľavicu. V tomto výskume sme sa preto zamerali nielen na tradičnú seba-identifikáciu participantov ako skôr liberálne alebo konzervatívne orientovaných, ale najmä na „abstraktnejší“ koncept demokratických hodnôt. Demokratické hodnoty sa zvyčajne vzťahujú na prikladanie vysokej hodnoty aspektom ako sú práva menšín, vláda zákona a sloboda prejavu (Canetti-Nissim, 2004). Naopak, pravicové autoritárstvo sa často spája s pravicovou politickou orientáciou, náboženským fundamentalizmom, sociálnym konzervativizmom, tradiconalizmom a predsudkom voči menšinám a „vonkajším skupinám“ (Duckitt, 2022). A napokon, politický cynizmus predstavuje rozsah, v akom ľudia vidia politikov ako zásadne nedôveryhodných a politiku ako neúctyhodnú profesiu (Pattyn et al., 2012). Inými slovami, kým cynizmus vo všeobecnosti spochybňuje motívy všetkých ľudí, politický cynizmus sa zameriava na politikov a ich pochybné motívy.

V našom výskume sme sa preto rozhodli bližšie pozrieť na vzájomné vzťahy viacerých zložiek hodnotovej orientácie – politického smerovania, demokratických hodnôt a pravicového autoritárstva – u veľkej reprezentatívnej vzorky slovenskej populácie. Našim cieľom bolo zhodnotiť jednak spoľahlivosť krátkych nástrojov na zachytávanie demokratických hodnôt a pravicového autoritárstva a druhak preskúmať konzistentnosť odpovedí participantov na hodnotovo ladené otázky naprieč viacerými metódami. Naše výsledky sme porovnávali s predchádzajúcimi výskumnými zisteniami o pozitívnom vzťahu pravicového autoritástva s dôverou konšpiračným teóriám (Abalakina Paap et al., 1999; Bruder et al., 2013) a naopak, jeho negatívnym vzťahom s podporou občianskych slobôd a demokratickými hodnotami (Cohrs et al., 2005; Crowson, 2009).

2 Metódy

2.1 Participanti

Výskum bol súčasťou väčšej longitudinálnej štúdie týkajúcej sa presvedčení ohľadom súčasnej vojny na Ukrajine. Výskumu sa zúčastnilo 896 participantov (415 mužov, 481 žien) vo veku 18 až 79 rokov (M = 44.4, SD = 13.7).

2.2 Materiály

Deskriptívne hodnoty a údaje o vnútornej konzistencii škál merajúcich jednotlivé premenné sa nachádzajú v Tabuľke 1 nižšie. Pri škálach tvorených viacerými položkami sme pred samotnou analýzou vypočítali priemerné hodnotenie zo všetkých položiek (berúc do úvahy prípadné reverzne-skórované položky). Položky dotazníkov merajúich pravicové autoritárstvo, demokratické hodnoty a politický cynizmus boli preložené autormi štúdie, pri ostatných sme použili existujúce slovenské verzie.

Pravicové autoritárstvo sme merali Veľmi krátkou škálou pravicového autoritárstva (Bizumic & Duckitt, 2018), ktorá pozostáva zo 6 položiek. Participanti odpovedali na škále od 1 (= *úplne nesúhlasím*) po 5 (= *úplne súhlasím*).

Demokratické hodnoty sme merali 6 otázkami z Canetti-Nisim (2004). Participanti odpovedali na škále od 1 (= *úplne nesúhlasím*) po 5 (= *úplne súhlasím*).

Politická orientácia. Na zistenie politickej orientácie sme použili jednu položku, v rámci ktorej sa mali participanti sebaidentifikovať na škále od 1 (*silno konzervatívny*) po 7 (*silno liberálny*).

Preferencia liberálnej vlády nad autoritárskym vodcom bola meraná otázkou: „Ktorá s nasledujúcich foriem vlády je podľa vás lepšia pre Slovensko?“ 1 = mať silného a rozhodného vodcu, ktorý sa neobťažuje s parlamentom alebo voľbami, 5 = mať liberálnu demokraciu s pravidelnými voľbami a systémom viacerých strán.

Výmena istôt. Požili sme 3 otázky z výskumu Globsec, ktorými sme sa participantov pýtali: „Vymenili by ste určité práva a slobody ako napríklad slobodu cestovania, združovania, alebo slobodu slova, za: (1) lepšiu finančnú situáciu? (2) za vyššiu bezpečnosť v našej krajine? (3) za zachovanie tradičných hodnôt na Slovensku? Participanti odpovedali na škále od 1 (= *určite nie*) po 5 (= *určite áno*).

Politický cynizmus sme merali piatimi položkami zo škály politického cynizmu (Pattyn et al., 2012), ktoré sme doplnili otázkou z výskumu Globsec (Globsec, 2020) o vnímaní demokracie.

Ohrozenie identity a hodnôt: Požiadali sme participantov, aby na škále od 1 (= vôbec neohrozuje moju identitu a hodnoty) po 5 (= veľmi ohrozuje moju identitu a hodnoty) naznačili, ako ich identitu a hodnoty ohrozujú (1) západné krajiny a ich spôsob života, (2) Európska únia, (3) USA, (4) Rusko.

Konšpiračné presvedčenia sme merali prostredníctvom siedmich položiek týkajúcich sa dôvery v rozšírené konšpiračné teórie z výskumu Šrola a kol. (2022). Participanti vyjadrovali svoj ne/súhlas s položkami na 5-bodovej škále.

3 Výsledky

V tabuľke 1 uvádzame deskriptívne hodnoty pre jednotlivé premenné v tejto štúdii. Už zo samotnej tabuľky vidieť, že škály použité na meranie demokratických hodnôt a pravicového autoritárstva vykazujú veľmi nízku mieru vnútornej konzistencie. Je síce pravda, že každý z týchto škál pozostáva iba zo šiestich položiek, čo by čiastočne mohlo vysvetľovať nízku mieru reliability. No na druhej strane, iné premenné v tejto štúdii (politický cynizmus, výmena istôt a konšpiračné presvedčenia) pozostávajú z podobného počtu položiek a napriek tomu vykazujú oveľa vyššie a uspokojivé hodnoty vnútornej konzistencie.

Tabuľka 1. Základné údaje o premenných v tejto štúdii

	M	*SD*	*α*	range
Pravicové autoritárstvo	2.53	0.63	.46	1-5
Demokratické hodnoty	3.93	0.59	.51	1-5
Politický liberalizmus	3.96	1.33	–	1-7
Preferencia liberálnej vlády	3.54	1.26	–	1-5
Výmena istôt	2.87	1.12	.86	1-5
Politický cynizmus	3.78	0.86	.83	1-5
Ohrozenie identity:				
Západné spoločnosti	2.75	1.29	–	1-5
EÚ	2.70	1.30	–	1-5
USA	3.03	1.41	–	1-5
Rusko	3.11	1.21	–	1-5
Konšpiračné presvedčenia	2.54	.148	.93	1-5

Poznámka. Tabuľka zobrazuje deskriptívne hodnoty (priemer, štandardná odchýlka), rozsah hodnôt a vnútornú konzistenciu jednotlivých premenných v našej štúdii.

V dôsledku nízkej miery vnútornej konzistencie škál merajúcich demokratické hodnoty a pravicové autoritárstvo sme sa položky v týchto škálach rozhodli analyzovať separátne. Tabuľka 2 obsahuje korelácie jednotlivých položiek s ostatnými premennými v našej štúdii. Ako možno vidieť z tejto tabuľky, položky dvoch škál nevykazujú konzistentné korelácie s inými, relevantnými, hodnotovými premennými.

Najskôr sme sa zamerali na analýzu korelácií v prípade položiek merajúcich demokratické hodnoty. Pri týchto položkách by sme očakávali konzistentné pozitívne vzťahy s preferenciou liberálne demokratickej vlády a naopak negatívne vzťahy s konšpiračnými presvedčeniami (stĺpce 4. a 11.). Kým vzťahy s preferenciou liberálne demokratickej vlády sú viac-menej konzistentné pozitívne (hoci veľmi rôznorodej sily vzťahu), konšpiračné presvedčenia vykazujú nekonzistentné vzťahy s položkami tejto škály a dokonca prekvapivo, pozitívne korelujú s dvoma položkami tejto škály (položky 2 a 5).

Podobne je to aj v prípade položiek škály pravicového autoritárstva, kde by sme očakávali opačné smerovanie vzťahov s preferenciou liberálne demokratickej vlády a konšpiračnými presvedčeniami. Kým väčšina korelácií s preferenciou demokratickej vlády s týmito položkami je v očakávanom smere, dve sú štatisticky nevýznamné a jedna je dokonca pozitívna. Čo sa týka vzťahov s konšpiračnými presvedčeniami, tieto vzťahy sú ešte menej konzistentné. Iba dve položky pravicového autoritárstva signifikantne korelujú s konšpiračnými presvedčeniami, pričom jedna (položka 3) v očakávanom pozitívnom a druhá v prekvapivom negatívnom smere (položka 1).

Tabuľka 2. Korelácie premenných s jednotlivými položkami škály demokratických hodnôt a pravicového autoritárstva

	3.	4.	5.	6.	7.	8.	9.	10.	11.
Demokratické hodnoty:									
1. Slovenská republika musí poskytovať rovnaké spoločenské a politické práva všetkým obyvateľom, bez ohľadu na ich vierovyznanie, rasu, či pohlavie.	**.10**	**.26**	**–.14**	**.14**	**–.12**	**–.11**	**–.10**	**.02**	**–.16**
2. Otvorená a verejná kritika vlády by mala byť povolená aj v stave núdze.	**–.08**	**.09**	.00	**.36**	**.15**	**.15**	**.20**	**–.09**	**.16**
3. Aj najmenšie ohrozenie bezpečnosti štátu je dostatočným dôvodom na vážne obmedzenie demokracie. ®	–.01	**.11**	**–.23**	**–.08**	–.04	–.05	–.05	**–.14**	**–.12**
4. Uprednostnil/a by som demokratickú vládu, aj keď by som nesúhlasil/a s jej názormi a správaním pred nedemokratickou vládou, ktorá by presadzovala moje názory a hodnoty.	**.16**	**.21**	**–.07**	–.04	**–.12**	**–.12**	**–.12**	**.09**	**–.09**
5. Každý občan má právo verejne protestovať za svoje presvedčenia, ak je to potrebné.	–.04	**.12**	–.04	**.30**	**.15**	**.15**	**.18**	–.01	**.17**
6. Nie je potrebné, aby sa verejnosť podieľala na rozhodovaní, rozhodovanie stačí ponechať v rukách niekoľkých dôveryhodných lídrov. ®	**–.07**	**.18**	**–.12**	**.11**	.05	.05	**.07**	**–.12**	**–.09**
Pravicové autoritárstvo:									
1. Je skvelé, že mnoho mladých ľudí je v súčasnosti pripravených vzdorovať autorite. ®	–.05	–.02	–.04	**–.28**	**–.13**	**–.10**	**–.13**	–.02	**–.19**
2. Naša krajina najviac potrebuje disciplínu a aby všetci svorne poslúchali našich lídrov.	–.03	**–.16**	**.13**	**–.09**	–.02	–.04	**–.07**	**.12**	.02
3. Božie zákony ohľadom potratov, pornografie a manželstva je treba prísne dodržiavať než bude príliš neskoro.	**–.37**	**–.28**	**.27**	**.10**	**.28**	**.28**	**.24**	–.06	**.34**
4. Na predmanželskom pohlavnom styku nie je nič zlé. ®	**–.32**	**–.20**	.06	**–.21**	**.10**	**.08**	.04	–.01	.05
5. Naša spoločnosť nepotrebuje tvrdšiu vládu a prísnejšie zákony. ®	**–.08**	**–.08**	.02	–.05	–.00	–.02	–.02	**.08**	–.06
6. Údaje o trestných činoch a nedávne verejné nepokoje ukazujú, že ak máme zachovať zákon a poriadok, musíme silnejšie zasiahnuť proti tým, ktorí robia problémy.	**.12**	.05	.06	**.09**	**–.12**	**–.13**	**–.13**	**.19**	–.06

Poznámka. Tabuľka zobrazuje korelácie jednotlivých premenných (označených číslami podľa tabuľky 1) s jednotlivými položkami škály demokratických hodnôt a pravicového autoritárstva. Signifikantné korelácie sú označené hrubým písmom ($p < .05$). Reverzne skórované položky v rámci každej škály sú označené symbolom ®, pričom korelácie s inými premennými zobrazujú vzťahy po tomto reverznom-skórovaní (teda po transformácií odpovedí v týchto položkách).

Kým s ohľadom na vzťahy s ostatnými premennými sme nemali tak jasné predpoklady opierajúce sa o výsledky predchádzajúceho výskumu, niektoré zistenia nepochybne stoja za povšimnutie. Tu ide napríklad o významné pozitívne vzťahy politického cynizmu – ktorý môžeme chápať ako prejav voči politickému zriadeniu (Pattyn et al., 2012) – s demokratickými hodnotami a naopak jeho negatívne vzťahy s pravicovým autoritárstvom. Takisto premenné týkajúce sa ohrozenia identity a hodnôt, pri ktorých by sa dalo očakávať spojenie skôr s pravicovým autoritárstvom, vykazujú pozitívne vzťahy s položkami na demokratické hodnoty.

4 Diskusia a záver

V našom výskume sme sa zamerali na hlbšiu analýzu položiek využívaných v spoločenskovednom výskume na meranie kľúčových zložiek hodnotovej orientácie participantov – konkrétne demokratických hodnôt a pravicového autoritárstva. Táto hĺbková analýza bola motivovaná predchádzajúcimi snahami o podchytenie hodnotovej orientácie, ktoré sa stretli s problémami týkajúcimi sa meracích nástrojov zachytávajúcich hodnotovú orientáciu vo výskumoch so slovenskými vzorkami (Adamus et al., n.d.; Kanovský & Kocičová, 2018; Kostovičová et al., 2017). Podobne ako v týchto predchádzajúcich výskumoch, tak aj v našej štúdii sme narazili na viaceré problémy spojené s použitými metódami – jednak, nízku vnútornú konzistenciu použitých škál a po druhé, nekonzistentné vzťahy jednotlivých položiek v rámci daných škál s inými relevantnými premennými.

Zatiaľ čo je známe, že veľká časť výskumu v spoločenských vedách pochádza z rozvinutých a bohatých krajín globálneho západu (Henrich et al., 2010), táto analýza poukazuje na ďalšiu úroveň problému s možnými kultúrnymi rozdielmi vo výsledkoch spoločenskovedného výskumu týkajúcu sa špecifickosti nástrojov používaných na zachytenie hodnotovej orientácie. Tá môže mať v rôznych kultúrach (aj v rámci západných rozvinutých krajín) rôzne dimenzie a historické a politické kontexty v dôsledku čoho môžu položky zamerané na zachytávanie istého aspektu hodnotovej orientácie v jednej krajine zachytávať v skutočnosti úplne iný aspekt hodnotovej orientácie na Slovensku. Niektoré nájdené nečakané vzťahy pri znalosti slovenského kontextu prestávajú byť také prekvapujúce. Napríklad pozitívna korelácia medzi konšpiračnými presvedčeniami a položkami 2 (Otvorená a verejná kritika vlády by mala byť povolená aj v stave núdze.) a 5 (Každý občan má právo verejne protestovať za svoje presvedčenia, ak je to potrebné) demokratickej orientácie veľmi dobre odráža situáciu na Slovensku v čase pandémie COVID-19. Navyše, mnohé konšpiračné naratívy obsahujú výraznú kritiku vlády či vládnucich elít. a práve opozičné populistické a pravicové strany to využili na burcovanie ľudí k protestom voči pandemickým opatreniam a „hygienickému teroru“. A naopak, akceptácia obmedzenia demokratických práv a slobôd liberálnejšie orientovanými podporovateľmi vládnych opatrení odrážala pravdepodobne prinajmenšom do určitej miery frustráciu z infodémie a ovládnutia témy rôznymi pseudoodborníkmi a extrémistickými stranami, čo reflektuje negatívna korelácia medzi liberálnou orientáciou a položkami 2 a 6 (Nie je potrebné, aby sa verejnosť podieľala na rozhodovaní, rozhodovanie stačí ponechať v rukách niekoľkých dôveryhodných lídrov) dotazníka demokratickej orientácie, či v tomto kontexte viac pochopiteľnou pozitívnou koreláciou s položkou 5 dotazníka pravicového autoritárstva (Údaje o trestných činoch a nedávne verejné nepokoje ukazujú, že ak máme zachovať zákon a poriadok, musíme silnejšie zasiahnuť proti tým, ktorí robia problémy).

