

**UNIVERZITA KARLOVA
FILOZOFICKÁ FAKULTA
KATEDRA PSYCHOLOGIE**



Diplomová práce

Kateřina Koppová

Zapamatování fragmentovaných scén – role detailů a extrapolace

**Visual Memory of Fragmented Scenes – Role of Detail and
Extrapolation**

Praha 2019

Vedoucí práce: Mgr. Jiří Lukavský, Ph.D.

Ráda bych na tomto místě poděkovala Mgr. Jiřímu Lukavskému, Ph.D. za trpělivé a laskavé provázení celým procesem od zrodu experimentu a za cenné rady nejen k diplomové práci samotné, ale také k vědecko-výzkumné činnosti v její obecnosti. Dále děkuji Mgr. Filipovi Děchtěrenkovi, Ph.D. za velikou podporu odbornou i hluboce lidskou. Děkuji také LABELS za umožnění provedení experimentu v jejích prostorách. Děkuji všem participantům za zájem zúčastnit se experimentů a přispět k vědeckému poznání. Z celého srdce děkuji také své rodině a nejbližším, kteří mě vždy nesmírně podporovali a nepřestali ve mně věřit.

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 24. dubna 2019

.....
Kateřina Koppová

Abstrakt

Každý den jsme vystaveni velkému množství vizuálních podnětů (scén). Není však stále jasné, jak dobře jsou tyto scény v paměti uloženy a pamatovány a jaká je role detailu vzhledem k paměti celého obrazu. Konkrétně, do jaké míry jsou pamatovány detaily a do jaké míry extrapolujeme z neviděného obsahu z prezentovaných částí. K prozkoumání této otázky jsme využili fragmentovaných scén (rozčleněných na 4x4 dílků), které se skládaly ze tří různých kategorií (vnitřní, vnější-přírodní a vnější-umělé) s různým počtem prezentovaných dílků (3, 5 a 8). Zajímá nás, jestli je nějaký vztah mezi počtem prezentovaných dílků a schopností rozpoznat, které dílky byly prezentované a které ne (přestože byly ze stejné fotografie). V analýze jsme se soustředili na charakteristiky SDT (signal detection theory), především na citlivost paměti (d') a bias. Byly provedeny dva experimenty a u obou byly zjištěny největší hodnoty správnosti v odpovědích u 3 zobrazených dílků (Exp.1: $d'(n3) = 0,67$; Exp.2: $d'(n3) = 0,66$), u kterých převažovala tendence biasu k záporným odpovědím (Exp.1: $c(n3) = 0,27$; Exp.2: $c(n3) = 0,16$). U počtu 5 a 8 dílků se správnost odpovědí snížila (Exp.1: $d'(n5) = 0,35$; $d'(n8) = 0,34$; Exp.2: $d'(n5) = 0,39$; $d'(n8) = 0,41$) a zároveň se bias přiklonil ke kladným odpovědím (Exp.1: $c(n5) = -0,11$, $c(n8) = -0,34$; Exp.2: $c(n5) = -0,16$, $c(n8) = -0,37$). V neposlední řadě nás také zajímal možný vliv mřížky na zapamatování, který se však neprokázal.

Klíčová slova: vizuální krátkodobá paměť, signal detection theory, fragmentované scény, mřížka, kategorie scén, viděné/neviděné dílky

Abstract

Every day we are exposed to huge amounts of visual stimuli (scenes). However, it is not yet clear how accurately these scenes are stored and remembered, and what is the role of detail relative to the memory for the whole image. More specifically, to what extent the details are remembered and to what extent we extrapolate the unseen content from the presented details. We used fragmented scenes (broken to 4x4 grids) to investigate this question. Fragmented scenes were composed of three different theme categories (indoor, natural, man-made), each with different number of presented patches (3, 5 and 8). Our main research question is, whether there is any relationship between the number of presented patches and the ability to recognize which patches were presented and which were not (but still patches from the same photograph). In analysis we focus on Signal Detection Theory characteristics, mainly memory sensitivity (d') and bias. We run two experiments and in both the highest scores for (d') were for 3 patches (Exp.1: $d'(n3) = 0,67$; Exp.2: $d'(n3) = 0,66$) with bias towards negative answers (Exp.1: $c(n3) = 0,27$; Exp.2: $c(n3) = 0,16$). For 5 and 8 patches the (d') was lower (Exp.1: $d'(n5) = 0,35$; $d'(n8) = 0,34$; Exp.2: $d'(n5) = 0,39$; $d'(n8) = 0,41$) and in the same time bias was towards positive answers (Exp.1: $c(n5) = -0,11$, $c(n8) = -0,34$; Exp.2: $c(n5) = -0,16$, $c(n8) = -0,37$). We have also examined the role of the grid for remembering, but we have found no effect.

