

Jihočeská univerzita v Českých Budějovicích
Pedagogická fakulta
Katedra pedagogiky a psychologie

Dodatek k disertační práci

Neurální koreláty aritmetických funkcí

Vypracovala: Mgr. Michala Plassová

České Budějovice 2019

V následujícím textu budou formou errata adresovány hlavní body negativního posudku mé disertační práce, přičemž cílem je blíže objasnit zvolený metodologický postup a poskytnout doplňující informace o zpracování získaných dat. Zejména budu pozornost věnovat palčivému problému neurověd a neurozobrazování, a to konkrétně nízké hladině replikovatelnosti a falešně pozitivním výsledkům. Poté se zaměřím na popis nastavení obtížnosti výzkumných úloh, které jsou skutečně v disertační práci chybně uvedeny, rovnou však předesílám, že chyba v popisu obtížnosti je pouze deskriptivní a na závěrečné výsledky nemá žádný vliv.

Dodatek ke kapitole 7.1

Na straně č. 76 je chybně uvedena obtížnost výzkumných úloh. Správná obtížnost je uvedena v tabulce č. 1.

Dodatek ke kapitole 9.2

Logaritmický rozdíl zobrazený v prvním řádku tabulky č. 1 odpovídá počtu prvků, kterými je větší množina objektů početnější, nežli druhá.

Tabulka 1 Vyjádření obtížnosti úloh.

Obtížnost	2.3	2.2	2.1	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1
Rozdíl	30	28	26	24	22	20	18	16	14	12	10	8	6

Dodatek ke kapitole 10.

Přejdeme-li nyní k metodologickým problémům neurovědy a neurozobrazování obecně, je vhodné na úvod zmínit, že v současné době probíhá velká diskuze o minimálních statistických standardech, a to i pro publikaci v impaktovaných časopisech (např. Woo et al., 2014; Rosier et al., 2016; Yeung, 2018). Yeung (tamtéž) ve své studii vzal všechny články z databáze PubMed v oblasti neurozobrazování pomocí funkční magnetické rezonance v roce 2017 a věnoval se jejich statistickému zpracování. Ačkoli průběžně narůstá počet studií, které používají statistické korekce, stále 53% z nich nepoužívá adekvátní statistický postup. Tyto studie v podstatě nereflktují naplnění minimálních standardů pro zvolené statistické

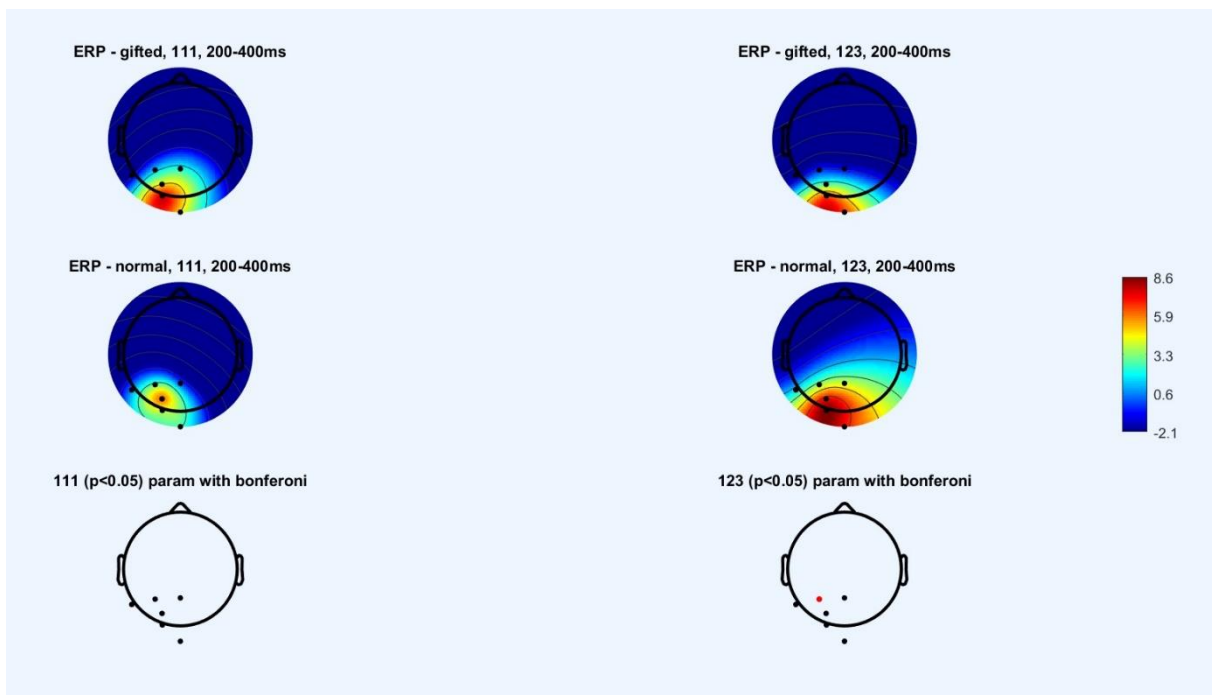
zpracování. Srovnatelné problémy se vyskytují i v oblasti zpracování elektroencefalografického (EEG) signálu. Kupříkladu Xu (et al., 2016) uvádí, že ve všech dostupných lékařských studiích věnujících se tématu epilepsie do roku 2015 (v databázích MEDLINE, EMBASE, PsychINFO, CINAHL a AMED) je pravděpodobně 71% pacientů diagnostikováno chybně. Buttonová (et al., 2013) jde dokonce ještě dál a na základě Ioannidisovy studie (2005) prohlašuje, že většina biomedicínského výzkumu je pravděpodobně falešně pozitivní. Příčinu autorka spatřuje v určitém tlaku na publikování, kdy tyto studie na malých vzorcích mají obecně nízkou velikost efektu, a dále pak opět zmiňuje minimální požadavky na statistické zpracování, kdy je hladina významnosti $p < 0.05$ považována za dostatečnou. Autorka poté dodává, že 97% studií v dané oblasti generuje minimálně jeden falešně pozitivní výsledek a replikace výzkumu se podaří v méně než 25% případů.