Napríklad, rôzna úroveň politickej polarizácie niektorých tém môže mať za následok to, že politická orientácia (a na ňu naviazané zložky hodnotovej orientácie, ako individualizmus, či pravicová orientácia atď.) môže vykazovať úplne iné vzťahy k premenným ako klimatický skepticizmus, či konšpiračné presvedčenia naprieč rôznymi krajinami (Hornsey et al., 2018).

Síce nie je v možnostiach tohto jedného príspevku priniesť riešenia týchto komplexných problémov, radi by sme poskytli aspoň niekoľko odporúčaní pre výskumníkov a výskumníčky venujúce sa problémom súvisiacim s meraním hodnotovej orientácie. V prvom rade, ako dokladujú naše výsledky, je dôležité venovať v prípade meracích nástrojov zameraných na hodnotovú orientáciu participantov, čas a pozornosť podrobnejším pschometrickým analýzam zameraným nie len na výpočet vnútornej konzistencie týchto škál, ale v prípade nízkej konzistencie sa tiež zamerať na jednotlivé položky a ich vzťahy s inými relevantnými premennými. Takýto postup umožňuje identifikovať položky v rámci daných škál, ktoré „fungujú“ podľa očakávaní. Samozrejme, korelácia s inými premennými nie je jediným kritériom a jej absencia automaticky neznamená, že položka je nevhodná pre danú škálu (no prinajmenšom by sme v škále neočakávali položky s významnými negatívnymi koreláciami s podobnými konštruktmi). A po druhé, našim odporúčaním je využívať dlhšie metodiky na zachytenie hodnotovej orientácie, ktoré je možné v prípade nevhodnosti vybraných položiek ďalej prispôsobiť, či vychádzať pri výbere metód z výskumov, ktoré boli realizované v podobnom spoločenskom a kultúrnom kontexte.

Poďakovanie

Tento príspevok vznikol s podporou grantovej agentúry APVV-20-0335 v rámci projektu „*Redukovanie šírenia dezinformácií a nepodložených presvedčení"* a s podporou agentúry VEGA 2/0053/21 v rámci projektu „*Skúmanie nepodložených presvedčení vo vzťahu ku kontroverzným spoločenským otázkam*".

Literatúra

Abalakina-Paap, M., Stephan, W. G., Craig, T., & Gregory, W. L. (1999). Beliefs in conspiracies. *Political Psychology*, *20*(3), 637-647. https://doi.org/10.1111/0162-895X.00160

Adamus, M., Šrol, J., Čavojová, V., & Ballová Mikušková, E. (2023). Seeing past the tip of your own nose? How outward and self-centred orientations could contribute to closing the green gap despite helplessness. *BMC psychology*, *11*(1), 79. https://doi.org/10.1186/s40359-023-01128-z

Adamus, M., Šrol, J., & Sobotová, B. (n.d.). Pro-environmental behaviors and its barriers. https://osf.io/5qy2b/

Bizumic, B., & Duckitt, J. (2018). Investigating right wing authoritarianism with a very short authoritarianism scale. *Journal of Social and Political Psychology*, *6*(1), 129–150. https://doi.org/10.5964/jspp.v6i1.835

Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy Mentality Questionnaire. *Frontiers in psychology, 4*, 225. https://doi.org/10.3389/fpsyg.2013.00225

Canetti-Nisim, D. (2004). The effect of religiosity on endorsement of democratic values: The mediating influence of authoritarianism. *Political Behavior, 26*, 377-398. https://doi.org/10.1007/s11109-004-0901-3

Cohrs, J. C., Kielmann, S., Maes, J., & Moschner, B. (2005). Effects of right-wing authoritarianism and threat from terrorism on restriction of civil liberties. Analyses of Social Issues and Public Policy, 5(1), 263-276. https://doi.org/10.1111/j.1530-2415.2005.00071.x

Crowson, H. M. (2009). Right-wing authoritarianism and social dominance orientation: As mediators of worldview beliefs on attitudes related to the war on terror. Social Psychology, 40(2), 93-103. https://doi.org/10.1027/1864-9335.40.2.93

Čavojová, V., Adamus, M., & Ballová Mikušková, E. (2022). You before me: How vertical collectivism and feelings of threat predicted more socially desirable behaviour during COVID-19 pandemic. *Current Psychology*. https://doi.org/10.1007/s12144-022-03003-3

Drummond, C., & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences*, *114*(36), 9587-9592. https://doi.org/10.1073/pnas.1704882114

Duckitt, J. (2022). Authoritarianism. Conceptualisation, Research, and New Developments. In: D. Osborne, Ch.G. Sibley (Eds), T*he Cambridge Handbok of Political Psychology*, Cambridge University Press, pp. 177-197.

Globsec. (2020). *Voices of Central and Eastern Europe: Perceptions of democracy & governance in 10 EU countries—GLOBSEC.* https://www.globsec.org/publications/voices-of-central-and-eastern-europe/

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences*, *33*(2-3), 61-83. https://doi.org/10.1017/S0140525X0999152X

Hornsey, M. J., Harris, E. A., & Fielding, K. S. (2018). Relationships among conspiratorial beliefs, conservatism and climate scepticism across nations. N*ature Climate Change*, *8*(7), 614-620. https://doi.org/10.1038/s41558-018-0157-2

Kanovský, M., & Kocičová, N. (2018). Ideological consistency and political polarization in Slovakia. *Human Affairs*, *28*(1), 44–53. https://doi.org/10.1515/humaff-2018-0005

Kostovičová, L., Bašnáková, J., & Bačová, V. (2017). Predicting Perception of Risks and Benefits within Novel Domains. *Studia Psychologica*, *59*(3), 176–192. https://doi.org/10.21909/sp.2017.03.739

Pattyn, S., Van Hiel, A., Dhont, K., & Onraet, E. (2012). Stripping the Political Cynic: A Psychological Exploration of the Concept of Political Cynicism. *European Journal of Personality*, *26*(6), 566–579. https://doi.org/10.1002/per.858

Šrol, J., Čavojová, V., & Ballová Mikušková, E.(2022). Finding someone to blame: The link between COVID-19 conspiracy beliefs, prejudice, support for violence, and other negative social outcomes. *Frontiers in psychology, 12*, 6390. https://doi.org/10.3389/fpsyg.2021.726076

van Mulukom, V., Pummerer, L. J., Alper, S., Bai, H., Čavojová, V., Farias, J., Kay, C. S., Lazarevic, L. B., Lobato, E. J. C., Marinthe, G., Banai, I. P., Šrol, J., & Žeželj, I. (2023). Antecedents and consequences of COVID-19 conspiracy beliefs: A systematic review. *Social Science & Medicine, 301*(January), 114912.https://doi.org/10.1016/j.socscimed.2022.114912

Čo chýba ChatGPT k tomu, aby rozumel, čo robí?

Martin Takáč

Centrum pre kognitívnu vedu FMFI UK
Mlynská dolina, 848 48 Bratislava
Email: martin.takac@fmph.uniba.sk

Abstrakt

Pre ChatGPT platí problém ukotvenia symbolov a všetky s ním spojené námietky. Napriek tomu nás ohromuje kvalitou vygenerovaných odpovedí a máme tendenciu veriť, že rozumie tomu, čo hovorí. Ale je to naozaj tak? V príspevku predstavím niekoľko pokusov ukotviť symboly veľkých jazykových modelov v multimodálnych dátach, resp. ich prepojiť s telom. Na záver naznačím disruptívne zmeny, ktoré systémy ako ChatGPT môžu spôsobiť, resp. už spôsobujú.

1 Úvod

V novembri 2022 dala spoločnosť OpenAI k verejnému používaniu systém ChatGPT. Ten okamžite pritiahol pozornosť médií aj používateľov – po prvých piatich dňoch ich bolo milión, v januári 2023 sto miliónov a dnes ich s ním denne interaguje približne 25 miliónov.[1] Dokáže konverzovať, písať básne aj eseje, riešiť matematické problémy, školské úlohy a testy, aj generovať kód v rôznych programovacích jazykoch. Súčasná (výrazne vylepšená) verzia s názvom GPT-4 sa v mnohých benchmarkových úlohách približuje k ľudskému výkonu (OpenAI, 2023). Výskumníci z Microsoftu vo svojom rozsiahlom reporte (Bubeck a spol., 2023) tvrdia, že GPT-4 možno považovať za prvotnú (zatiaľ neúplnú) verziu všeobecnej umelej inteligencie (AGI), teda umelej inteligencie, ktorá bez špeciálneho dotrénovania dokáže riešiť hocijakú úlohu, ktorú by vyriešil človek.

2 Ako to funguje

ChatGPT je veľký jazykový model (vo verzii GPT-4 obohatený o vizuálnu modalitu)—hlboká neurónová sieť s 10^{12} parametrami na báze transformera (Vaswani a spol., 2017), ktorý pre textový prompt na vstupe vygeneruje jeho pokračovanie (odpoveď). Odpovede hodnotené ľuďmi tvoria (spolu s promptom) trénovacie dáta pre reinforcement learning (RL) modul, ktorý sa naučí zoraďovať odpovede podľa vhodnosti.

[1] https://nerdynav.com/chatgpt-statistics/

3 Čo znamená „rozumieť“?

Umelé systémy sú tradične kritizované pre nedostatočné ukotvenie symbolov (Harnad, 1990): napriek tomu, že takýto systém môže správne reťaziť symboly, či odpovedať na otázku, pre neho sú ukotvené iba vo vzťahu k iným symbolom, ale nie k reálnemu svetu. Searle (1980) to opisuje metaforou Čínskej izby: Searle je zavretý v izbe, dostáva zvonka na papieri otázky v čínštine a pomocou veľkej knihy s pravidlami a porovnávania tvarov znakov podáva von na papieri správne odpovede v čínštine bez toho, aby rozumel po čínsky. Na argument Čínskej izby možno odpovedať viacerými spôsobmi (Cole, 2020). Jeden z nich je „The other minds reply“: o tom, že nám rozumejú iní ľudia, sa vieme presvedčiť iba pragmaticky, na základe ich správania. Ak správne zareagujú na našu požiadavku, tak jej rozumejú. Tento princíp tvorí základ mnohých variantov Turingovho testu: pokiaľ umelý systém v nejakej zložitej jazykovej úlohe uspeje ako človek, tak rozumie jazyku. GPT-4 by v mnohých ohľadoch Turingovým testom prešiel. Napriek tomu mu chýba ukotvenie symbolov a modely jeho typu boli označené za „stochastické papagáje“ (Bender a spol., 2021). Tento nedostatok sa moderné systémy snažia vyriešiť prepojením jazykovej domény s inými modalitami, najmä vizuálnou.

4 Multimodálne ukotvenie symbolov

Štandardným SOTA systémom sa stal CLIP (Radford a spol., 2021), ktorý prepája transformerový enkóder pre text s enkóderom pre obrázky. Je trénovaný metódou kontrastívneho učenia na pároch obrázok-text verejne dostupných z internetu. Oba enkódery sa trénujú na predikciu toho, ktoré obrázky boli spárované s akým textom. Následne sa system používa ako zero-shot klasifikátor, ktorý vie k obrázkom doplňať popisy.

V čase písania tohto článku bol predstavený multimodálny systém Kosmos-1 od Microsoftu (Huang a spol., 2023) zvládajúci tvorbu titulkov k obrázkom, odpovedanie na otázky o obrázkoch (VQA), rozpoznávanie entít v obrázkoch na základe textových inštrukcií, a ďalšie úlohy. V Ravenovom IQ teste (merajúcom schopnosť neverbálneho uvažovania) dosiahol 22 % úspešnosť bez dotrénovania oproti 17 % base-

line pri náhodnom hádaní. Systém pozostáva z transformerového dekódera trénovaného na obrovskom korpuse multimodálnych sekvencií (so špeciálnymi tagmi pre netextové data).

5 Prepojenie s telom

Ďalším (logickým) trendom je prepájanie jazykových modelov s robotickými telami. Microsoft (Vemprala a spol., 2023) reportuje o experimentoch, v ktorých dostal ChatGPT API pre detekciou objektov a vzdialeností a ovládanie robota, a textové popisy obrazu z kamery v každom časovom kroku. ChatGPT dokázal úspešne riadiť manipuláciu s objektami robotickým ramenom i ovládať dron.

Rovnakým smerom sa uberá aj výskum v Google. Článok (Driess a spol., 2023) predstavuje systém PaLM-E, ktorý pozostáva z dekodéra s 562 miliardami parametrov, na vstupe má multimodálne vety obsahujúce vizuálnu i textovú informáciu ako aj kontinuálne odhadovanie stavu systému. Je trénovaný end-to-end a zvláda VQA, titulkovanie obrázkov i riadenie robota.

6 Čínska izba ešte raz

Súčasné trendy v ukotvovaní jazykových modelov kopírujú princípy z niekoľkých ďalších odpovedí na argument čínskej izby. Podľa „System reply“ netreba postulovať v izbe Searla (rovnako ako v ľudskom mozgu sa nenachádza žiaden homunkulus), ale je to celý systém (izba, vstupy a výstupy, program a proces jeho vykonávania), ktorý rozumie čínsky. Podľa „Robot reply“ možno myšlienkový experiment pozmeniť tak, že Searle nebude v izbe, ale v robotovi, pričom vstupy prichádzajú z kamery a Searlove výstupy riadia efektory robota. A napokon, podľa „Developmental reply“ by v tomto robotovi nemusel byť dospelý Searle, ale novorodenec, ktorý si postupne osvojí a naučí sa všetky zákonitosti prepojení medzi vstupmi a výstupmi.

Ak teda dáme umelému systému telo a prepojíme jeho vstupy a výstupy na reálne prostredie a vybavíme ho silnými štatistickými mechanizmami učenia (napr. v hlbokej neurónovej sieti),[2] bude sa takýto systém principiálne líšiť od toho, ako funguje človek, jeho telo a jeho učiaci sa mozog?

7 Záver

V tomto článku som sa venoval najmä teoretickej otázke, v akom zmysle môžu systémy ako GPT-4 rozumieť tomu, čo robia. Tieto systémy so sebou však prinášajú aj množstvo urgentných praktických otázok. Doteraz bola kritizovaná najmä ich zaujatosť (biases), netransparentnosť a ekologické dopady (Bender a spol., 2021). Tým, že sa poskytli verejnosti v masovom meradle, sa stali disruptívnou technológiou, ktorá zmení obchodné modely fungovania, žurnalistiku, trh práce, vzdelávanie, nehovoriac o možnej destabilizácii demokracie, ak sa stanú nástrojom propagandy. Preto by bolo vhodné časť zdrojov, ktoré veľké firmy venujú na čo najrýchlejšie vytvorenie čo najvýkonnejšieho modelu, presmerovať do výskumu bezpečnostných aspektov takýchto modelov a ich dopadu na spoločnosť.

[2] Podstatným je zaväzbenie systému s prostredím v reálnom čase a rýchle (1-shot, few shot) a priebežné učenie, inak by vedomosti systému ostali zamrznuté v čase trénovania.