Key words: visual short-term memory, signal detection theory, fragmented scenes, grid, categories of scenes, seen/unseen patches

Obsah

Úvod	11
Teoretická část	
1) Vizualní paměť	12
1.1 Vizualní sensorická paměť	12
1.1.1 Viditelná perzistence	13
1.1.2 Informační perzistence	13
1.2 Vizualní krátkodobá paměť	13
1.2.1 Vymezení vizualní krátkodobé paměti mezi ostatními paměťovými systémy	14
1.2.2 Kapacita	16
1.2.3 Podoba reprezentací	20
1.3 Vizualní dlouhodobá paměť	23
2) Vnímání scény a její uspořádání	26
2.1 Vizualní scéna	26
2.2 Vnímání scény na základě vlastností fotografie	27
2.3 Kategorizace scén	28
2.4 Memorabilita fotografie	29
2.5 Vnímání scény na základě jejich jednotlivých prvků	32
2.5.1 Okraje a kontrast	32
2.5.2 Umístění a barevnost	32
2.5.3 Velikost objektů	33
2.6 Vnímání scény na komplexnější úrovni	34
2.6.1 Podstata scény (gist)	34
2.6.2 Přístup zaměřený na objekty a na celou scénu	35
2.6.3 Centrální a periferní vidění	35
2.6.4 Důležitost kontextu ve scéně	35
2.7 Narušení vztahů ve scéně a její fragmentace	36
2.7.1 Neobvyklé prostorové umístění objektu	37
2.7.2 Fragmentace scény	37
2.7.3 Mřížka	40
2.8 Extrapolace	41

Empirická část

3) Cíl výzkumu a výzkumné otázky	43
4) Experiment 1	45
4.1 Metody	45
4.1.1 Participanti	45
4.1.2 Design a stimuly.....	45
4.1.3 Procedura	50
4.2 Výsledky.....	53
4.2.1 Reakční čas	53
4.2.2 Signal detection theory (SDT).....	54
4.2.3 Regresní analýza	61
4.3 Diskuze.....	61
5) Experiment 2	63
5.1 Metody	63
5.1.1 Participanti	63
5.1.2 Design a stimuly.....	63
5.1.3 Procedura	63
5.2 Výsledky.....	63
5.2.1 Reakční čas	63
5.2.2 Signal detection theory	64
5.2.3 Regresní analýza	69
5.2.4 ANCOVA.....	70
5.3 Diskuze.....	72
Závěr.....	74
Seznam použité literatury.....	75

Seznam zkratek

CR	<i>(correct rejection)</i>	správně záporná odpověď na stimul, který nebyl přítomen
Exp.1		Experiment 1
Exp.2		Experiment 2
F	<i>(false alarm rate)</i>	pravděpodobnost kladných odpovědí v případě nepřítomného stimulu
FA	<i>(false alarm)</i>	nesprávně kladná odpověď na stimul, který nebyl přítomen
FN	<i>(false negative)</i>	nesprávně záporná odpověď na stimul, který byl přítomen (označení také MISS)
FP	<i>(false positive)</i>	viz FA
I	<i>(indoor)</i>	scéna kategorie vnitřní-umělá
H	<i>(hit rate)</i>	pravděpodobnost kladných odpovědí v případě přítomného stimulu
LABELS		Laboratoř behaviorálních a lingvistických studií
M	<i>(man-made)</i>	scéna kategorie vnější-umělá
N	<i>(natural)</i>	scéna kategorie vnější-přírodní
RT	<i>(reaction time)</i>	reakční čas
S	<i>(seen)</i>	patch, který byl součástí prezentované fragmentované fotografie a zároveň tak jedné ze tří původních fotografií
SDT	<i>(signal detection theory)</i>	teorie detekce signálu
STM	<i>(short-term memory)</i>	krátkodobá paměť
SUN, SUN397	<i>(Scene UNderstanding)</i>	databáze k porozumění scén („SUN database“)
TN	<i>(true negative)</i>	viz CR
TP	<i>(true positive)</i>	správně kladná odpověď na stimul, který byl přítomen (označení také HIT)
U	<i>(unseen)</i>	patch, který nebyl součástí prezentované fragmentované fotografie, avšak byl součástí jedné ze tří původních fotografií
VLTM	<i>(visual long-term memory)</i>	vizuální dlouhodobá paměť
VSM	<i>(visual sensory memory)</i>	vizuální sensorická paměť