Během psaní své disertační práce jsem měla tyto problémy na zřeteli. Má hlavní snaha byla naplnit především minimální požadavky pro zvolenou analýzu EEG dat, tj. metodu evokovaných potenciálů. Nicméně naplnit veškeré požadavky na zvolenou statistickou metodu bylo díky problémům, jež uvádím níže, velmi náročné a zvolená prezentace výsledků je pak jistým kompromisem. Vystává tak trochu kacířská otázka, zdali má být hodnocení disertační práce v rámci zvolené statistiky přísnější, než jsou nároky na publikování ve většině neurovědních časopisů.

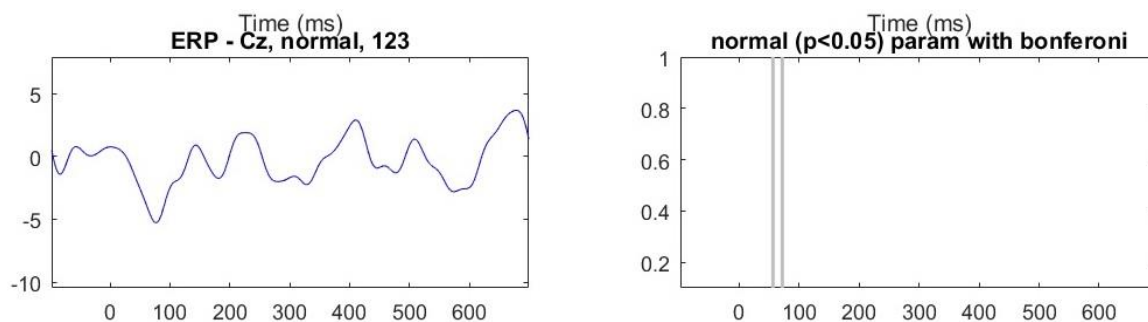
Přestoupím-li k problémům s naplněním požadavků na zvolenou statistickou metodu, pak jsem se zejména potýkala s nízkým množstvím naměřeného signálu u předškolních dětí (zhruba 30 minut na dítě) a nízkým počtem probandů. Mám-li reflektovat nízký počet probandů, kdy návratnost oslovení rodičů dětí prostřednictvím rozdávaní letáčků v mateřských školách byla nižší než 1 %, pak je pravděpodobně na vině má snaha o měření výhradně v dopoledních hodinách (pro minimalizaci alfa vlny a optimální kognitivní výkon v inteligenčním testu), kdy se rodiče museli uvolňovat z práce, a dále pak fakt, že rodičům bylo z měření poskytováno pouze malé množství informací, a to pouze v případě měřeného inteligenčního testu. Zde bych ráda adresovala možná nejasnou formulaci ohledně poskytování výsledků inteligenčního testu rodičům dětí. Rodiče byli vždy poučeni ohledně problematiky zastaralého testu bez české standardizace, jak jsem ve své disertační práci uvedla. Zdůrazním tedy, že výsledky byly vždy prezentovány jako pouhý odhad a z mé strany byla prioritou především nepoškodit zavádějícím výsledkem dítě. Reálně pak inteligenční pásmo v podstatě nebylo uvedeno, spíše jsem se snažila rodičům zhodnotit, jak si dítě v testu vedlo. Hodnotila jsem klinickou škálu a hovořila o případných problémech v konkrétním subtestu. Vzhledem k faktu, že si děti a jejich rodiče pro mé měření vyšetřili dvě dopoledne bez nároku na jakoukoli finanční kompenzaci, považuji