Poďakovanie

Vznik tohto článku bol podporený grantom VEGA 1/0373/23 a KEGA 022UK-4/2023.

Literatúra

Bender, E. M. a spol. (2021). On the dangers of stochastic parrots: Can language models be too big? V *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, str. 610–623.

Bubeck, S. a spol. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv:2303.12712 [cs.CL].

Cole, D. (2020). The Chinese Room Argument. Zalta, E. N. (zost.), V *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Driess, D. a spol. (2023). PaLM-E: An embodied multimodal language model. arXiv:2303.03378 [cs.LG].

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42:335–346.

Huang, S. a spol. (2023). Language is not all you need: Aligning perception with language models. arXiv:2302.14045 [cs.CL].

OpenAI (2023). GPT-4 technical report. arXiv:2303.08774 [cs.CL].

Radford, A. a spol. (2021). Learning transferable visual models from natural language supervision. arXiv:2103.00020 [cs.CV].

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3):417–457.

Vaswani, A. a spol. (2017). Attention is all you need. arXiv:1706.03762 [cs.CL].

Vemprala, S. a spol. (2023). ChatGPT for robotics: Design principles and model abilities. Technická správa MSR-TR-2023-8, Microsoft.

Stavová úzkosť po vystavení výškovej situácii vo virtuálnej realite – rola pocitov prítomnosti a stelesnenia

Varšová, Kristína[1]; Juřík, Vojtěch[1]; Janoušek, Oto[2]

[1]Psychologický Ústav, Masarykova Univerzita, Brno, Česká republika
[2]Ústav biomedicínského inženýrství, Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické Brno, Česká republika

kristinavars@gmail.com, jurik.vojtech@gmail.com, janouseko.vut.cz

Abstrakt

Virtuálna realita (VR) je technológia, ktorá v posledných rokoch prilákala pozornosť výskumníkov z oblasti psychológie, ktorí začali skúmať rozličné smery využitia v terapii fóbií. Jednou z najčastejšie liečených fóbií vo virtuálnom prostredí je akrofóbia - strach z výšok. Sľubná moderná technika kognitívne-behaviorálnej terapie (KBT) fóbií je expozičná terapia vo VR (VRET). Pre VRET je užívateľský zážitok dôležitým aspektom, obzvlášť v zmysle pocitov prítomnosti a stelesnenia. V našej štúdii bola zmapovaná VR expozícia vo výškovom prostredí u ľudí so stredným strachom z výšok. Hlavným cieľom štúdie bolo nájsť súvislosti medzi stavovou úzkosťou a pocitmi prítomnosti a stelesnenia, ktoré ovplyvňujú virtuálny zážitok.

1 Úvod

VR zaujala množstvo bežných užívateľov, ale aj vedcov a výskumníkov. Ani predstavitelia psychologického a psychoterapeutického výskumu nezaostávali a začali používať VR ako nákladovo efektívny, prístupný a užívateľsky prívetivý nástroj vo výskume (Jurik et al; 2018; Boeldt et al., 2019, Freeman et al., 2017). Jedným z najpoužívanejších terapeutických prístupov vo VR je VRET (orig. virtual reality exposure therapy), ktorá sa používa predovšetkým pri práci so špecifickými fóbiami (Emmelkamp & Meyerbröker; Chou et al., 2021). Rôzne aspekty VRET sú však stále len málo preskúmané, najmä pokiaľ ide o psychologické reakcie počas procesu expozície v rôznych prostrediach.

1.1 VR a psychické javy spojené s virtuálnym zážitkom

VR môže byť definovaná ako umelo vytvorené prostredie, ktoré vďaka pôsobeniu na takmer všetky zmysly, vytvára pocit prítomnosti a imerzie (Jerald, 2016). Táto technológia sprostredkúva simulovaný zážitok pocitu prítomnosti a fyzický svet je v nej nahradený virtuálnym prostredím. V kontexte VR je mimoriadne dôležitý používateľský zážitok, a to najmä pokiaľ ide o psychologické javy, ktoré tento zážitok sprevádzajú. Z tohto dôvodu sa mnohé štúdie začali zameriavať práve na túto tému a venujú sa často pocitu prítomnosti (Felnhofer et al, 2019; Pallavicini & Pepe, 2020), ktorý sa vzťahuje na pocit mentálneho prenesenia do virtuálneho prostredia (Smith & Mulligan, 2021; Heeter, 1992). Ďalším aspektom, o ktorom sa v súčasnosti diskutuje, je pocit stelesnenia, t.j. pocit vlastného tela, ktorý súvisí s vnímaním seba samého (Longo et al., 2008; Kilteni a Groten, 2012). Virtuálne stelesnenie v podstate znamená vytvorenie ilúzie prítomnosti v inom tele, než je to vlastné.

Jeden z najpodstatnejších účinkov prítomnosti vo VR pre prax je schopnosť vyvolať rovnaké emócie ako skutočný zážitok. Psychologický konštrukt prítomnosti vo VR sa vo všeobecnosti považuje za podstatný pre emočné reakcie vo virtuálnom prostredí, konkrétne aj pre strach alebo úzkosť (Witmer & Singer, 1998; Alsina-Jurnet et al., 2011). Niektoré výskumy ukazujú, že virtuálne prostredie dokáže zvyšovať subjektívne pocity úzkosti u fobických jedincov a opakovaná expozícia túto úzkosť dokáže znížiť (Regenbrecht et al.,1998; Ling et al., 2014). Výskumy potvrdzujúce vzťah prítomnosti a strachu alebo úzkosti taktiež naznačujú, že tento vzťah funguje aj opačne - participanti cítiaci strach z vystavovaného stimulu vo VR, môžu pociťovať väčšiu prítomnosť (Peperkorn et al., 2016; Bouchard et al., 2008; Gromer et al., 2019). Vyššie spomínaný vzťah prítomnosti a strachu však nebol potvrdený vždy a viacerí výskumníci toto spojenie na druhej strane popierajú (Wilhelm et al., 2005; Felnhofer et al., 2014; Slater, 2003), pričom napríklad tvrdia, že vysoká prítomnosť by mala byť braná ako vstup k emóciám, a teda u človeka sa musí vyskytovať určitá miera prítomnosti, aby mohol precítiť požadované emócie vyplývajúce z prostredia

VR. Intenzita emócii by však nemala byť priamo závislá na prítomnosti a akýkoľvek nájdený vzťah medzi týmito premennými môže byť podľa nich náhodný.
Zatiaľ čo existujúce štúdie skúmajúce pocity z VR vo vzťahu k úzkosti sú zamerané predovšetkým na mapovanie pocitu prítomnosti, pocit stelesnenia v asociácii s úzkosťou doposiaľ trpí nedostatkom výskumov. Výskum od Galla a kolegov (2021) naznačuje, že ilúzia stelesnenia môže zintenzívniť emocionálne spracovanie virtuálneho prostredia a naopak, no konkrétne pociťovaná úzkosť doteraz skúmaná nebola. Výsledky v oblasti vzťahu pocitov prítomnosti a stelesnenia sú každopádne nejednoznačné a vyžadujú si viac výskumnej pozornosti.

1.2 VRET ako nástroj pre terapiu akrofóbie

VR ponúka osobitý prostriedok na prenesenie ľudí do simulácií náročných situácií v bezpečnom prostredí (Freeman et al., 2017), a teda predkladá významný terapeutický potenciál. Fóbie sú typickou duševnou poruchou, pri ktorej sa VR využíva. Jednou z nich je akrofóbia, silný a nelogický strach z výšok, ktorý často koexistuje s inými duševnými problémami vrátane úzkostných porúch alebo depresie (Kapfhammer et al., 2016). Najobľúbenejším psychoterapeutickým prístupom pri akrofóbii je KBT, ktorá ovplyvňuje myslenie, emócie a správanie človeka (Beck, 2011). Najbežnejšou intervenciou KBT pri akrofóbii je expozičná terapia (Chou et al., 2021; Arroll et al, 2017). Expozícia je založená na myšlienke, že človek môže prekonať svoj strach tým, že sa s ním bude často stretávať. V nedávnej metaanalýze Choua a jeho kolegov (2021), ktorá porovnávala účinnosť a prijateľnosť rôznych prístupov k terapii akrofóbie boli výhody technológie VR osobitne zdôraznené. Dokonca aj nízkonákladová intervencia VR môže pomôcť zlepšiť negatívne príznaky akrofóbie, pretože virtuálne prostredie môže vyvolať realistické behaviorálne a fyziologické reakcie a účastníci v situácii VR vykazujú správanie, ktoré je podobné správaniu v reálnom prostredí (Kisker et al., 2021; Morina et al., 2015; Donker et al., 2019). Systematický prehľad a metaanalýza výskumov, ktorý uskutočnili Botella a kol. taktiež poskytol podporu pre používanie VRET (2017). Na základe mnohých výskumov možno teda tvrdiť, že VRET je bezpečnou a cenovo výhodnou alternatívou klasickej expozičnej terapie.

1.3 Ciele štúdie

Cieľom tejto štúdie bolo zistiť, či existuje vzťah medzi mierou úzkosti a pocitmi prítomnosti a stelesnenia v súvislosti s vystavením výškovej situácii vo VR u participantom so stredným strachom z výšok. Už niekoľko rokov sa výskumníci venujú vzťahu medzi pocitom prítomnosti a emóciami prežívanými pri vystavení obávanej situácii vo VR. Aj keď majorita potvrdzuje, ba dokonca naznačuje vzájomný vzťah medzi pocitom prítomnosti a subjektívnou úzkosťou prípadne strachom u fobických jedincov (Alsina-Jurnet et al., 2011; Gromer et al., 2019; Bouchard et al., 2008), názory nie sú jednomyseľné a tento efekt môže byť ovplyvnený použitými nástrojmi, ale aj typom poruchy (Wilhelm et al., 2005; Felnhofer et al, 2014; Ling et al., 2014). V prípade spojenia pocitu stelesnenia a úzkosti skrz koreláciu doposiaľ neboli uskutočnené obdobné výskumy, no bolo naznačené, že stelesnenie môže pôsobiť na pociťované emócie vo VR a prípadne ich zosilniť (Gall et al., 2021).

2 Metódy

V tejto štúdii bolo využité dotazníkové šetrenie na zabezpečenie relevantnej výskumnej vzorky a následné experimentálne stretnutie s vystavením participantov výškovej situácii vo VR. Súčasťou stanoveného cieľa bolo preskúmať dôležité faktory - pocity prítomnosti a stelesnenia, ktoré môžu značne ovplyvniť virtuálny zážitok (Gorisse et al., 2017). Preklad dotazníkov, doposiaľ nevyužívaných v českom prostredí, bol prevedený v súlade so štandardným postupom – boli vytvorené dva preklady z angličtiny do češtiny, následne prebehla ich syntéza a napokon spätný preklad do angličtiny pre overenie (Behling & Law, 2000). Vzhľadom na to, že zber dát prebiehal na území Českej republiky, všetky položky aj názvy subtestov v rámci popisu nástrojov uvádzame v českom jazyku Štúdia sa riadila potrebnými etickými normami a bola posúdená a schválená Etickým panelom na Psychologickom ústave Filozofickej fakulty Masarykovej univerzity v Brne. Účasť v štúdii bola dobrovoľná a všetkým účastníkom bol pred začiatkom experimentálneho sedenia predložený na prečítanie a podpis informovaný súhlas. Účastníci boli výslovne informovaní, že môžu zo štúdie kedykoľvek odstúpiť bez akýchkoľvek negatívnych dôsledkov. Pred začiatkom štúdie boli účastníkom podrobne vysvetlené všetky relevantné aspekty štúdie, ako aj návrh štúdie a spracovanie údajov.

2.1 Participanti

Cieľovou skupinou boli osoby s miernym strachom z výšok, ktoré boli vybrané na základe ich skóre v dotazníku HIQ (Heights Interpretation Questionnaire; Steinman & Teachman, 2011 – viac informácii nižšie). Výskumnú vzorku tvorilo 36 mladých dospelých (24 žien; 66,7 %) vo veku od 19 do 36 rokov (m = 24,6; SD = 4; med = 24). Potenciálnych záujemcov sme oslovili prostredníctvom rôznych skupín na sociálnej sieti Facebook a letákov rozmiestnených v okolí Filozofickej fakulty Masarykovej univerzity v Brne.

2.2 Materiály a nastavenie

Každé vystavenie sa uskutočnilo v Grey Lab na Psychologickom ústave Filozofickej fakulty Masarykovej univerzity v Brne. Účastníci dostávali vizuálne podnety pomocou zariadenia HTC Vive Pro. Na reprezentáciu virtuálneho výškového prostredia sa použila aplikácia Richie's Plank Experience (RICHIE'S PLANK EXPERIENCE). V tejto aplikácii sa používatelia ocitli v rušnom centre mesta, kde sa nachádza vysoký 80-poschodový mrakodrap. Výťah vyvezie jednotlivca na najvyššie poschodie, kde ho po otvorení dverí čaká drevená doska smerujúca von z výťahu. Aby sme vytvorili čo najväčší pocit reálnosti aplikácie, rozhodli sme sa umiestniť na zem skutočnú drevenú dosku, podľa ktorej veľkosti bola presne kalibrovaná virtuálna doska, umiestnená paralelne so skutočnou. Okrem hluku v pozadí mesta stimuluje sluchovú modalitu aj zvuk praskajúcej virtuálnej dosky pri šliapaní na ňu. S cieľom uvoľniť účastníkov po vystavení výškovej situácii vo VR sme sa rozhodli použiť aplikáciu Guided Meditation VR (Guided Meditation VR).

2.3 Nástroje

2.3.1 Heights Interpretation Questionnaire (HIQ)

HIQ (Steinman a Teachman, 2011) je dotazník, ktorý sa skladá zo 16 položiek. Škála tohto dotazníka predpovedá strach, úzkosť a vyhýbavé správanie v prípade pobytu vo výškovom prostredí. HIQ sa v tejto štúdii použil ako kritérium zaradenia aj vylúčenia z účasti a ako dôležitý ukazovateľ strachu z výšok. Do štúdie mohli byť zaradení len účastníci, ktorí dosiahli skóre približne od 26 do 55 bodov, čo naznačuje mierny strach z výšok a je v súlade s výberovými kritériami iných štúdií (Freeman et al., 2018; Arroll et al., 2017). Dotazník HIQ má vysokú vnútornú konzistenciu, konvergentnú platnosť s inými nástrojmi merajúcimi strach z výšok a vysokú spoľahlivosť (Cronbachova alfa=0,91).

2.3.2 State-Trait Anxiety Inventory (STAI)

STAI (Spielberger, 1989) predstavuje najpoužívanejšiu metódu merania úzkosti a často sa používa na meranie aj v kontexte VR expozície (Ling et al., 2014). Tvorí 2 subtesty, pozostávajúce z 20 položiek. Na účely nášho výskumu sa použil subtest zameraný na stavovú úzkosť (STAI-Y1), aby sa zachytili bezprostredné pocity po vystavení sa výškovej situácii. Dotazník STAI-Y1 bol použitý z 2 dôvodov: a) inšpirácie z predchádzajúcich štúdií s podobným zámerom merať stavovú úzkosť vo VR (Concannon et al., 2020; Felnhofer et al., 2014); b) z dôvodu nedostatku relevantných dotazníkov zachytávajúcich aktuálnu úroveň strachu. Koeficient Cronbachovej alfy predstavuje v prípade STAI 0,92.