VSTM	(<i>visual short-term memory</i>)	vizuální krátkodobá paměť
WM	(<i>working memory</i>)	pracovní paměť

Seznam obrázků

Obrázek I. Dělení VSTM (Squire, 2004, 173)

Obrázek II. Typy scén kategorií a podkategorií

Obrázek III. Příklad konkrétních fotografií se zařazením do podkategorie a kategorie

Obrázek IV. Konečná podoba fragmentované fotografie vytvořené z předchozích příkladů na Obrázku III.

Obrázek V. Příklad vybraného patche z fragmentované fotografie, na který byl následně participant dotazován. Oranžové ohraničení a šipka jsou pouze názornou pomůckou a nebyly součástí experimentálních stimulů

Obrázek VI. Struktura experimentu v PsychoPy

Obrázek VII. Exp.1: Histogramy četnosti odpovědí vzhledem k reakčním časům

Obrázek VIII. Confusion matrix: rozložení HIT, FA, MISS a CR

Obrázek IX. Exp.1: Grafy průměrných hodnot HIT a FA vzhledem k počtu zobrazených dílků

Obrázek X. Exp.1: Grafy průměrných hodnot MISS a CR vzhledem k počtu zobrazených dílků

Obrázek XI. Exp.1: Graf průměrné přesnosti odpovědí vzhledem k počtu zobrazených dílků

Obrázek XII. Exp.1: Graf průměrných hodnot bias vzhledem k počtu zobrazených dílků

Obrázek XIII. Exp.1: Graf hodnot d' vzhledem k počtu zobrazených dílků

Obrázek XIV. Exp.2: Histogramy četnosti všech odpovědí vzhledem k reakčnímu času

Obrázek XV. Exp.1 a Exp.2: Graf průměrné hodnoty HIT vzhledem k počtu zobrazených dílků

Obrázek XVI. Exp.1 a Exp.2: Graf průměrné hodnoty FA vzhledem k počtu zobrazených dílků

Obrázek XVIII. Exp.1 a Exp.2: Graf průměrné hodnoty CR vzhledem k počtu zobrazených dílků

Obrázek XIX. Exp.1 a Exp.2: Graf průměrné přesnosti v odpovědích vzhledem k počtu zobrazených dílků

Obrázek XX. Exp.1 a Exp.2: Graf průměrného biasu vzhledem k počtu zobrazených dílků

Obrázek XXI. Exp.1 a Exp.2: Graf průměrného d' vzhledem k počtu zobrazených dílků

Obrázek XXII. Exp.1 a Exp.2: ANCOVA odhadovaných mezních hodnot průměru d' vzhledem k typu experimentu a počtu zobrazených dílků

Obrázek XXIII. Exp.1 a Exp.2: ANCOVA odhadovaných mezních hodnot průměru bias vzhledem k typu experimentu a počtu zobrazených dílků

Seznam tabulek

Tabulka I. Exp.1: koeficienty logistické regrese

Tabulka II. Exp.2: koeficienty logistické regrese

Tabulka III. Exp.1 a Exp.2: ANCOVA, párové srovnání d'

Tabulka IV. Exp.1 a Exp.2: ANCOVA, párové srovnání bias

Úvod

Tato práce se zaměřuje na vizuální paměť pro fragmentované scény, roli detailů a extrapolace. V první kapitole se věnuji typům vizuální paměti, ve druhé kapitole náležitostem scény a jejich vnímání a pamatování. Empirická část je zaměřena na popis dvou provedených výzkumů, které se pomocí SDT charakteristik snažili zodpovědět otázku, jakou roli detaily a extrapolace při pamatování fragmentovaných scén hrají.

Teoretická část

1) Vizuelní paměť

Paměť, mluvíme-li o paměti v její obecnosti a celistvosti, je složena z mnoha podčástí a vizuelní paměť je jednou z nich. Na základě smyslových modalit můžeme rozlišit i další podtypy paměti, např. echoickou (sluch) či olfaktorickou (čich). Jako takové všechny podléhají stejným procesům, které se v paměti odehrávají, ačkoliv jejich průběh může být z hlediska druhu paměti odlišný: kódování (ukládání informací), retence (udržení informace), reprodukce (vybavování si informací) a zapomínání (ztráta informací).