za spíše neetické nepodat žádnou zpětnou vazbu. Kvalitu této zpětné vazby jsem pak nastavila podle svého nejlepšího vědomí a svědomí.

Vrátíme-li se k naplnění standardů metody evokovaných potenciálů, musely vzhledem k výše uvedeným problémům přijít kompromisy. Hlavním z nich bylo opuštění původní myšlenky jednotlivých úrovní obtížnosti, protože nebylo možné zajistit alespoň 30 triálů na dítě, což je dle Lucka (2014) stále robustní počet. Pokud si tedy ilustrujeme na výsledcích získaných pro jednotlivé obtížnosti (viz obrázek č. 1), vidíme, že lze získat statisticky signifikantní výsledky i po Bonferroniho korekci. Takové výsledky jsou ale principiálně chybné, jelikož jsou získané na nízkém počtu triálů, což posléze vidíme na obrázku č. 2, kde je jasně patrné, že získaná fyziologická křivka reflektuje tento problém. Zároveň byla zahrnuta pouze malá část parietálních kanálů, díky čemuž byla uměle zvýšena statistická síla zvolené korekce. Yeung (2018) podobné trendy v manipulaci s daty přílehně označuje jako v překladu „vybírání třešniček“ (angl. *cherry picking*), čehož jsem se ve své práci snažila vyvarovat. Vždy uvádím kompletní skalpové projekce a do topografií jsou vždy zahrnuty všechny kanály. Navíc dávám k dispozici konkrétní rozložení p hodnoty k nahlédnutí.

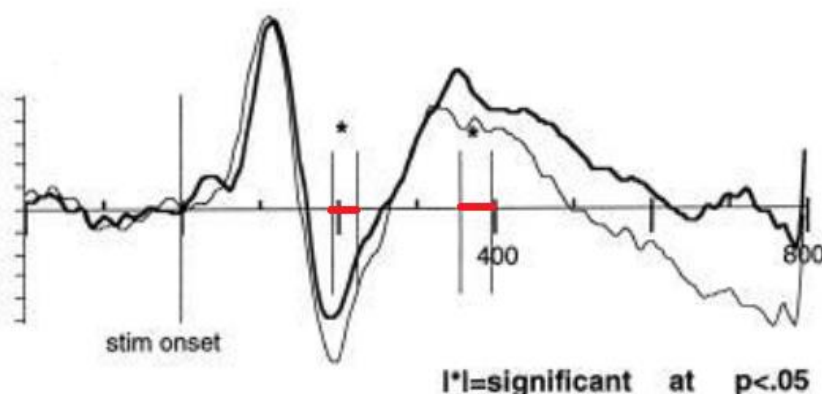


Obrázek 1 Úroveň obtížnosti 2.3 pro liché parietální elektrody s Bonferroniho korekcí na hladině významnosti $p < 0.05$.

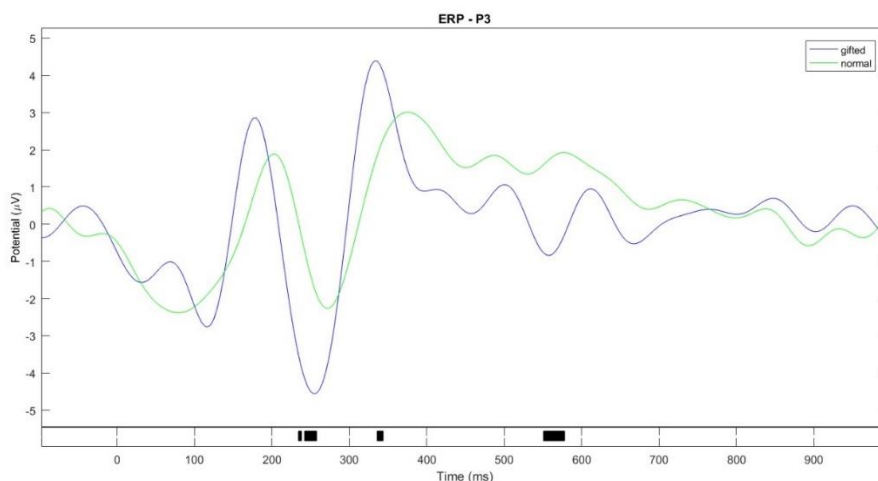


Obrázek 2 Evokovaný potenciál pro úroveň obtížnosti 2.3 na elektrodě CZ s Bonferroniho korekcí, $p < 0.05$