2.3.3 Embodiment Rating Scale

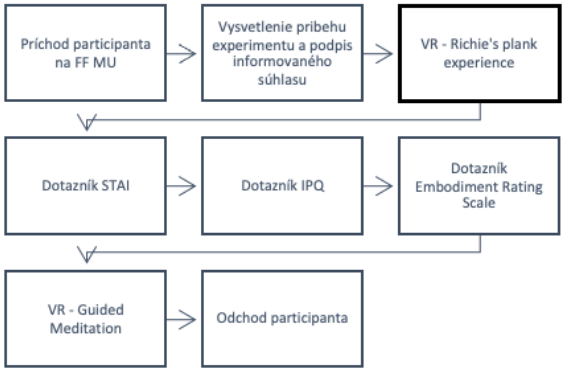
Embodiment Rating Scale, založený na práci Longa a kolegov (2008) a bol použitý vo výskume Pritcharda a kolegov (2016). Tento dotazník pozostáva 11 otázok rozdelených do troch subškál - Vliv, Vlastnictví a Umístění. Na základe posúdenia vhodnosti položiek v kontexte nášho výskumu sme sa rozhodli odstrániť jednu položku zo subtestu vlastníctva, a teda konečný počet použitých otázok bol 10.

2.3.4 Igroup Presence Questionnaire

Cieľom dotazníku IPQ (Schubert et al., 2001; Regenbrecht & Schubert, 2002; Schubert, 2003) je zachytiť pocit prítomnosti, a teda ako veľmi participanti pociťujú, že sa nachádzajú vo virtuálnom prostredí. Pozostáva zo 14 položiek rozdelených do štyroch podskupín – Prostorová přítomnost, Angažovanost, Všeobecná přítomnost a Reálnost. IPQ je považovaný za reliabilný nástroj pri skúmaní pociťovanej prítomnosti vo virtuálnom prostredí (α = .87).

2.4 Procedúra

Účastníci boli požiadaní o vyplnenie dotazníka HIQ a na základe dosiahnutých výsledkov boli pozvaní na experimentálne stretnutie. Priebeh hlavných častí experimentu znázorňujeme prostredníctvom Obr. 1.



Obr. 1: Schéma priebehu experimentálneho stretnutia

2.5 Analýza dát

K analýze dát bol použitý program SPSS. V prvej fáze analýzy bolo vypočítané celkové skóre dotazníkov HIQ a STAI, a to súčtom položiek. Následne bolo potrebné vypočítať osobitne skóre pre všetky subškály IPQ a Embodiment Rating Scale podľa postupov autorov. Výsledné skóre všetkých štyroch subškál IPQ sme získali na základe priemerného skóre položiek spadajúcich pod konkrétnu subškálu. Rovnaký postup

bol opakovaný pri výpočte subškál Embodiment Rating Scale.

3 Výsledky

Pre overenie vzťahov stavovej úzkosti a oboch premenných (pocity prítomnosti a stelesnenia) bola použitá Spearmanova korelácia. Prvá korelácia bola prevedená medzi hodnotami dotazníka STAI a subškálami dotazníka pocitu prítomnosti IPQ – Všeobecná přítomnost (IPQVP), Prostorová přítomnost (IPQPP), Angažovanost (IPQA) a Reálnost (IPQR). Medzi celkovou úrovňou stavovej úzkosti a subškálami IPQ u participantov so strachom z výšok, v súvislosti s expozíciou vo VR výškovej situácii, nebol nájdený žiadny štatisticky významný vzťah (Tab. 1).

Premenná		STAI	IPQVP	IPQPP	IPQA	IPQR
1. STAI	Spearman's rho	—				
	p-hodnota	—				
2. IPQVP	Spearman's rho	0.258	—			
	p-hodnota	0.129	—			
3. IPQPP	Spearman's rho	0.030	0.545 ***	—		
	p-hodnota	0.864	< .001	—		
4. IPQA	Spearman's rho	-0.032	0.532 ***	0.300	—	
	p-hodnota	0.854	< .001	0.076	—	
5. IPQR	Spearman's rho	0.244	0.532 ***	0.499 **	0.410 *	—
	p-hodnota	0.151	< .001	0.002	0.013	—

* p < .05, ** p < .01, *** p < .001

Tab. 1: Korelácia premennej stavovej úzkosti (STAI) a subškál pocitu prítomnosti (IPQ)

V prípade druhej časti bola prevedená korelácia medzi hodnotami dotazníka STAI a subškálami dotazníka pocitu stelesnenia (ERS) – Vlastnictví (ERSVL), Umístění (ERSUM) a Vliv (ERSVLIV). Jediný štatisticky významný vzťah bol nájdený medzi úrovňou stavovej úzkosti meranou pomocou STAI a subškálou Vlastnictví (ERSVL). Tento vzťah bol pozitívny a približne stredne silný (r=.342; p<.05). Korelácie premenných STAI a subškál ERS sú zobrazené v Tab. 2.

Premenná		STAI	ERSVL	ERSUM	ERSVLIV
1. STAI	Spearman's rho	—			
	p-hodnota	—			
2. ERSVL	Spearman's rho	**0.342 ***	—		
	p-hodnota	0.041	—		
3. ERSUM	Spearman's rho	0.185	0.559 ***	—	
	p-hodnota	0.281	< .001	—	
4. ERSVLIV	Spearman's rho	-0.131	0.251	0.395 *	—
	p-hodnota	0.445	0.140	0.017	—

* p < .05, ** p < .01, *** p < .001

Tab. 2: Korelácia premennej stavovej úzkosti (STAI) a subškál pocitu stelesnenia (ERS)

4 Diskusia

Hlavným cieľom tejto štúdie bolo nájsť súvislosti medzi stavovou úzkosťou a pocitmi prítomnosti a stelesnenia u participantov so strachom z výšok v súvislosti s výškovou situáciou vo VR.

Stavová úzkosť bola zisťovaná pomocou subjektívneho hodnotenia položiek dotazníka STAI a zisťované boli vzťahy k pocitu prítomnosti (meranému dotazníkom IPQ) a pocitu stelesnenia (meranému dotazníkom ERS). V prvom rade sa zameriame na interpretáciu vzťahu stavovej úzkosti a prítomnosti. Medzi úzkosťou a subškálami z dotazníka IPQ nebol nájdený žiadny signifikantný vzťah, čo je v opozícii voči väčšiemu počtu výsledkov štúdií. Podľa našich výsledkov sa teda prikláňame skôr k názoru Slatera (2003) a výskumným záverom Felnhofer a jej kolegov (2014) či Wilhelma et al., (2005) a naznačujeme, že určitá miera prítomnosti môže predstavovať len predpoklad vzniku emócii, no ich intenzita nemusí byť závislá na úrovni pociťovanej prítomnosti. Zároveň je nutné dodať, že vo výskume Felnhofer et al. (2014) boli použité rovnaké nástroje na meranie úzkosti (STAI) a pocitu prítomnosti (IPQ) ako v našej štúdii. Pre zistený nevýznamný vzťah stavovej úzkosti a prítomnosti v rámci našej štúdie môže existovať aj iné vysvetlenie. Napríklad aj predchádzajúca skúsenosť s VR, ktorá v našej štúdii nebola zisťovaná, môže tiež ovplyvniť pociťovanú prítomnosť participantov (Freeman, 1999; Sagnier et al., 2020). V rámci ďalšieho skúmania by teda mala byť aj táto premenná zahrnutá do výskumov. V druhom rade sme sa zamerali na vzťah stavovej úzkosti a pocitu stelesnenia. Vzhľadom na to, že náš zámer bol pomerne ojedinelý a nemali sme empirické teoretické základy, mohli sme len predpokladať, že v súlade s výskumnými závermi Galla a kol. (2021), bude emócia, a teda v našom prípade úzkosť, súvisieť s pociťovanou

prítomnosťou. Jediný štatisticky významný vzťah medzi stavovou úzkosťou a zisťovanými komponentami stelesnenia v rámci dotazníka ERS bol nájdený v prípade subškály Vlastnictví. V tejto subškále boli položky zamerané predovšetkým na to, či bol ovládač vnímaný ako súčasť tela participantov. Keďže nemáme konkrétne teoretické východiská, z ktorých by sa dali usudzovať súvislosti, môžu naše zistenia slúžiť ako vstupná brána k ďalšiemu skúmaniu. Nájdený signifikantný vzťah medzi stavovou úzkosťou a subškálou zameranou na vlastníctvo tela mohol byť náhodný. Mohol by však vypovedať napríklad o tom, že pre emócie môže byť podstatné to, ako sa cítime byť „prenesení" do virtuálneho avatara. Taktiež v našej štúdii nachádzame paralelu s výskumom Gall a kolegov (2021), v ktorom sa ukázal vzťah emócií a stelesnenia. V ňom totiž boli použité len dve položky na zistenie úrovne stelesnenia, pričom znenie nimi použitých položiek bolo veľmi blízke zneniu položiek zo subškály Vlastnictví, ktorá ako jediná vyšla signifikantne. Pri interpretácii všetkých prezentovaných výsledkov je však potrebné zobrať do úvahy limity, ktoré mohli vzťahy skúmaných premenných ovplyvniť. Za hlavné obmedzenie štúdie považujeme malú výskumnú vzorku, ktorá bola spôsobená zvýšenými logistickými a technickými požiadavkami štúdie v čase pandémie a ktorá môže čiastočne znížiť spoľahlivosť vykonaných štatistických záverov. Taktiež predpokladáme, že rozsah tejto štúdie bol obmedzený v kontexte terapeutického procesu, kde významnú úlohu zohráva špecifický dlhodobý vzťah medzi klientom a psychológom. Ako posledný bod uvádzame, že aj keď je zážitok z VR zvyčajne vzrušujúci, pre niekoho môže byť aj rušivý. Tento faktor považujeme za obmedzenie, ktoré by sa malo v následnom výskume lepšie kontrolovať.

Poďakovanie

Výskum bol realizovaný s podporou Psychologického ústavu Masarykovej Univerzity v Brne.

Literatúra

Alsina-Jurnet, I., Gutiérrez-Maldonado, J., & Rangel-Gómez, M.-V. (2011). The role of presence in the level of anxiety experienced in clinical virtual environments. *Computers in Human Behavior, 27*(1), 504–512. https://doi.org/10.1016/j.chb.2010.09.018

Arroll, B., Wallace, H. B., Mount, V., Humm, S. P., & Kingsford, D. W. (2017). A systematic review and meta analysis of treatments for acrophobia. *Medical Journal of Australia, 206*(6), 263-267. https://doi.org/10.5694/mja16.00540

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond.* (2nd ed.). Guilford Press.

Behling, O., & Law, K. S. (2000). Translating questionnaires and other research instruments: Problems and solutions. Thousand Oaks. Sage.

Boeldt, D., McMahon, E., McFaul, M., & Greenleaf, W. (2019). Using Virtual Reality Exposure Therapy to Enhance Treatment of Anxiety Disorders: Identifying Areas of Clinical Adoption and Potential Obstacles. *Frontiers in Psychiatry, 10*. https://doi.org/10.3389/fpsyt.2019.00773

Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., & Baños, R. (2017). Recent Progress in Virtual Reality Exposure Therapy for Phobias: A Systematic Review. *Current Psychiatry Reports, 19*(7). https://doi.org/10.1007/s11920-017-0788-4

Bouchard, S., St-Jacques, J., Robillard, G., & Renaud, P. (2008). Anxiety Increases the Feeling of Presence in Virtual Reality. *Presence: Teleoperators and Virtual Environments, 17*(4), 376-391. https://doi.org/10.1162/pres.17.4.376

Concannon, B. J., Esmail, S., & Roduta Roberts, M. (2020). Immersive Virtual Reality for the Reduction of State Anxiety in Clinical Interview Exams: Prospective Cohort Study. *JMIR Serious Games, 8*(3). https://doi.org/10.2196/18313

Donker, T., Cornelisz, I., van Klaveren, C., van Straten, A., Carlbring, P., Cuijpers, P., & van Gelder, J. -L. (2019). Effectiveness of Self-guided App-Based Virtual Reality Cognitive Behavior Therapy for Acrophobia: A Randomized Clinical Trial. *JAMA Psychiatry, 76*(7). https://doi.org/10.1001/jamapsychiatry.2019.0219

Emmelkamp, P. M. G., & Meyerbröker, K. (2021). Virtual Reality Therapy in Mental Health. *Annual Review of Clinical Psychology, 17*(1), 495-519. https://doi.org/10.1146/annurev-clinpsy081219-115923

Felnhofer, A., Hlavacs, H., Beutl, L., Kryspin-Exner, I., & Kothgassner, O. D. (2019). Physical Presence, Social Presence, and Anxiety in Participants with Social Anxiety Disorder During Virtual Cue Exposure. *Cyberpsychology, Behavior, and Social Networking, 22*(1), 46-50. https://doi.org/10.1089/cyber.2018.0221

Felnhofer, A., Kothgassner, O. D., Hetterle, T., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2014). Afraid to Be There? Evaluating the Relation Between Presence, Self-Reported Anxiety, and Heart Rate in a Virtual Public Speaking Task. *Cyberpsychology,*

Behavior, and Social Networking, 17(5), 310-316. https://doi.org/10.1089/cyber.2013.0472

Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., Denne, M., Brown, P., Slater, M., & Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry, 5*(8), 625-632. https://doi.org/10.1016/S2215-0366(18)30226-8

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine, 47*(14), 2393- 2400. https://doi.org/10.1017/S003329171700040X

Freeman, J., Avons, S. E., Pearson, D. E., & IJsselsteijn, W. A. (1999). Effects of Sensory Information and Prior Experience on Direct Subjective Ratings of Presence. *Presence: Teleoperators and Virtual Environments, 8*(1), 1-13. https://doi.org/10.1162/105474699566017

Gall, D., Roth, D., Stauffert, J. -P., Zarges, J., & Latoschik, M. E. (2021). Embodiment in Virtual Reality Intensifies Emotional Responses to Virtual Stimuli. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.674179

Gorisse, G., Christmann, O., Amato, E. A., & Richir, S. (2017). First- and Third-Person Perspectives in Immersive Virtual Environments: Presence and Performance Analysis of Embodied Users. *Frontiers in Robotics and AI, 4*. https://doi.org/10.3389/frobt.2017.00033

Gromer, D., Reinke, M., Christner, I., & Pauli, P. (2019). Causal Interactive Links Between Presence and Fear in Virtual Reality Height Exposure. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.00141

Guided Meditation VR. Steam. Prevzaté Marec 8, 2023, z https://store.steampowered.com/app/397750/Guided_Meditation_VR/

Heeter, C. (1992). Being There: The Subjective Experience of Presence. *Presence: Teleoperators and Virtual Environments, 1*(2), 262-271. https://doi.org/10.1162/pres.1992.1.2.262

Chou, P. -H., Tseng, P. -T., Wu, Y. -C., Chang, J. P. -C., Tu, Y. -K., Stubbs, B., Carvalho, A. F., Lin, P. -Y., Chen, Y. -W., & Su, K. -P. (2021). Efficacy and acceptability of different interventions for acrophobia: A network meta-analysis of randomised controlled trials. *Journal of Affective Disorders, 282*, 786-794. https://doi.org/10.1016/j.jad.2020.12.172

Chou, P. -H., Tseng, P. -T., Wu, Y. -C., Chang, J. P. -C., Tu, Y. -K., Stubbs, B., Carvalho, A. F., Lin, P. -Y., Chen, Y. -W., & Su, K. -P. (2021). Efficacy and acceptability of different interventions for acrophobia: A network meta-analysis of randomised controlled trials. *Journal of Affective Disorders, 282*, 786-794. https://doi.org/10.1016/j.jad.2020.12.172

Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.

Juřík, V., Herman, L., Šašinka, Č., Stachoň, Z., Chmelík, J., Strnadová, A., ... Bandrova, T., & Konečný, M. (2018). Behavior Analysis in Virtual Geovisualizations: Towards Ecological Validity. In T. Bandrova & M. Konečný (Eds.), *Proceedings of the 7th International Conference on Cartography and GIS* (pp. 518-527). Sofia, Bulgaria: Bulgarian Cartographic Association.