Vizuelní paměť je na podkladě modálního modelu paměti Atkinsona a Shiffrina (1968) tradičně rozdělována do tří hlavních podčástí: vizuelní sensorická paměť (*visual sensory memory, VSM*), vizuelní krátkodobá paměť (*visual short-term memory, VSTM*) a vizuelní dlouhodobá paměť (*visual long-term memory, VLTM*) (Luck & Hollingworth, 2008). V této kapitole si blíže představíme jednotlivé podčásti vizuelní paměti a zaměříme se na to, co o nich doposud bylo zjištěno a jakým způsobem jsou pojímány.

1.1 Vizuelní sensorická paměť

Jak můžeme usuzovat již podle samotného názvu, sensorická paměť přijímá a uchovává informace pocházející přímo ze smyslů. Podle toho, o jaké modalitě mluvíme, označujeme sensorickou paměť například jako *ikonickou*, v případě vizuelní sensorické paměti uchovávající informace ze zrakových podnětů, nebo jako *echoickou*, v případě audiální sensorické paměti uchovávající informace ze sluchových podnětů (Plháková, 2004). Tato terminologie byla poprvé užita Neisserem (2014). V českém názvosloví se můžeme setkat i s užíváním označení *ultrakrátká paměť*, jakožto ekvivalentu sensorické paměti obecně (viz např. Nakonečný, 2016; Plháková, 2004). Tímto označením se klade důraz spíše na dobu podržení informace v paměti než na kvalitu a podobu přijímané a uchovávané informace.

Neisser (2014, 16) upozorňuje na mylné předpoklady, které se často implicitně při výzkumech VSM vyskytují a shrnuje je pod pojmem „*naivní realismus*“:

- 1) „*přesvědčení, že vizuelní zkušenost přímo zrcadlí podobu podnětu*“
- 2) *přesvědčení, že vizuelní zkušenost začíná s počátkem vystavení podnětu a končí s jeho ukončením*
- 3) *přesvědčení, že vizuelní zkušenost, která je sama o sobě pasivní kopií podnětu, se může zrcadlit ve verbálním popisu této zkušenosti“*

(Neisser, 2014, 16)

V případě VSM dochází k rychlému poklesu udržení informace v paměti v čase, a to v průběhu jedné vteřiny. Zároveň je VSM považována za nízkourovňový jev, a jako taková je citlivá k následnému maskování podnětu (Luck & Hollingworth, 2008). Maskování podnětu znamená překrytí podnětu vizuálním šumem, který se obvykle skládá z náhodně rozmístěných bílých a černých čtverců, čímž dochází k přerušení vizuálního toku (Quinlan & Dyson, 2008). Takové maskování se nazývá *zpětné (backward masking)* (Quinlan & Dyson, 2008). Zpracování vizuálních informací na nízké úrovni znamená jejich zpracování např. z hlediska barvy, jasu nebo kontrastu (Groen, Silson, & Baker, 2017). S ohledem na podobu přetrvávající vizuální informace jsou rozlišovány dva typy perzistence, *viditelná* a *informační* (Luck & Hollingworth, 2008).

1.1.1 Viditelná perzistence

Viditelná perzistence (někdy také označována jako perzistence fenomenologická) je jev, ke kterému dochází po skončení působení vizuálního stimulu. Poté, co byl vizuální podnět prezentován a následně odstraněn, zůstává i přesto nadále viditelný, a to přibližně 100 ms. Toho se často využívá u různých technických zařízení, např. u televizí či počítačů (Luck & Hollingworth, 2008). Peirce (2018) uvádí, že obnovovací kmitočet obrazu je u většiny plochých počítačových monitorů 60 Hz. To znamená, že za 1 minutu dojde 60x k obnovení obrazovky. Takže zatímco člověk má dojem, že to, co vnímá, je vizuální nepřerušovaný tok informací, ve skutečnosti jsou tyto informace poskytovány přerušovaně, jako jednotlivé obrazy jdoucí ve vysoké rychlosti za sebou. A právě viditelná perzistence umožňuje v mysli vytvořit zdání jednolitého celku (Luck & Hollingworth, 2008).

1.1.2 Informační perzistence

Jedná se o jev, ke kterému dochází také po skončení působení vizuálního stimulu, ale oproti viditelné perzistenci zde nedochází k jeho viditelnosti. V paměti zůstává zhruba po dobu 500 ms pouze vyabstrahovaná informace o stimulu (Luck & Hollingworth, 2008).