Fakt, že nepoužívám statistické korekce ve své disertační práci, nijak neskrývám, naopak absenci korekcí u výsledků sama uvádím. Jde samozřejmě o problematický krok a připadá tak v úvahu námitka oponenta o falešně pozitivních výsledcích, se kterou se neurověda, jak jsem již nastínila, potýká. Při zpracování signálu jsem se soustředila na naplnění standardů pro zpracování zvolené EEG analýzy. Při svém rozhodnutí nepoužít korekci jsem se opírala zejména o konkrétní podobu získaných výsledků a jejich porovnání s již realizovanými studiemi, které v práci uvádím. Na základě jejich konsensu jsem se rozhodla korekce nepoužít. Uveďme si na konkrétním případu studie Posnera (In Tempelová, 1998). Na obrázku č. 3 vidíme podobu evokovaného potenciálu získaného na parietální elektrodě v časovém okně zhruba -200 až 800 ms a na obrázku č. 4 pak námi získaný výsledek. Srovnání je samozřejmě problematické, jelikož v samotné disertaci publikuji získané výsledky na více než 23 stranách dle různých obtížností úlohy, nicméně pro povahu ilustrace podobnosti křivky si srovnáme alespoň na jednom příkladu.

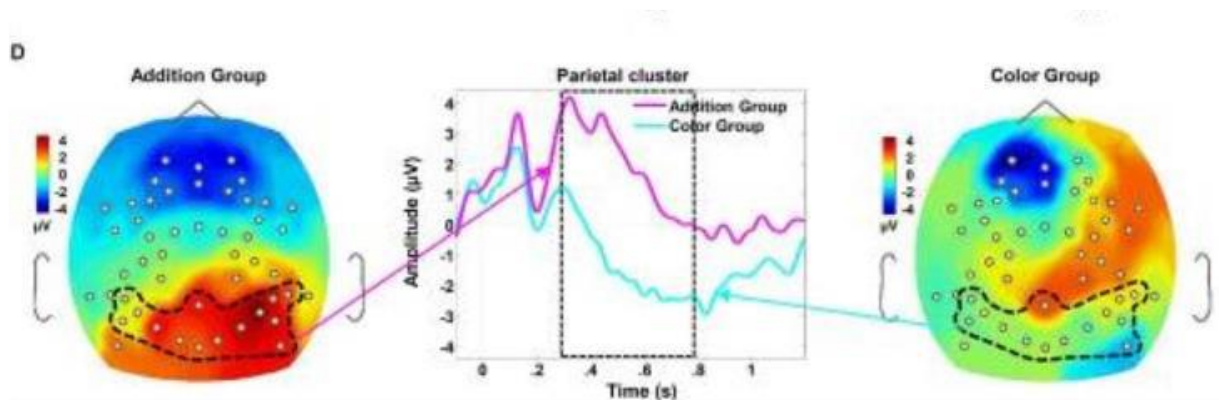


Obrázek 3 Získané ERP u dětí (In Templová, Posner, 1998).



Obrázek 4 ERP u předškolních dětí na elektrodě P3 pro lehkou obtížnost ($p < 0.05$).

Posner (tamtéž) uvádí nalezené signifikance okolo neurokomponenty N2 a P2p. Ve své práci adresují problém s jitterem (tj. nežádoucí časová odchylka v signálu), nicméně i přes tento fakt se všechny signifikance neurokomponent objevují v časovém rozmezí, která jsou pro tyto vlny typická. Goulet (2018) pak na základě studie u školně zralých dětí přináší výsledky, na kterých je jasně patrná signifikance ($p < 0.01$) v pozdních posteriorních komponentách (viz obrázek č. 5). Přičemž autor používá také permutační statistiku, jako je to i v mém případě, a zároveň používá i FDR korekci. To vše u školně zralých dětí, což samozřejmě představuje určitý problém, protože, jak jsem již v disertační práci uvedla, všechny pozorované neurokomponenty procházejí určitou maturací, nicméně všechny mnou publikované výsledky se objevují v časech, které jsou z literatury jasně specifické pro očekávanou mozkovou aktivitu. Navíc získané fyziologické křivky dosahují srovnatelného tvaru i výšky amplitudy, jakožto u již publikovaných výsledků předchozích studií.