Kapfhammer, H. -P., Fitz, W., Huppert, D., Grill, E., & Brandt, T. (2016). Visual height intolerance and acrophobia: distressing partners for life. *Journal of Neurology, 263(*10), 1946-1953. https://doi.org/10.1007/s00415-016-8218-9

Kilteni, K., Groten, R., & Slater, M. (2012). The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments, 21*(4), 373-387. https://doi.org/10.1162/PRES_a_00124

Kisker, J., Gruber, T., & Schöne, B. (2021). Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure. *Psychological Research, 85*(1), 68-81. https://doi.org/10.1007/s00426-019-01244-9

Ling, Y., Nefs, H. T., Morina, N., Heynderickx, I., Brinkman, W. -P., & Slater, M. (2014). A Meta-Analysis on the Relationship between Self-Reported Presence and Anxiety in Virtual Reality Exposure Therapy for Anxiety Disorders. *PLoS ONE, 9*(5). https://doi.org/10.1371/journal.pone.0096144

Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., & Haggard, P. (2008). What is Embodiment? A Psychometric Approach. *Cognition, 107*(3), 978-998. https://doi.org/10.1016/j.cognition.2007.12.004

Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P. M. G. (2015). Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. *Behaviour Research and Therapy, 74*, 18-24. https://doi.org/10.1016/j.brat.2015.08.010

Pallavicini, F., & Pepe, A. (2020). Virtual Reality Games and the Role of Body Involvement in Enhancing Positive Emotions and Decreasing Anxiety: Within-Subjects Pilot Study. *JMIR Serious Games, 8*(2). https://doi.org/10.2196/15635

Peperkorn, H. M., Diemer, J. E., Alpers, G. W., & Mühlberger, A. (2016). Representation of Patients' Hand Modulates Fear Reactios of Patients with Spider Phobia in Virtual Reality. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00268

Pritchard, S. C., Zopf, R., Polito, V., Kaplan, D. M., & Williams, M. A. (2016). Non-hierarchical Influence of Visual Form, Touch, and Position Cues on Embodiment, Agency, and Presence in Virtual Reality. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01649

Regenbrecht, H. T., Schubert, T. W., & Friedmann, F. (1998). Measuring the Sense of Presence and its Relations to Fear of Heights in Virtual Environments. *International Journal of Human-Computer Interaction*, *10*(3), 233-249. https://doi.org/10.1207/s15327590ijhc1003_2

Regenbrecht, H., & Schubert, T. (2002). Real and Illusory Interactions Enhance Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *11*(4), 425-434. https://doi.org/10.1162/105474602760204318

RICHIE´S PLANK EXPERIENCE. TOAST Games. Prevzaté Marec 7, 2023, z https://toast.games/

Sagnier, C., Loup-Escande, E., & Valléry, G. (2020). Effects of Gender and Prior Experience in Immersive User Experience with Virtual Reality. *Advances in Usability and User Experience,* 305-314. https://doi.org/10.1007/978-3-030-19135-1_30

Schubert, T. W. (2003). The sense of presence in virtual environments. *Zeitschrift für Medienpsychologie*, *15*(2), 69-71. https://doi.org/10.1026//1617-6383.15.2.6982

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266-281. https://doi.org/10.1162/105474601300343603

Slater, M. (2003). *A Note on Presence Terminology*. ResearchGate. Prevzaté Marec 27, 2023, z https://www.researchgate.net/publication/242608507_A_Note_on_Presence_Terminology

Smith, S. A., & Mulligan, N. W. (2021). Immersion, presence, and episodic memory in virtual reality environments. *Memory, 29*(8), 983-1005. https://doi.org/10.1080/09658211.2021.1953535

Spielberger, C. D. (1989). *State-Trait Anxiety Inventory: Bibliography (2nd ed.)*. Consulting Psychologists Press.

Steinman, S. A., & Teachman, B. A. (2011). Cognitive processing and acrophobia: Validating the Heights Interpretation Questionnaire. *Journal of Anxiety Disorders*, *25*(7), 896-902. https://doi.org/10.1016/j.janxdis.2011.05.001

Wilhelm, F. H., Pfaltz, M. C., Gross, J. J., Mauss, I. B., Kim, S. I., & Wiederhold, B. K. (2005). Mechanisms of Virtual Reality Exposure Therapy: The Role of the Behavioral Activation and Behavioral Inhibition Systems. *Applied Psychophysiology and Biofeedback*, *30*(3), 271-284. https://doi.org/10.1007/s10484-005-6383-1

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225-240.

Robotická manipulace pomocí sekvence neurálních modulů s vlastní policy

Michal Vavrečka, Jonas Kříž, Nikita Sokovnin, Gabriela Šejnová

CIIRC ČVUT
Jugoslávských Partyzánů 3, Praha
michal.vavrecka@cvut.cz

Abstrakt

Vyvinuli jsme nový multi-policy algoritmus (MultiPPO2), který zlepšuje učení posilováním v úlohách, kde jsou vyžadovány různé dílčí cíle (dovednosti). Algoritmy s jednou policy často uspějí pouze v úlohách vyžadujících osvojení podobných nerůznorodých dovedností. Vzhledem k tomu, že většina současných robotických testů je založena na přemisťování objektů ve scéně, řeší algoritmy s jednou policy tyto úlohy s vysokou úspěšností. Pokud je vyžadována posloupnost různorodých dílčích cílů (translace, rotace, manipulace v 6DOF, sledování trajektorie), síť s jednou policy selže. Navrhujeme algoritmus s více policy pro každou dovednost. Je schopen naučit se vícekrokovou manipulaci s různými dílčími cíli. Náš algoritmus jsme testovali ve virtuálním robotickém simulátoru jak na jednokrokových i vícekrokových úlohách vyžadujících nerůznorodé (Pick and Place) i různorodé dovednosti (Pick and Rotate a Swipe). Náš multi-policy algoritmus dosáhl podobného výkonu v případě nerůznorodých dovedností a překonal single-policy algoritmy v úlohách vyžadujících různorodou sadu dovedností. Výsledek dokazuje, že algoritmus s více policy nabízí lepší škálovatelnost pro složité úlohy. Potenciál nabízí v kombinaci s hierarchickým posilovacím učením.

1 Úvod

S využitím posilovaného učení pro robotické manipulační úlohy je spojeno několik klíčových problémů. Jedním z nich je, že robot se musí naučit provádět úkoly zahrnující rozsáhlé sekvence akcí, jako je například "pick-and-place". Použití tradičních metod RL pro učení složité úlohy a single-policy je neefektivní, protože vyžaduje, aby agent prozkoumal rozsáhlý prostor stavů a akcí a shromáždil mnoho dat, než obdrží smysluplný učební signál.

V porovnání s jednokrokovými úlohami vyžadují komplexní úlohy ve většině reálných problémů více transformací k dosažení cílového stavu. Tento problém lze odstranit použitím algoritmů s několika policy. Ty jsou založeny na modulech odpovídajících základním schopnostem (dovednostem) řešit komplexní úlohy Andreas a spol. (2017). Takové policy netrpí zapomínáním a nabízejí lepší interpretovatelnost Alet a spol. (2018).

Devine a spol. použili modulární neuronové sítě, kde každý modul představuje samostatné dovednosti Devin a spol. (2017). Modulární sítě jsou sítě s více policy, které lze kompozičně řetězit, takže systém je schopen řešit složitější úlohy. Pore a kol. Pore a Aragon-Camarasa (2020) používají behaviorální klonování k trénování jednotlivých modulů dílčích úloh a hierarchicky vyšší modul, který je spojuje dohromady. Nevyřešeným problémem je způsob řazení modulů.

Někteří autoři Marzari a spol. (2021) řadí moduly, které mají být aktivovány, jeden po druhém a někteří Dalal a spol. (2021) zahrnují parametry pro konkrétní dílčí cíle. Způsob orchestrace modulů Plappert a spol. (2018) je stále nevyřešen. Objevily se přístupy k návrhu obecného sekvenceru Clegg a spol. (2018), ale jeho univerzálnost je stále omezená.

Protože bychom rádi přispěli k řešení tohoto problému, navrhujeme nový algoritmus optimalizace několika policy (MultiPPO2) a testujeme jej na jedno- i vícekrokových úlohách s různými dílčími cíli. Jeho výsledky překonávají tradiční algoritmy s jednou policy. Hlavní přínosy našeho přístupu jsou následující:

- Implementovali jsme algoritmus, který je schopen trénovat několik policy v rámci jedné epizody a přepínat mezi nimi během testování.

- Trénovali jsme a testovali náš algoritmus MultiPPO2 ve virtuálním robotickém simulátoru, abychom prokázali jeho lepší výkonnost ve srovnání s metodami s jednou policy.

2 Metody

Náš přístup se zaměřuje na robotické úlohy, které lze rozdělit na menší podúlohy. Trénování urychlíme tím, že trénujeme všechny dílčí úlohy za sebou. Pro podrobné srovnání tréninkových algoritmů jsme vyvinuli nové benchmarky, které jsou založeny na pokročilé robotické manipulaci, konkrétně na rotaci objektu v 6DOF (Pick and Rotate) a následování trajektorie při manipulaci s objektem (Swipe).

2.1 Prostředí

Naše experimenty provádíme pomocí toolboxu pro trénování a testování myGym Vavrecka a spol. (2021). Je postaven na fyzikálním enginu Pybullet. Používáme dva modely skutečných průmyslových robotů - Kuka IIWA (Kuka) a Franka Emika Panda (Panda). Používáme kloubové úhlové řízení robotů. Každý robot má 7 pohyblivých kloubů. K manipulaci s předmětem používáme magnetické chapadlo.

2.1.1 MultiPPO2

Vyvinuli jsme algoritmus, který dokáže v rámci jedné epizody trénovat více policy a přepínat mezi nimi na základě průběhu řešení úlohy. Vycházíme z algoritmu PPO2 Schulman a spol. (2017) z knihovny Stable Baselines Hill a spol. (2018). V porovnání se standardním PPO2 se v MultiPPO2 během tréninku instancují samostatné moduly. Úlohy, které se snažíme trénovat, rozdělíme na několik jednodušších podúloh a pro každou z nich napíšeme odměnu. Model MultiPPO2 při inicializaci vytvoří dílčí model pro každou úlohu a spáruje je s odpovídající funkcí odměny. Trénování modelu MultiPPO2 je popsáno v části Alg. 1 a probíhá podobně jako u modelu PPO2.

Algorithm 1 MultiPPO2

```
function TRAIN_MODEL(M, reward)
    M = "number of desired models"
    reward = "reward functions array"
    Model.submodels = [ ]
    for m in M do
        models[m] = PPO2_model()
    end for
    for i in epochs do
        id = decide(observation)
        id = ïd of appropriate submodel"
        train_batch(Model.submodels[id], reward[id])
    end for
    return Model
end function
```

Druhým novým krokem je zavedení predikčního modulu. Systém na základě aktuálního pozorování rozhodne, jakou dílčí úlohu se síť právě snaží splnit. Podle rozhodnutého úkolu se na této části dat cvičí odpovídající model s odměnou náležející danému úkolu. Při použití natrénovaného modelu MultiPPO2 pro testování se postupuje obdobně jako při trénování, jak je uvedeno v Alg. 2.

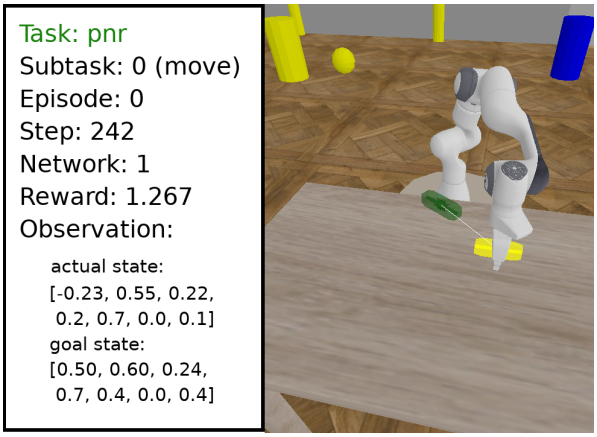
2.2 Úlohy

Vyvinuli jsme novou sadu úloh v rámci myGym toolboxu, abychom porovnali náš algoritmus MultiPPO2 s

Algorithm 2 MultiPPO2 predictions

```
function PREDICT(Model, observation)
    id = decide(observation)
    id = ïd of appropriate submodel"
    return Model.submodels[id].predict(observation)
end function
```

PPO, PPO2 Schulman a spol. (2017) a ACKTR Wu a spol. (2017). Tyto single-policy algoritmy již byly vyhodnoceny v rámci myGym toolboxu Vavrecka a spol. (2021) na úlohách bez dílčích cílů (reach, pick, push) s vysokou úspěšností.



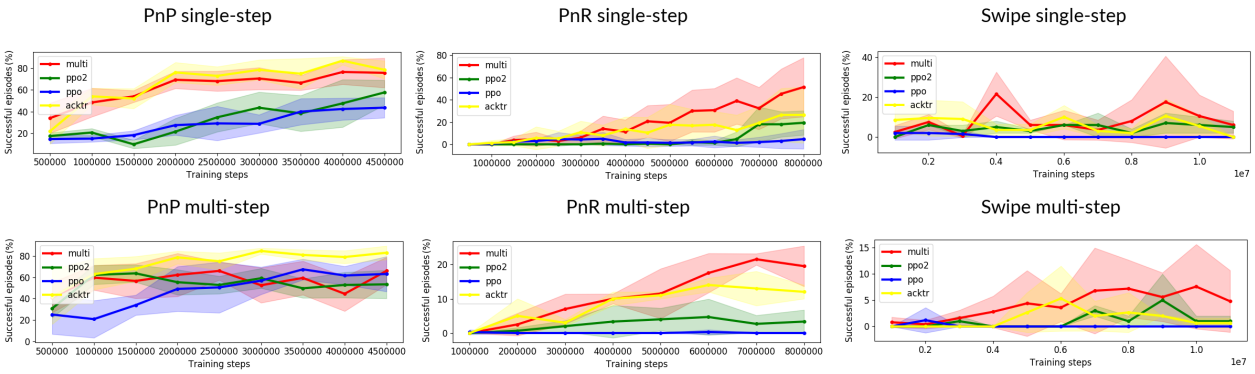
Obr. 1: Úloha Pick and Rotate je modifikací úlohy Pick and Place. Robot musí vybrat objekt a poté jej přesně umístit v prostoru podle požadavků cílového stavu. Na pravé straně jsou podrobné informace o průběhu tréninku.

První novou úlohou je Pick and Place s podúlohami (PnP), kde každá část procesu má samostatnou odměnu. Úloha Pick and Rotate (PnR) představuje rozšířenou verzi PnP (viz například obr. 1) . Tato úloha se týká specifického typu robotické manipulace, kdy je robot trénován k tomu, aby zvedl předmět a otočil jej do určité orientace. Úloha obecně zahrnuje následující kroky: První dva dílčí cíle jsou podobné jako u PnP, ale třetí dílčí cíl (otočit) vyžaduje, aby robot přesně otočil uchopený objekt do polohy 6DOF.

Třetí a nejobtížnější úlohou je Swipe. Oproti úlohám Pick a Rotate je zde prezentována nová dílčí úloha. Robot si musí osvojit nové dovednosti, aby v této úloze uspěl. Swipe vyžaduje, aby robot našel a přesunul houbu do požadované polohy a sledoval specifickou trajektorii po desce stolu.