1.2 Vizuální krátkodobá paměť

VSTM je spolu s VSM a VLTM podsystémem vizuální paměti. Informaci je schopna udržet po dobu několika vteřin až minut (Nakonečný, 2016). Z toho plyne i její funkce, udržet informaci přes sakády (oční pohyby) (Xu & Chun, 2006), aby byl mozek schopen skládat komplexnější informace po delší časový úsek, než opět přesuneme svůj pohled z místa na místo, čímž bychom pouze na základě VSM všechny informace vzápětí ztratili.

Protože jsme se při našem výzkumu zaměřili právě na spojitost mezi fragmentací scény a zapamatováním ve VSTM, bude této podčásti věnovaný náležitý prostor a pozornost. Nejprve si jednotlivé podsystémy přiblížíme vzájemným srovnáním, a to jak terminologicky, tak i na základě jejich podobností a rozdílností v jejich fungování. Dále se blíže zaměříme na kapacitu VSTM neboli množství a kvalitu informací, které je VSTM schopna uchovat najednou. Badatelé se prozatím neshodují v teoretickém ukotvení fungování kapacity VSTM, neboť mnohdy ve svých studiích dospívají k výsledkům, které podporují rozdílné teoretické koncepty. V závěru této podkapitoly nahlédneme na podobu reprezentací ve VSTM, a tím se blíže seznámíme s tím, v jaké podobě jsou informace ve VSTM udržovány a jak dochází k vykompenzování poměrně nízké kapacity, kterou VSTM disponuje.

1.2.1 Vymezení vizuální krátkodobé paměti mezi ostatními paměťovými systémy

Jak jsem naznačila již v úvodu této kapitoly, vizuální paměť můžeme členit na VSM, VSTM a VLTM. V této části se zaměříme zejména na odlišnosti těchto systémů, a tudíž si nastíníme, proč je přínosné uvažovat o těchto konceptech jako o do jisté míry oddělených jednotkách a také jako o svébytných systémech, které však mezi sebou velmi úzce spolupracují. V poslední době se můžeme setkat také s pojmem *křehké vizuální krátkodobé paměti (fragile visual short-term memory)*, kterou někteří výzkumníci vnímají jako mezistupeň mezi VSM a VSTM (např. Pinto, Sligte, Shapiro, & Lamme, 2013). Pro potřeby této práce však není nutné se uchylovat k tomuto členění.

Mimo jiné se také v mnohých výzkumech můžeme setkat s rozdílným pojmoslovným aparátem, kterým je uchopována problematika krátkodobé paměti. Jedná se především o pojmy *krátkodobé paměti (short-term memory, STM)* a *pracovní paměti (working memory, WM)*. Proto se nejprve blíže podíváme právě na tyto terminologické nejasnosti, abychom se vyhnuli případným nesrovnalostem plynoucích z nejednoznačného vymezení pojmosloví.

Pracovní a krátkodobá paměť

Zatímco v některých výzkumech jsou STM a WM v zásadě užívány jako ekvivalenty téhož pojmu, v jiných případech jsou od sebe striktně rozlišovány, ať už jako pojmy soupeřící či pouze protichůdné. Tato zmatenost pramení z nepochopení možného teoretického propojení těchto dvou konceptů při představení jednoho z nich, pracovní paměti, jakožto nového pojmu Baddeleyem a Hitchem (1974). Jak ukazují Luck a Hollingworth (2008), tyto dva koncepty nemusí být nutně ve vzájemném konfliktu, ale na VSTM lze nahlížet jako na úložiště vizuálních informací, které je součástí širšího systému pracovní paměti (WM). V této práci budu pojmoslovně udržovat takto vymezenou linii.

Srovnání vizuální sensorické, krátkodobé a dlouhodobé paměti

Luck a Hollingworth (2008) představují srovnání VSTM a VLTM na základě čtyř vlastností, ve kterých se vzájemně odlišují. Jedná se o rychlost vytváření reprezentací, způsob udržování těchto reprezentací, kapacita jednotlivých typů paměti a množství informací, které jsou schopny pojmout. Pro přehlednější srovnání předkládám i informace k VSM, pokud tak lze z podstaty uváděných informací smysluplně učinit.