Obrázek 5 Komponenta P2p. Modře skupina s AAT tréninkem, fialově bez tréninku (In Gouet. 2018).

Navíc jsou již dostupné výsledky Kratochvílové (2019), která za použití stejných výzkumných úloh v rámci své bakalářské práce zkoumala aproximální numerický systém u vysokoškolských studentů psychologie. V rámci jejího zjištění vyplývá, že všechny nalezené signifikance ($p < 0.01$ bez korekce) se objevují v podobných časových úsecích jako u mnou naměřených skupin. Bohužel při psaní tohoto dodatku se mi nepodařilo získat statistické srovnání časů a amplitudy s ostatními studii. Narazila jsem jednoduše na problém, že v současné chvíli není dostupný standardizovaný postup, jak relevantně porovnávat evokované potenciály. V rámci obdrženého posudku jsem tak narazila na závažný problém v metodologickém úzu pro srovnávání s předchozím výzkumem. **Jeho řešením by patrně mohla být změna zvolené EEG analýzy z evokovaných potenciálů na analýzu spektrální, která následně umožňuje převod EEG dat po Fourierově transformaci do cvs souboru. Taková data jsou následně přístupna klasickému statistickému zpracování.** V současné chvíli neexistuje dostatek studií, které by se frekvenční analýzou aproximálního numerického signálu u dětí zabývaly, tudíž pro potřeby srovnání byly v disertační práci zvoleny evokované potenciály. Získat později spektrální analýzu bylo mým záměrem již při sběru a předzpracování EEG dat, což je patrné v nastavení lowpass filtru na 120 Hz, ačkoli je zvykem pro evokované potenciály tento filtr nastavovat na 30 Hz (případné distorze signálu zmiňuji v disertační práci). Aktuálně spolupracuji s odborníky na statistickém zpracování signálu i s odborníky pro práce v prostředí programu Matlab, jelikož potřebná hloubka znalosti programovací syntaxe není v mé kompetenci.

Vrátíme-li se však k výsledkům Kratochvílové (2019), pak **všechny autorkou získané výsledky podporují závěry mé disertační práce, a to tedy, že výška amplitudy a čas nástupu neurokomponent N1, N2, P2p a pozdní posteriorní komponenty mohou sloužit jako indikátory inteligence a zjevně se tento závěr nevztahuje pouze na předškolní děti.**

Nicméně pro zobecnění těchto výsledků je potřeba pečlivé statistické analýzy, na které v této chvíli usilovně pracujeme.

Literatura

1. Buttonová, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews, Neuroscience*.14.
2. Gouet, C., Gutiérrez Silva, C. A., Guedes, B., Peña, M. (2018). Cognitive and Neural Effects of a Brief Nonsymbolic Approximate Arithmetic Training in Healthy First Grade Children. *Frontiers in Integrative Neuroscience*, 12, 28.
3. Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med.* 2, e124.
4. Kratochvílová, D. (2019). *Numerozita u vysokoškolských studentů*. Č. Budějovice, 2019. bakalářská práce (Bc.). Jihočeská univerzita v Českých Budějovicích. Pedagogická fakulta
5. Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press.
6. Nakagawa, Shinichi (2004). "A farewell to Bonferroni: the problems of low statistical power and publication bias". *Behavioral Ecology*. 15 (6): 1044–1045.
7. Rosier, J. P., Linden, D. E., Gorno – Tempinin, M. L., Moran, R. J., Dickerson, B. C., Grafton, S. T. (2016). Minimum statistical standards for submissions. *Neuroimage: Clinical*. 12: 1045–1047.
8. Templová, E., Posner, M. (1998). Brain mechanisms of quantity are similar in 5-year-olds and adults. *Proceedings of the National Academy of Sciences*. 95 (13), 7836–7841.
9. Woo, C. W., Krishnan, A., Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *Neuroimage*. 91: 412–419 .
10. Xu, Y., Nguyen, D., Mohamed, A., Carcel, C., Li, Q., Kutlubaev, M. A., Anderson, C. S., Hackett, M. L. (2016). Frequency of a false positive diagnosis of epilepsy: A systematic review of observational studies. *Seizure*. 41: 167–174.
11. Yeung, A. W. K. (2018). An Updated Survey on Statistical Thresholding and Sample Size of fMRI Studies. *Frontiers in Human Neuroscience*. 12 (16).