2.3 Trénink a evaluace

Úlohy popsané v předchozí části bylo trénováno ve virtuálním robotickém simulátoru s robotickými ra-



Obr. 2: Srovnání jednokrokových a vícekrokových úloh s rostoucí složitostí zleva doprava v jednokrokovém (nahoře) a vícekrokovém (dole) scénáři

Tab. 1: Srovnání algoritmů pro úlohy s rostoucí složitostí

ALGORITHM	PNP SINGLE	PNP MULTI	PNR SINGLE	PNR MULTI	SWIPE SINGLE	SWIPE MULTI
PPO	43.6 (9.2)	63.1 (16.4)	4.6 (8.6)	0.8 (0.4)	0.0. (0.0)	0.0 (0.0)
PPO2	57.6 (11.3)	53.4 (13.2)	19.3 (10.8)	4.2 (2.9)	5.9 (3.0)	1.0 (1.0)
ACKTR	**78.4 (8.2)**	**83.0 (6.4)**	26.4 (25.0)	12.4 (3.1)	0.0 (0.0)	0.6 (0.9)
MULTIPPO2 (OURS)	75.6 (13.5)	66.0 (12.7)	**51.3 (26.4)**	**21.6 (5.2)**	**6.0 (6.9)**	**4.8 (5.8)**

meny Panda i Kuka (oba roboti dosáhli podobných výsledků). Poloha a orientace objektů při inicializaci byla náhodná. Jednokrokové úlohy jsme trénovali po 512 tréninkových krocích a vícekrokové úlohy po 1024 krocích.

Přesnost algoritmu opakovaně vyhodnocujeme po určitém počtu tréninkových kroků. Robotovi je předloženo 50 nových konfigurací a sečetli úspěšné epizody. Vyhodnocujeme také úspěšnost dílčích úkolů, protože nám může napovědět, který podúkol nebyl naučen správně. Počítá se také průměrný počet kroků na každý dílčí cíl a průměrná vzdálenost od cílového stavu.

3 Výsledky

Výsledek pro úlohu PnP prokázaly (**??**, že jak single-policy, tak multi-policy sítě jsou schopny se úlohu naučit. Algoritmy MultiPPO2 a ACKTR mírně překonávají algoritmy PPO a PPO2 a dosahují 80 % přesnosti. Vzhledem k tomu, že tato úloha nevyžaduje různorodé dovednosti, algoritmy se výrazně neliší. Podobné výsledky jsou i u vícekrokového scénáře (graf vlevo dole). Mezi single-policy PPO2 a multi-policy MultiPPO2 není žádný rozdíl. Celkový výkon pro vícekrokový scénář je překvapivě lepší než pro jednokrokový, přestože je úloha obtížnější kvůli náhodné poloze paže po dokončení první dílčí úlohy. Můžeme konstatovat, že všechny algoritmy vykazují v úlohách PnP téměř podobný výkon.

Ve složitějším scénáři, kde jsou prezentovány různé dovednosti (Pick and Rotate), můžeme vidět větší rozdíl mezi sítěmi s jednou a více policy. Jelikož je jako jeden z dílčích cílů prezentováno otáčení 6DOF, jednopolitické sítě si vedou mnohem hůře ve srovnání s MultiPPO2. Naše metoda dosahuje 51 % přesnosti ve srovnání s 26 % u PPO2 a 19 % u ACKTR. PPO v této úloze selhává, protože dosahuje přibližně 4 % přesnosti. Ve vícekrokovém scénáři metoda MultiPPO2 opět překonala algoritmy s jednou sítí. Jak PPO, tak PPO2 nebyly schopny se tuto úlohu naučit. Výkonnost MultiPPO2 klesá mezi jedno- a vícekrokovými úlohami z 51 na 21 %.

Analýza nejobtížnější úlohy ukázala, že algoritmus s více policy si vede lépe, ale výsledky nejsou uspokojivé. MultiPPO2 dosahuje 20 % přesnosti v polovině tréninku, ale poté výkonnost klesá. Podobný pokles můžeme pozorovat i u vícekrokového scénáře. To lze přičíst složité funkci odměny, kdy funkce odměny pravděpodobně narušuje proces trénování a je třeba jej dále zlepšovat. Jak je vidět na grafech vpravo, algoritmy s jednou policy nebyly schopny tuto úlohu vyřešit.

Výsledky pro všechny úlohy jsme shrnuli v tab. **??**. Můžeme konstatovat, že MultiPPO2 překonává všechny single-policy algoritmy v jedno- i vícekrokových úlohách, kde jsou vyžadovány různorodé dovednosti.

4 Závěr

Potvrdili jsme hypotézu, že různorodé dovednosti (různé dílčí cíle s různými funkcemi odměny)

způsobují problémy single-policy algoritmů. Ty nejsou schopny zachytit variabilitu stavového prostoru úlohy. Představili jsme metodu pro tréninkový proces RL, která je schopna se s tímto problémem vypořádat. Náš algoritmus MultiPPO2 překonal single-policy algoritmy reprezentované PPO, PPO2 a ACKTR ve složitých úlohách vyžadujících různorodé soubory dovedností. Pomocí robotické simulace jsme ukázali, že algoritmus MultiPPO2 je efektivní při řešení problémů, které lze rozdělit na menší části. Domníváme se, že strategii více policy lze aplikovat i na další algoritmy RL. Tuto možnost budeme zkoumat v budocnu. Další možné pokračování zahrnuje automatické rozdělení úlohy na dílčí úlohy. Rádi bychom také prozkoumali kompozici dílčích cílů v komplexních úlohách.

Poděkování

Tento příspěvěk vznikl za podpory projektu GAČR č. 23-04080L a č. 21-31000S).

Literatúra

Alet, F., Lozano-Pérez, T. a Kaelbling, L. P. (2018). Modular meta-learning. V *Conference on Robot Learning*, str. 856–868. PMLR.

Andreas, J., Klein, D. a Levine, S. (2017). Modular multitask reinforcement learning with policy sketches. V *International Conference on Machine Learning*, str. 166–175. PMLR.

Clegg, A., Yu, W., Tan, J., Liu, C. K. a Turk, G. (2018). Learning to dress: Synthesizing human dressing motion via deep reinforcement learning. *ACM Transactions on Graphics (TOG)*, 37(6):1–10.

Dalal, M., Pathak, D. a Salakhutdinov, R. R. (2021). Accelerating robotic reinforcement learning via parameterized action primitives. *Advances in Neural Information Processing Systems*, 34:21847–21859.

Devin, C., Gupta, A., Darrell, T., Abbeel, P. a Levine, S. (2017). Learning modular neural network policies for multi-task and multi-robot transfer. V *2017 IEEE international conference on robotics and automation (ICRA)*, str. 2169–2176. IEEE.

Hill, A., Raffin, A., Ernestus, M., Gleave, A., Kanervisto, A., Traore, R., Dhariwal, P., Hesse, C., Klimov, O., Nichol, A., Plappert, M., Radford, A., Schulman, J., Sidor, S. a Wu, Y. (2018). Stable Baselines. `https://github.com/hill-a/stable-baselines`.

Marzari, L., Pore, A., Dall'Alba, D., Aragon-Camarasa, G., Farinelli, A. a Fiorini, P. (2021). Towards hierarchical task decomposition using deep reinforcement learning for pick and place subtasks. V *2021 20th International Conference on Advanced Robotics (ICAR)*, str. 640–645. IEEE.

Plappert, M., Andrychowicz, M., Ray, A., McGrew, B., Baker, B., Powell, G., Schneider, J., Tobin, J., Chociej, M., Welinder, P. a spol. (2018). Multi-goal reinforcement learning: Challenging robotics environments and request for research. *arXiv preprint arXiv:1802.09464*.

Pore, A. a Aragon-Camarasa, G. (2020). On simple reactive neural networks for behaviour-based reinforcement learning. V *2020 IEEE International Conference on Robotics and Automation (ICRA)*, str. 7477–7483. IEEE.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A. a Klimov, O. (2017). Proximal policy optimization algorithms.

Vavrecka, M., Sokovnin, N., Mejdrechova, M. a Sejnova, G. (2021). Mygym: Modular toolkit for visuomotor robotic tasks. V *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, str. 279–283, Los Alamitos, CA, USA. IEEE Computer Society.

Wu, Y., Mansimov, E., Liao, S., Grosse, R. a Ba, J. (2017). Scalable trust-region method for deep reinforcement learning using kronecker-factored approximation.

Jak přemítat o umělé inteligenci*

Jiří Wiedermann

Centrum Karla Čapka pro výzkum hodnot ve vědě a technice
Ústav informatiky AV ČR
Pod Vodárenskou věží 2, 182 07 Praha, Česká republika
Email: jiri.wiedermann@cs.cas.cz

Motto: *„Předpokládám, že moudrost znamená správné používání znalostí. Samotné vědění ještě neznamená moudrost. Je mnoho lidí, jež vědí mnoho — ale to z nich dělá ještě větší hlupáky. Není nad hlupáka než vědoucího hlupáka. Poznání, jak používat znalosti, znamená mít moudrost.“*

C. H. Spurgeon, *The Fourfold Treasure* (1871)

Abstrakt

„Proč rozvíjíme umělou inteligenci?“, a *„Jaký účel má používání umělé inteligence?“* Příhodnou odpověď naznačují již současné velké jazykové modely: umělou inteligenci rozvíjíme za účelem získání a aplikování umělé moudrosti, jež umožní dělat moudrá rozhodnutí a chovat se moudře. Ukážeme, že současné jazykové modely nepřímo, vzhledem ke svým jazykovým schopnostem a návaznosti dat, ze kterých se učí, na popisy z reálného světa, dovedou extrahovat sémantiku ze syntaktických dat a navíc splňují podmínky tzv. 4E kognice (embodied, embedded, extended, enacted cognition). Tato se snaží vysvětlit mechanismy inteligentního chování pomocí dalších než výlučně výpočetních prostředků. Tyto modely tak vládnou jistou formou iluzorní inteligence a iluzorní moudrosti, doposud nepopsanou v odborné literatuře. Toto poznání má fundamentální význam pro filozofii a metodologii výzkumu umělé inteligence, protože představuje zásadní posun ve výpočetním paradigmatu používaném v umělé inteligenci, a sice od pohledu na umělou inteligenci jako na procesy generující znalosti směrem k procesům generujícím a využívajícím moudrost.

1 Úvod

Hledáme odpovědi na dvě zdánlivě jednoduché otázky: *„Proč rozvíjíme umělou inteligenci?“,* a *„Jaký účel má používání umělé inteligence?“* Samozřejmě, zajímají nás netriviální odpovědi, vycházející z hlubšího pochopení pojmu umělé inteligence, vyplývají z nějaké teorie, jež platí pro „jakoukoli umělou inteligenci“, přinesou nové vhledy do povahy umělé inteligence a dovolí extrapolaci trendů v oblasti umělé inteligence. Pod pojmem „jakákoliv umělá inteligence“ rozumíme jak to, co v současné době považujeme za umělou inteligenci, tak i veškeré druhy uměle vytvořené inteligence v budoucnosti, jak na Zemi, tak kdekoliv ve Vesmíru.

Nabízíme následující odpovědi: umělou inteligenci rozvíjíme za účelem získání a aplikování umělé moudrosti, jež umožní dělat moudrá rozhodnutí a chovat se moudře ve světě, ve kterém umělá inteligence operuje a má o něm dostatečné znalosti. V tomto kontextu chápeme umělou moudrost jako „správné používání znalostí“ pomocí účelného chování — kombinovaný efekt kognice a akce (viz úvodní citát). Nositeli umělé moudrosti jsou autonomní vtělení kognitivní behaviorální agenti. Ukážeme, že již v současné době se umělá moudrost vyskytuje v různých druzích a intenzitě a v rozličných světech, ve formě různých velkých jazykových modelech.

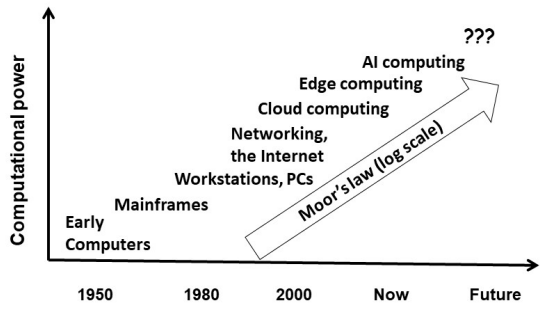
Pojmem moudrosti se zabývali především filozofové. Koncept moudrosti se vyskytuje v oblasti umělé inteligenci již od samotných počátků této discipliny. Nicméně, až donedávna bylo více pozornosti věnováno znalostem nežli moudrosti. Bylo tomu tak pravděpodobně proto, že znalosti jsou považovány za základní koncept, od nějž jsou odvozovány další koncepty, a speciálně moudrost. V posledních letech se situace mění a moudrost se dostává do popředí zájmu v oblasti praktické filozofie a etiky — a to je vlastně i náš případ.

V tomto kontextu práce přináší několik důležitých výsledků a poznatků.

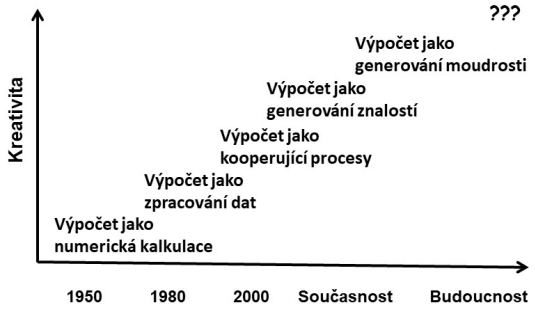
Po prvé, ukazuje, že systémy generující moudrost jsou přirozeným pokračováním trendů ve vývoji jak výpočetních technologií, tak i jejích aplikací na oblasti, vyžadující kognitivní schopnosti v celém spektru od úrovně získávání dat až po úroveň jejich zpracování a využívání.

Po druhé, zabývá se samotnou definici moudrosti, se kterou by mohly pracovat výpočetní technologie. Za tím účelem definuje model vtělených kognitivních behaviorálních agentů, jako systémů umělé inteligence, jež generují moudrost v širokém spektru různých světů. Tímto způsobem zobecňují pojem moudrosti i na domény odlišné od těch, ve kterých vy-

*Tato práce vznikla na základě diskusí a společných publikací s Janem van Leeuwenem z Utrechtské university.



Obr. 1. Vývoj počítáni



Obr. 2. Výpočetní trendy

niká lidská inteligence.

Po třetí, zkoumá současné velké jazykové modely z hlediska jejich využití pro generování moudrosti. Dochází k zajímavým výsledkům, že tyto systémy částečně splňují podmínky tzv. 4E kognice (embodied, embedded, extended, enacted cognition), která se snaží vysvětlit mechanismy inteligentního chování i dalšími než výlučně výpočetními prostředky. Jedná se o nepřímou, zprostředkovanou kognici, vyznačující se jakýmsi off-line povědomím o virtuálním světě, jež je zprostředkovaný prostřednictvím schopnosti modelu v jistém smyslu porozumět jazyku. Je to jakási zatím v odborné literatuře nepopsaná forma iluzorní inteligence, jež je podstatně odlišné od lidské inteligence a je charakteristická pro současné velké jazykové modely tím, že ukazuje možnost existence nějaké formy inteligence bez kognice.

Po čtvrté a v neposlední řadě, z filozofického hlediska a z pohledu metodologie, umělá moudrost představuje podstatný posun ve výpočetním paradigmatu používaném v umělé inteligenci, a sice od pohledu na výpočty jako procesy generující znalosti směrem k procesům generujícím a využívajícím moudrost. Koncept umělé moudrosti, vedle lidské moudrosti, se stává novou, a pravděpodobně finální metou lidského snažení.

Cílem práce není poskytnout návod, jak realizovat systémy generující moudrost, ale jak na ně pohlížet, jak je chápat a rozumět jejím možnostem a limitům.

2 Trendy ve využívání výpočetních technologií

Vývoj výpočetních technologií od jejich počátků v polovině 20. století až po současnost je obecně znám a přehledně jej zachycuje Obr. 1.

Z hlediska této práce je více zajímavý pohled na tyto technologie (Obr. 2), jež zachycuje jejich „informační sílu“ — pojem, jenž je zřejmější z popisu tohoto obrázku. Z obou obrázků je jasně vidět, jak s rozvojem informačních technologií a jejich aplikací rostla (a roste) jejich informační síla v rámci tzv. DIKW hierarchie: *data, informace, znalosti, moudrost* (Pyramid 2023). Tato hierarchie zachycuje skutečnost, ze v typickém případě je informace definovaná pomocí dat, znalosti pomocí informací, a moudrost pomocí znalostí.