V případě rychlosti vytváření reprezentací se jedná o proces konsolidace a uchování reprezentací v paměti. U VSM po skončení prezentovaného stimulu dochází k postupné ztrátě informací, které jsou detekovatelné pouze ve formě perzistencí, nicméně nedochází k vytvoření reprezentací jako takových (Luck & Hollingworth, 2008). U VSTM se již reprezentace vyskytují, a to ve formě komplexních objektů spíše než jejich jednotlivých vlastností (Logie, Brockmole, & Vandenbroucke, 2009). Ve VSTM dochází k uložení informací zhruba v čase 50 ms na jednu položku (Vogel, Woodman, & Luck, 2006). Oproti tomu jsou reprezentace ve VLTM vytvářeny pomaleji (Luck & Hollingworth, 2008), nicméně jsou mnohem detailnější (Hollingworth & Henderson, 2002) a robustnější (Hollingworth, 2005b). Důvodem může být právě to, že reprezentace ve VLTM jsou založené na reprezentacích z VSTM, které jsou komplexní (Logie et al., 2009).

VSTM a VLTM se liší také ve způsobu udržování těchto reprezentací. Jak uvádí Luck a Hollingworth (2008), zatímco ve VSTM dochází k zániku reprezentací společně s aktivním udržováním těchto reprezentací (tzn., že reprezentace se vyskytují v paměti, dokud se je tam aktivně snažíme udržet), v případě VLTM reprezentace mohou přetrvávat neomezeně i bez našeho aktivního úsilí. To je způsobeno odlišnými mechanismy na nervové úrovni (Luck & Hollingworth, 2008).

Rozdíly nacházíme i v případě kapacity u jednotlivých typů paměti. U VSTM se co do počtu pamatovaných jednotek zdá kapacita nejnižší, a to v rozmezí 3-5 položek (Cowan, 2010). Liu a Jiang (2005) navíc zdůrazňují význam času prezentovaného materiálu pro množství zapamatovaných položek. V případě zobrazení stimulu na dobu 1 s si lidé pamatovali sotva 1 objekt, zatímco při zvýšení doby zobrazení na 15 s se zvýšil i počet zapamatovaných objektů na 5 (Liu & Jiang, 2005). Kapacita VSM a VLTM je oproti tomu považována za nesrovnatelně vyšší. U VLTM lze mluvit až o tisících velmi detailně uložených objektech (Brady, Konkle, Alvarez, & Oliva, 2008) a v případě VSM jsou uložené informace také velmi detailní (Sligte, Vandenbroucke, Scholte, & Lamme, 2010) a její kapacita se zdá takřka neomezená (Luck & Hollingworth, 2008). Zároveň Bradley a Pearson (2012) poukazují na to, že existuje postupný přechod mezi VSM a VSTM, kdy postupně dochází ke snižování počtu objektů, které jsou jednotlivé typy paměti schopny pojmout.

S výše uvedeným tak souvisí i poslední vlastnost, a to množství informací, které jsou jednotlivé typy paměti schopny pojmout. Zatímco u VSTM je tato schopnost velmi omezená, u VSM A VLTM je nesmírná (Luck & Hollingworth, 2008).

1.2.2 Kapacita

Teoretické přístupy ke kapacitě VSTM

Přestože by se z předchozího srovnání jednotlivých typů paměti mohlo zdát, že VSTM je schopna pojmout pouze omezený malý počet jednotek, situace není tak jednoznačná. Luck a Hollingworth (2008) rozlišují dva teoretické přístupy, jak lze kapacitu VSTM pojímat.

Prvním z nich je pohled na kapacitu jako na fixní a neměnnou vlastnost VSTM. V tomto přístupu je kapacita připodobňována k přesně vymezenému množství políček, do kterých je možno jednotlivé reprezentace ukládat. Drobné změny v počtu uložených jednotek jsou poplatné jejich dalším vlastnostem, jako je např. detailnost rozlišení uložené reprezentace. V takto omezeném prostoru dochází ke konfliktu reprezentací, které tak mezi sebou soupeří o omezený úložný prostor paměti (Luck & Hollingworth, 2008).

Druhý přístup považuje kapacitu za variabilní a flexibilní vlastnost VSTM. Přestože se předpokládají omezené zdroje pro schopnost uchovávat informace, jsou tyto zdroje navzájem sdílené a jsou přizpůsobovány dané situaci. Ve výsledku by tak např. při zvýšení počtu objektů, které je potřeba uchovat v paměti, docházelo současně i ke snížení detailnosti rozlišení uložených reprezentací (Luck & Hollingworth, 2008).

Konceptualizace kapacity VSTM ve výzkumech

Ve výzkumech se často používá dvou praktických přístupů, jak kapacitu VSTM měřit. V prvním případě se jedná o přístup *vysoké prahové hodnoty (high-threshold)*, kde se předpokládá, že v paměti reprezentace buďto je, a potom velmi přesně, anebo není vůbec. Ve druhém případě se jedná o výzkumy založené na *teorii detekce signálu (signal detection theory, SDT)*. V těchto případech se předpokládá, že reprezentace je důsledkem intervencí proměnných narušena. Reprezentace tak není uchována jako přesná kopie podnětu a musí být dosaženo určité prahové hodnoty k tomu, aby mohlo být rozhodnuto o tom, zda se jedná o prezentovaný stimul, nebo ne (Luck & Hollingworth, 2008).

a) *Fixní versus flexibilní přístup*

Franconeri, Alvarez a Cavanagh (2013) představují alternativní pohled na kapacitu VSTM. Oproti fixnímu přístupu, který je založen na přesně vymezených vzájemně nezávislých políčkách pro fixní počet položek je alternativním modelem přístup flexibilní, ve kterém je kapacita pojímána jako společný prostor pro informace, mezi kterými dochází k soupeření o omezený paměťový prostor (Franconeri et al., 2013).

Proti fixnímu modelu stojí zjištění Bradyho a Alvareze (2015b), že participanti si jsou schopni zapamatovat jednotlivé informace pouze z 1-2 komplexních objektů. V opozici k těmto zjištěním vystupují se svými výsledky Awh, Barton a Vogel (2007), kteří zastávají názor, že počet reprezentací ve VSTM je fixní bez ohledu na jejich komplexnost, a tudíž bez ohledu na detailnost uložených informací. Toto tvrzení podkládají zjištěním, že přestože i v jejich studii docházelo ke snížení kapacity s narůstající složitostí ukládaného materiálu, přisuzují tento jev procesuálním chybám při zpracovávání podobných podnětových setů, spíše než schopnosti paměti jako takové (Awh et al., 2007). Alvarez a Cavanagh (2004) ve svém výzkumu nicméně ukazují, že kapacita je limitována jak počtem objektů, tak i informačním zatížením, které daná položka ve VSTM vyvolává.

Zhang a Luck (2008) odmítají vysvětlení kapacity na základě flexibilní alokace zdrojů a přiklánějí se spíše k fixnímu modelu. Ve svém dalším výzkumu Zhang a Luck (2011) testovali, zda jsou participanti skutečně schopni zvýšit množství zapamatovaných položek snížením detailnosti, ve které si je budou pamatovat, jak předpokládá flexibilní přístup. Toho se však participantům v jejich experimentu dosáhnout nepodařilo (Zhang & Luck, 2011). Machizawa, Goh a Driver (2012) nicméně ukázali, že při malém počtu položek je naopak možné zvýšit detailnost jejich reprezentace.

Další, kteří svými zjištěními podporují fixní model VSTM jsou také Xu a Chun (2009), kteří poukázali na způsob ukládání jednotlivých objektů do paměti nejprve skrze vyčlenění fixního počtu objektů z percipované scény a následného detailnějšího zpracování těchto vybraných objektů. S tím je v souladu i zjištění Fougnieho, Asplunda a Maroise (2010), že z čím více vlastností se objekt skládal, tím více se zhoršila detailnost reprezentací, přestože počet uložených objektů ve VSTM zůstal stejný bez ohledu na množství prezentovaných znaků.

Přestože dosud nebylo jednoznačně vysvětleno, na jakém principu kapacita VSTM funguje (Luck & Hollingworth, 2008), Xu a Chun (2006) poukazují na důležitost různých neuronálních mechanismů, které se společně zapojují do fungování VSTM a skrze které je kapacita VSTM ovlivňována, a to jak v rámci počtu reprezentací, tak i jejich detailnosti rozlišení.

Suchow, Fougny, Brady a Alvarez (2014) nastiňují celkovou debatu mezi fixním a flexibilním přístupem a pomocí doplňujících otázek zaměřených např. na možný počet položek uložených ve VSTM, na přesnost jejich uložení nebo na podobu reprezentací, se snaží lépe objasnit teoretický prostor, ve kterém se tyto dva přístupy nacházejí jako pouhé dva body v mnohem širším prostoru.