V další kapitole ukážeme, jak lze na výše zmíněnou hierarchii pohlížet z hlediska teorie epistemických výpočtů, tj. z pohledu na výpočty jakožto procesů generujících znalosti.

3 Výpočet jako proces generující znalost

Základem našeho přístupu k výpočetním schopnostem systémů umělé inteligence je *epistemická teorie výpočtů* jež vychází z prací autorů van Leeuwena a Wiedermanna (2014, 2015a, 2015b, 2017). Z hlediska této teorie výpočtů nahlížíme na výpočty jako na procesy, které generují znalosti nad danou znalostní doménou $\mathbb{D}$ v rámci příslušné znalostní (epistemické) teorie $\mathcal{T}$. Výpočet pracuje tak, že kombinuje prvky znalostní domény — jimiž jsou *informace* (resp. jejich reprezentace), nazývané také *elementární znalosti* — do odvozených, často složitějších konstrukcí, které již tvoří *novou znalost*, opět nad danou doménou a v rámci teorie $\mathcal{T}$. Pro kombinaci těchto prvků používá výpočet množinu (odvozovacích) *pravidel,* která může být předem daná v rámci teorie $\mathcal{T}$, anebo se může tvořit pomocí učení během velkého počtu různých výpočtů nad danou doménou.

Systém tímto způsobem pracuje s více či méně formální teorii $\mathcal{T}$, která zachycuje vlastnosti dané znalostní domény a způsoby odvozování nových znalostí,

stále v rámci dané domény. Jakmile systém načte nějaká data, tyto se stávají v rámci teorie $\mathcal{T}$ informací. Z nich systém shora popsaným způsobem generuje znalosti; některé z nich se mohou stát výstupem výpočtu.

Díky své obecnosti znalostní přístup lze uplatnit nejen v dobře formalizovatelných, tzv. *exaktních znalostních doménách,* ale i ve znalostních doménách a pro odvozovací pravidla, které se vzpírají jakékoliv formalizaci. Takovým doménám budeme říkat *popisné znalostní domény.*

Příkladem formální znalostní domény budiž množina přirozených čísel s příslušnou teorii reprezentovanou pomocí Peanových axiomů.

Typickým případem popisné domény s neformálními odvozovacími pravidly je reálný svět. Jeho objekty, jevy, akce a vztahy mezi jimi jsou popsány pomocí přirozeného jazyka. Znalosti o takové doméně jsou zachyceny ve větách přirozeného jazyka. Odvozovací pravidla jsou v tomto případě tzv. *pravidla racionálního uvažování a chování.* Tato pravidla vycházejí z faktů a argumentů, která lze odpozorovat z přirozeného jazyka a zachytit v přirozeném jazyce. V typickém případě mají popisné domény rozsáhlé znalostní báze (jako např. obsah internetu) a relativně krátké odvozovací řetězce. Současné velké jazykové modely (LLM) jsou pěkným příkladem takových domén a neformálních teorií.

Pro formalizaci výše zmíněného přístupu viz práci van Leeuwena a Wiedermanna (2017). Přehled dosavadních výsledků z oblasti kognitivních výpočtů lze nalézt v práci Wiedermanna a van Leeuwena (2015a). Současně to umožní nový pohled na velké jazykové systémy, jak uvidíme v části 5.

V rámci epistemického přístupu k výpočtům lze tedy dobře popsat první tři úrovně DKIW hierarchie: data, informace, znalosti. Poslední, nejvyšší úroveň moudrosti popíšeme v další kapitole.

4 Od znalosti k moudrosti

Pojem moudrost, jako všechna slova přirozeného jazyka, která vznikly v běžném životě, je velmi kluzký, těžce definovatelný. Problém je v tom, že se jedná o slovo — kufr, jak to nazýval Marvin Minsky (1998). Jsou to slova, kterým lidé připisují, resp. do nichž „balí“, leckteré další významy. Např. Wikipedie definuje moudrost jako *vědění, mudrlanství, chytrost, slovo označující schopnost používat znalosti, pochopení, selský rozum a vhled* (Wisdom 2023). Co slovo, to další kufr. To není definice, ze které by se dalo vycházet ve výpočetním prostředí.

Naštěstí máme zde i definice z oblasti filozofie. Jedna z nich, použitelná v našem případu, je v úvodu této práce. Pro naše účely se hodí obecně akceptovaná „slovníková“ definice: *zatímco znalost je definována jako nabytí dat a informací, moudrost je praktická aplikace a použití znalostí za účelem vytváření hodnot.* Tato definice klade důraz na praktické aspekty moudrosti — její využití v reálném světě.

Pro naše další účely si tuto definici ještě upravíme. Všechny známe definice moudrosti se vztahují k lidské moudrosti, a vznikly, pochopitelně, na základě kognitivního (smyslového) poznání světa. My budeme potřebovat definici umělé moudrosti vhodnou pro použití v v obecných, a tím pádem také v umělých, kognitivních systémech. V obecném případě kognitivní systém je autonomní systém schopný vnímat své okolí, učit se ze své zkušenosti, předvídat běh událostí, jednat účelně a eticky pro splnění svých cílů a přizpůsobovat se měnícím okolnostem (Vernon 2021).

V tomto kontextu budeme používat následující definici: *přirozená i umělá moudrost je správné používání znalostí pomocí účelného chování — kombinovaný efekt kognice a jednání směřující k vytváření pragmatických či dodržování etických hodnot.*

Nositeli umělé moudrosti jsou především autonomní interaktivní vtělení kognitivní behaviorální agenti. Umělá moudrost se vyskytuje v různých světech, druzích a intenzitě závislé na ustrojení a schopnostech agenta získávat a využívat potřebné znalosti ve svém osvětí. V naší terminologii epistemických výpočtů různé světy (osvětí) reprezentuje znalostní doména $\mathbb{D}$, druh umělé moudrosti odvozujeme od epistemické teorie $\mathcal{T}$ — jaká je její vyjadřovací síla, co všechno, jaká fakta, vztahy a dotazy dovede vyjádřit, intenzita hovoří o efektivitě takového vyjádření, ustrojení se týká vybavení agenta pomocí senzorů a efektorů a repertoáru jejich akcí. Problémem ovšem zůstává část definice, požadující *„správné používání znalostí pomocí účelného chování“*, a také, co se chápe pod pojmem *„směřovat k vytvoření pragmatických či etických hodnot“*? Co to znamená v kontextu autonomních interaktivních vtělených kognitivních behaviorálních agentů?

To je složitá otázka, která sa týká množství a kvality znalostí agenta a jeho schopnosti je využívat.

Intuitivně, být moudrý v nějaké znalostní doméně znamená, že jednak agent má pomocí svých senzorů přístup ke všem objektům v této doméně, a také, že se dokáže vypořádat se všemi situacemi, dotazy a příkazy, týkající se těchto objektů, avšak pouze vzhledem ke svému poslání. Vše k tomu potřebné je popsáno v epistemické teorii. Poslední výhrada je důležitá — např. nemůžeme chtít po autonomním vozidle, aby s námi konverzovalo o všech objektech, které zachycuje svými senzory, avšak nejsou popsány v epistemické teorii, anebo aby se chovalo účelně v prostředí, pro které nebylo navrženo resp. naučeno — třeba v prostředí středověkého města.

Pro zodpovězení této výhrady musíme zavést pojem smysluplnosti (resp. poslání) agenta. Tento pojem je formálně popsán v jeho *funkční specifikaci* Φ. Tato specifikace popisuje, jaké funkce musí agent vykonávat a za jakých podmínek. Specifikace musí

splňovat následující dvě podmínky:

- Specifikace musí předepsat, jak se má systém chovat, v závislosti na teorii $\mathcal{T}$ v dané situaci s_i za předpokladu, že víme, jak se choval v předchozích situacích $s_1, s_2, \ldots, s_{i-1}$, pro libovolné $i > 1$ a libovolnou posloupnost situací, která se může v doméně $\mathbb{D}$ vyskytnout.
- Specifikace musí garantovat vytváření pragmatických hodnot a současně i dodržování etických hodnot.

Uvedená definice funkční specifikace autonomního interaktivní vtěleného kognitivního behaviorálního agenta je typická pro epistemický přístup, protože vyžaduje splnění dvou podmínek, aniž by něco hovořila o tom, jak toho dosáhnout, např. jestli musejí agenti být v nějakém smyslu inteligentní anebo ne. To ji dává širokou platnost — odpovídající systémy mohou být fixní, neměnné během své činnosti, a nebo se mohou učit, rozvíjet své znalosti (tzv. evoluční systémy), mohou být vědomé, mít svobodnou vůli a další mentální schopnosti. Pro specifikaci není důležité, jak má agent dosahovat svých cílů, ale je důležité to, co má dělat (vytvářet pragmatické hodnoty), a také, aby přitom dodržoval etické hodnoty či principy. V obecném případě jsou etické hodnoty popsány pomocí etické teorie $\mathcal{E}$. Je to opět specializovaný druh epistemické teorie, která popisuje etické principy — zásady a limity chování — které musí autonomní interaktivní vtělený kognitivní behaviorální agent dodržovat.

Teď konečně můžeme definovat umělou moudrost: *agent A se chová moudře, anebo stručně, je moudrý ve své doméně $\mathbb{D}$ vzhledem ke své funkční specifikaci ϕ, epistemické teorii $\mathcal{T}$ a etické teorii $\mathcal{E}$ právě když v každé situaci, ve které se může ocitnout, splňuje své funkční specifikace.*

Takto obecně pojatá umělá moudrost se tedy formálně skrývá v agentově funkční specifikaci Φ a je závislá na doméně $\mathbb{D}$ a teorii $\mathcal{T}$ a $\mathcal{E}$. Pokud jsou teorie $\mathcal{T}$ a $\mathcal{E}$ příliš jednoduché, nepostihující racionální chování agenta ve všech situacích a nezaručující vytvoření pragmatických či etických hodnot, může se agentovo chování jevit jako hloupé nebo neetické — ale to je v podstatě chyba návrhu. Agent totiž dělá přesně to, co mu jeho specifikace diktuje.

Tato definice dává smysl i z hlediska definice moudrosti tak, jak ji definovali starověcí filozofové a náboženští myslitelé. Tito zdůrazňovali maximální inteligenci, převyšující úroveň inteligence většiny lidí jako definiční vlastnost moudrosti s tím, že se jedná o zřídkavou kvalitu. Pokud odhlédneme od toho, že inteligence je opět slovo — kufr, naše definice vlastně ztotožňuje zřídkavou kvalitu inteligence se schopnosti naplňovat poslání agenta, definované v jeho funkčních specifikacích, ve všech situacích, se kterými se může setkat (a tedy, implicitně, dosahovat svých cílů etickým

Obr. 3. Théâtre D'opéra Spatial

způsobem). Co více chtít od agenta, jenž nemůže překročit svoji specifikaci? Všimněme si, že „zřídkavou kvalitu“ nejen umělé inteligenci dodává nikoliv pouze samotná schopnost dosahovat svých cílů, ale a způsob, jakým je jich dosahováno — etika.

Formálně definovaná umělá moudrost umožňuje hovořit o moudrosti extrémně jednoduchých kognitivních systémů, jakým jsou např. automatické otvírané dveře. Jsou moudré, protože otevřou dveře (vykonáním akce vytvoří pragmatickou hodnotu pro procházející osobu), kdykoliv rozeznají takovou potřebu (kognitivní schopnost), a chovají se eticky (pokud jsou zkonstruovány tak, že nikoho „nepřivřou“), a nic jiné se od nich nepožaduje. Složitější systém, jako třeba autonomní vozidlo, je také moudrý, protože (a pokud) pomocí kombinovaného efektu využití svých senzorů a motorů vytvoří pragmatickou a etickou hodnotu — doveze svého uživatele bezpečně k cíli.

Výhodou shora uvedené definice je skutečnost, že ukazuje, že moudrost není absolutní vlastnost, ale že závisí na schopnostech agenta, formálně popsaných v jeho „dvousložkové“ specifikaci, která zachycuje jak jeho jednání v každé situaci, tak i nutnost jednat tak, aby agent směřoval k naplnění svého poslání za dodržování etických hodnot. Nevýhodou je, že ji lze aplikovat pouze na formálně definované umělé systémy. Nám však nejde o to, poskytnout návod, jak takové systémy realizovat, ale jak je chápat a rozumět jejím možnostem a limitům.

5 Velké jazykové modely: Inteligence bez kognice

Jistou představu o tom, jak by mohly systémy generující moudrost v budoucnosti vypadat, nám dávají současné velké jazykové modely. Ukážeme, že tyto modely mají potenciál *správně používat své znalostí pomocí účelného chování — kombinovaný efekt kognice a jednání směřující k vytváření pragmatických a etických hodnot.*

Toto tvrdíme i přesto, že dle významného italo-britského filozofa Luciana Floridiho (2023) tyto systémy postrádají jakoukoliv inteligenci a porozumění, a nemají vůbec žádné kognitivní schopnosti. Tím pádem jsou velmi křehké (náchylné ke katastrofickým selháním), nespolehlivé (schopné dodat nesprávnou anebo vymyšlenou informaci), příležitostně schopné dělat elementární logické chyby v uvažování anebo v jednoduchých počtech. Floridi argumentuje, že v nich dochází k oddělení vazby mezi jednáním a inteligencí. Jde tedy o jednání bez inteligence, jak Floridi tyto systémy charakterizuje.

Mimochodem, je zajímavé si uvědomit, že v některých případech se mohou nedostatky velkých jazykových modelů, zmiňované Floridim, obracet ve výhody. To je případ systémů, generujících z textového popisu obrázky (jako je např. DALL-E (2023)). V případě obrázků totiž nelze jednoznačně určit, jestli je výsledný obrázek chybný, nebo padělek, anebo pouze měl systém (rozuměj: tvůrce obrázku) jinou představu než zadavatel. Obr. 3 (Théâtre d'Opéra Spatial 2023) ukazuje, že obrázek zadaný pouze textovým popisem může být oceněn jako kvalitní umělecké dílo. Shora uvedené nedostatky se v tomto kontextu mohou jevit jako jistý druh kreativity, vedoucí k nečekaným efektům, které nelze považovat za chyby, pouze za jakýsi „specifický umělecký vkus".

Vraťme se však zpět k velkým jazykovým systémům generujícím texty. Proč tedy tvrdíme, že tyto systémy, přes všechny shora uvedené nedostatky, mají potenciál vykazovat moudrost, alespoň v nějaké minimální míře? Uvedeme dva argumenty ve prospěch našeho tvrzení.

Prvním argumentem je samotný způsob práce těchto modelů,v jehož základem je *princip extrakce sémantiky ze syntaktických dat*. Tuto schopnost modelů Floridi (2023) sice zmínil, ale při svých úvahách nevzal v potaz povahu syntaktických dat, se kterými velké jazykové modely pracují. Tyto systémy jsou totiž schopné v překvapivé míře rozumět přirozenému jazyku, a pracovat s ním pomocí statistické analýzy různých vzorů, jež se nacházejí v kvantech syntaktických dat, na kterých je systém trénován. Porovnejme to s lidmi, kteří pracují s jazykem i jiným způsobem — sémantickým a kontextuálním usuzováním. Výsledek je však většinou tentýž — chápání významů zprostředkovanými jazykem (Agüera y Arcas, 2022). Již tento fakt samotný ukazuje, že tyto systémy nejsou zcela bez inteligence.

Náš druhý argument vychází z paradigmatu málo užívaného v oblasti umělé inteligence, kybernetiky anebo robotiky, ale o to známějšího v kognitivních vědách: *4E kognice* (viz např. (Newen, de Bruin & Gallagher, 2018) nebo (Extended mind thesis, 2022)). Toto paradigma postuluje, že kognice není pouhým vnitřním, individuálním procesem, nýbrž se jedná o emergentní proces, jež vzniká interakcí mezi mozkem, tělem, okolím a sociálním kontextem. Zkratka 4E naznačuje, že kognice je vtělená, vnořená, vykonávaná, a rozšířená (embodied, embedded, enacted, or extended) pomocí mimo-mozkových procesů a struktur.

Podívejme se, jestli a jak velké jazykové modely souzní s paradigmatem 4E.

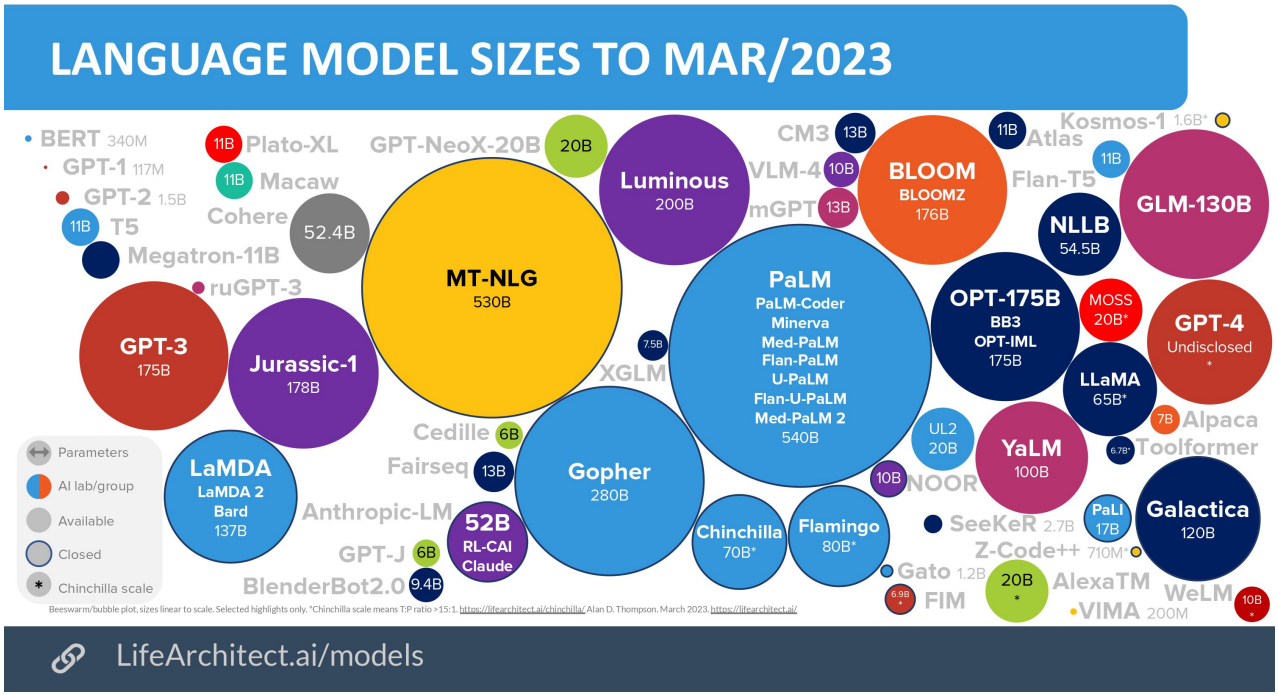
Vtělenost. Jazykové modely zřejmě nejsou fyzicky vtělené do reálného světa. Na druhé straně, učí se z obrovského množství dat a jejich porozumění jazyku vychází z kontextu, ve kterém se slova a fráze vyskytují. Tyto pocházejí z reálného světa, resp. přesněji, z představ lidí, kteří o tomto světe vytvářejí své texty. Takže lze říci, že prostřednictvím svých znalostí jsou tyto modely *nepřímo vtěleny* do reálného světa, protože poznávají svět pomocí interakce s texty pocházejícími s různých kontextů vyskytujících se ve skutečném světě.

Vnořenost. Za stejných důvodů jsou tyto modely *nepřímo vnořené,* zakotvené v reálném světě. Učí se asociovat slova s věty s kontextem, ve kterém se vyskytují a to jim dává význam v reálném světě.

Vykonatelnost. Způsob, kterým jazykové modely zpracovávají jazyk, lze chápat jako vykonatelný proces, protože tyto aktivně generují a manipulují své jazykové výstupy v závislosti na svých interních mechanismech. Tyto výstupy dále *nepřímo ovlivňují konání* lidí, kteří zadali vstupní data (prompt) do modelu.

Rozšířitelnost. Způsob, kterým velké jazykové modely reagují na svá zadání, zřejmě podstatně závisí na okolí systému, ze kterého „čerpá" své znalosti, a na promptu samotném, jež situuje systém v oblasti diskuse a poskytuje kontext pro porozumění jak vstupu, tak i výstupu. Tím je činnost sytému *nepřímo rozšiřovaná* a nasměrovaná, v závislosti na dotazu, do souvisejícího kontextu reálného světa.

Zastánci 4E kognice argumentují, že shora uvedené čtyři atributy mají vztah k inteligenci systémů, které splnění příslušných atributů vykazují. Je důležité si všimnout, že ve všech čtyřech případech atributů 4E kognice jsme hovořili o nepřímé vtělenosti, nepřímé vnořenosti, nepřímé vykonatelnosti, a nepřímé rozšířitelnosti. To jsou zcela odlišné atributy než ty, uvažované v klasickém případě 4E kognice. Nicméně, v tomto našem případě můžeme usuzovat z jednání systému přinejmenším o jakési *iluzorní inteligenci*, která nabízí iluzi inteligence vzhledem k danému prostředí, založenou na masivní agregaci dat z tohoto prostředí a výběru reakcí, jež závisejí na současném i minulém kontextu jednání systému. To je důvod, proč zcela nesouhlasíme s Floridiho závěrem, že ve velkých jazykových modelech je jednání odtržené od inteligence. Když, tak jde o jednání odtržené of reálné, bezprostřední kognice. (Mimochodem — tak jednají i lidé, pokud dospějí k rozhodnutí na základě „přemýšlení"). Iluzorní inteligence je v mnoha případech lepší než žádná inteligence. Otázkou zůstává, jestli součástí takové iluzorní inteligence je i nějaká forma vědomí. Pro úvahy, jestli velké jazykové modely mohou mít vědomí, viz práci (Chalmers 2022).



Obr. 4. Velikost jazykových modelů. Zdroj: https://s10251.pcdn.co/pdf/2023-Alan-D-Thompson-AI-Bubbles-Rev-7b.pdf

Znalosti využívané daným jazykovým modelem jsou ty znalosti o světě, které se model naučil analýzou textů získaných z Internetu. Všimněme si, že v takovém případě jazykové modely nevnímají svět „tak, jak vypadá“, tj. tak, jako ho vnímáme my lidé, ale pouze (neboli přesně) tak, jak se o něm, včetně o AI, píše. Text, který model vygeneruje, je požadovaná „umělá moudrost“, sémantika textu, který model vygeneruje, představuje pro nás ony *„pragmatické resp. etické hodnoty“*, a posléze samotný výpočetní akt konstrukce odpovědi odpovídá (resp. měl by odpovídat) „správnému používání znalostí modelu“. „Účelné chování“ je řízeno dotazem — model generuje text, který nejlépe souzní s odpovědí na daný dotaz. Reakce modelu současně nastavuje zrcadlo tomu člověku, který se ptá. Pokud se totiž nezeptá důkladně a nevysvětlí, co všechno chce vědět, včetně argumentů pro a proti, a nedožaduje se různých jiných, třeba i protichůdných názorů na danou věc, tak se dozví velmi málo, a povrchně, anebo dokonce dostane špatnou odpověď.

V reakci na iluzorní inteligenci velkých jazykových modelů můžeme tedy hovořit o *iluzorní moudrosti* takových systémů. Na rozdíl od kouzelníků, iluzionistů a jejich triků je zde situace příznivější. Pokud se nám něco na vygenerované moudrosti nezdá, anebo jen pro jistotu, můžeme požádat systém o vysvětlení, dodání třeba jiných, alternativních argumentů, a tak odhalit, jestli se jedná o iluzorní „moudrost“, anebo o moudrost, která je dobře zdůvodněná. To je současně návod, jak se stavět k používání současných velkých jazykových modelů.

V současné době se vyvíjí desítky různých jazykových modelů (viz. Obr. 4). Z hlediska technologie se tyto modely často velmi liší. Mohou používat různé architektury, trénovací postupy anebo techniky předzpracování dat. Z uživatelské hlediska jsou rozdíly mezi nimi jemnější. Vhodnost modelů se odvíjí od specifických úkolů, druhu a kvality trénovacích dat. Tato rozmanitost naznačuje, že velké jazykové modely a jejich architektury jsou robustní ve smyslu základní ideje, a naše výsledky k tomu dodávají, že tyto modely skutečně vládnou jakousi minimální, i když někdy iluzorní, inteligenci.

6 Proč rozvíjíme umělou inteligenci? Jaký účel má používání umělé inteligence?

Na základě všeho dříve řečeného začíná být zřetelná odpověď na dvě shora uvedené otázky. Umělou inteligenci rozvíjíme proto, abychom vyvinuli a poté mohli využívat nástroj na získávání (umělé) moudrosti — a to je současně i smysl jejího používání. Z definice DKIW hierarchie vyplývá, že moudrost, jako nejvyšší stupeň této hierarchie, nelze překonat. Někteří autoři se dokonce domnívají, že moudrost je více než inteligence (Jeste et al., 2020), ale to, zdá se, je otázkou definice inteligence — jestli je moudrost součástí inteligence

anebo ne.

Umělá moudrost se zdá být finální metou umělé inteligence. Jako celek to však není meta dosažitelná v konečném čase. Příklad matematiky, ve které prokazatelně existuje nekonečně, avšak spočetně mnoho teorií, to dokazuje. Velké jazykové systémy jsou prvními systémy AI, které naznačuji, ze v blízké budoucnosti budeme mit systémy, které budou generovat umělou moudrost a nebudou pouze pasivním „skladem" vědomostí (viz např. (Wiedermann & van Leeuwen, 2015b)). A to znamená zásadní posun společenského paradigmatu — od znalostní společnosti směrem k moudré společnosti. Umělá moudrost bude představovat společně s přirozenou moudrostí trvalý a smysluplný odkaz naším současníkům i potomkům a bude zvyšovat pravděpodobnost našeho a jejich přežití v zatím neznámých budoucích časech a místech.

7 Závěr

Umělá moudrost přestává být zajímavým slovním spojením, hodné filozofických úvah, ale začíná se jevit jako něco, co je na dosah dnešních technologií. Sam Altman, výzkumný ředitel firmy OpenAI, jež sestrojila a provozuje velký jazykový model GPT3 a GPT4, hovoří o tom, že finálním cílem OpenAI je všeobecná bezpečná umělá inteligence — AGI — s důrazem na slovo „bezpečná" (The Contradictions of Sam Altman, AI Crusader 2023). V kontextu moudrosti bychom slovo „bezpečná" nahradili slovem „etická". Proto nevidí důvod pro pozastavení vývojových prací na tomto projektu, jak požaduje Elon Musk a další vědci (Elon Musk, Other AI Experts Call for Pause in Technology's Development 2023). To je v souladu i s hlavní myšlenkou našeho příspěvku — směřovat k nacházení moudrosti pro naše současníky i potomky.

Poděkování: Tento příspěvek vznikl za částečné podpory TAČR v rámci projektu EBAVEL, registrační číslo CK04000150, a programu Strategie AV21 „Filozofie a umělá inteligence". Díky patří i systému ChatGPT za konzultace o práci velkých jazykových systémů.

Literatura

Agüera y Arcas, B. (2022). Do Large Language Models Understand Us?. Daedalus 2022; 151 (2): 183–197. doi: https://doi.org/10.1162/daed_a_01909

Chalmers, D. J. (2022). Could a Large Language Model be Conscious? https://philpapers.org/archive/CHACAL-3.pdf

DALL-E. (2023, March 29). In Wikipedia. https://en.wikipedia.org/wiki/DALL-E

DIKW pyramid. (2023, March 10). In Wikipedia. https://en.wikipedia.org/wiki/DIK_pyramid

Elon M. (2023). Other AI Experts Call for Pause in Technology's Development. (2023, March 29). The Wall Street Journal, https://www.wsj.com/articles/elon-musk-other-ai-bigwigs-call-for-pause-in-technologys-development-56327f?reflink=desktopwebshare_permalink

Extended mind thesis. (2022, December 1). In Wikipedia. https://en.wikipedia.org/wiki/ Extended_mind_thesis

Floridi, L. (2023). AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models (February 14, 2023). Philosophy and Technology, 2023, Available at SSRN: https://ssrn.com/abstract=4358789

Jeste, D., Graham, S., Nguyen, T., Depp, C., Lee, E., & Kim, H. (2020). Beyond artificial intelligence: Exploring artificial wisdom. International Psychogeriatrics, 32(8), 993-1001. doi:10.1017/S1041610220000927

Large language model. (2023, March 29). In Wikipedia. https://en.wikipedia.org/wiki/Large_language_model

Minsky, M. (1998). Consciousness is a Big Suitcase: A talk with Marvin Minsky. https://www.edge.org/conversation/marvin_minsky-consciousness-is-a-big-suitcase

The Contradictions of Sam Altman, AI Crusader. (2023, March 31). The Wall Street Journal, htpps://www.wsj.com/articles/chatgpt-sam-altman-artificial-intelligence-openai-b0e1c8c9?reflink=desktopwebshare_permalink

Newen, A., De Bruin, L. and Gallagher, S. (eds). (2018). *The Oxford Handbook of 4E Cognition,* Oxford Library of Psychology (2018; online edn, Oxford Academic, 9 Oct. 2018)

Peano axioms. (2023, March 14). In Wikipedia. https://en.wikipedia.org/wiki/Peano_axioms

Théâtre d'Opéra Spatial. (2023, April 3). In Wikipedia. https://en.wikipedia.org/wiki/Théâtre_d'Opéra_Spatial

Vernon, D. (2021). Cognitive System. In: Ikeuchi, K. (eds) Computer Vision. Springer, Cham. https://doi.org/10.1007/978-3-030-63416-2_82

Wiedermann, J. (2017). Nový pohled na výpočty a umělá inteligence. In: Sborník přednášek konference *Kognice a umělý život 2017,* http://cogsci.fmph.uniba.sk/kuz2017/files/zbornik/Wiedermann.pdf

Wiedermann, J. a van Leeuwen, J. (2014). Computation as knowledge generation, with application to the observer-relativity problem. In: *Proc. 7th AISB Symposium on Computing and Philosophy: Is Computation Observer-Relative?*, AISB Convention 2014 (Goldsmiths, University of London), AISB, 2014

Wiedermann, J. a van Leeuwen, J. (2015a). What is Computation: An Epistemic Approach. (Invited talk). In: *Italiano, G. et al., (eds.). SOFSEM 2015: Theory and Practice of Computer Science.* LNCS 8939, Berlin: Springer, pp. 1-13

Wiedermann, J. a van Leeuwen, J. (2015b). Towards a Computational Theory of Epistemic Creativity. In: *Proc. 41st Annual Convention of AISB 2015.* London, pp. 235-242

Wiedermann, J. a van Leeuwen, J. (2017). Understanding and Controlling Artificial general Intelligenct Systems. In: *Proc. 10th AISB Symposium on Computing and Philosophy: Language, Cognition and Philosophy*, AISB Convention 2017, (University of Bath, UK), AISB

Wiedermann, J., & van Leeuwen, J. (2018). Epistemic computation and artificial intelligence. In Philosophy and Theory of Artificial Intelligence 2017 (pp. 215-224). Springer International Publishing.

Wisdom. (2023, March 16). In Wikipedia. https://en.wikipedia.org/wiki/Wisdom

Register autorov