Na výzkumném poli provedli zastánci fixního i flexibilního přístupu mnoho dalších experimentů na související témata, která dále prohlubují možné propojení s fixní či flexibilní hypotézou a přispívají tak k celkovému lepšímu propojení poznatků.

b) Individuální rozdíly a vývoj kapacity v průběhu života jedince

Rouder et al. (2008), kteří jsou zastánci fixního pojetí kapacity VSTM, z důvodu některých odlišných výsledků jedinců ve své studii upozorňují na důležitost individuálních rozdílů ve VSTM. Tyto individuální rozdíly můžeme pozorovat již na úrovni odlišných očních pohybů při prohlížení scén (Hayes & Henderson, 2017). Navíc jak poukazují Hayes a Henderson (2017), tyto rozdíly v odlišném způsobu prohlížení fotografií ze 40 % vysvětlují také odlišnosti v individuálních rozdílech na úrovni kapacity VSTM a inteligence jedinců. Spojením mezi fluidní inteligencí a kapacitou VSTM se zabývali také Fukuda, Vogel, Mayr a Awh (2010), kteří zjistili, že jejich vztah je sycen pouze počtem reprezentací, které mohou být uloženy ve VSTM, nikoli jejich kvalitou rozlišení.

Z hlediska vývoje kapacity VSTM v průběhu života jedince můžeme pozorovat veliké změny již mezi 6,5měsíčními a 7,5měsíčními kojenci (Oakes, Ross-Sheehy, & Luck, 2006). Zatímco v 6,5 měsících si kojenci nebyli schopni zapamatovat zároveň barvu a umístění objektu, v 7,5 měsících již toho schopni byli, a to stejně dobře jako kojenci ve 12,5 měsících (Oakes et al., 2006). Oakes et al. (2006) tento obrat dávají do souvislosti s vývojem zadní parietální kůry, která je zásadní pro přijímání sensorických informací a vykonávání pohybu (Orel & Procházka, 2017). Cowana (2017) zajímá vývoj VSTM od dětství do dospělosti a navrhuje zaměřit další výzkumy na prozkoumání toho, zda a jak se v průběhu těchto období mění počet a kvalita rozlišení jednotlivých reprezentací. Sám pak navrhuje hypotézu, že zatímco počet reprezentací se v průběhu života mění jen málo, k výrazným změnám dochází ve schopnosti uchovat detailnější reprezentace (Cowan, 2017).

Fougny, Suchow a Alvarez (2012) poukazují nejen na existenci interindividuálních rozdílů kapacity VSTM, ale také na kvalitativní rozdíly intraindividuální, tudíž rozdílnou schopnost zapamatování u jednoho jedince napříč různými pokusy ve stejném věkovém období. Tuto variabilitu dávají Fougny et al. (2012) do souvislosti se stochastickými procesy (*stochastic processes*) v mozku, při nichž dochází k náhodnému pálení neuronů. Tyto stochastické procesy jsou výhodné pro schopnost detekovat signál nebo pro rozhodovací procesy, u kterých vedou k pravděpodobnostnímu chování

(Deco, Rolls, & Romo, 2009). To znamená, že jako rušivý element vstupují do kognitivních procesů a převažují je k jedné či druhé alternativě, čímž zabraňují zablokování celého systému (Deco et al., 2009) a vytvářejí tak prostor pro rozdílné výkony u jednoho jedince. Tento rozměr fluktuace začlenili do svého modelu VSTM Van Den Berg, Shin, Chou, George a Ma (2012), kteří se tak na základě svých zjištění přiklánějí spíše k pohledu VSTM jakožto souvislého a proměnlivého zdroje nežli fixního počtu jednotek.

Právě neuronální podklad je častou příčinou individuálních rozdílů. Na základě odlišné neuronální aktivity je možno rozlišovat jedince s nízkou a vysokou kapacitou VSTM (Vogel & Machizawa, 2004). Ukazuje se, že zatímco jedinci s vysokou kapacitou VSTM jsou schopni ukládat do paměti pouze podstatné informace, jedinci s nízkou kapacitou VSTM ukládají do paměti i informace nepodstatné (Vogel, McCollough, & Machizawa, 2005). Tak dochází paradoxně k situaci, při které si jedinci s nízkou kapacitou VSTM ukládají do paměti více informací než jedinci s vysokou kapacitou VSTM (Vogel et al., 2005), nicméně kvůli tomu, že se jedná o větší množství nepodstatných informací, může být následně jejich výkon zkomplikován.

c) Předchozí znalosti a učení

Brady, Störmer a Alvarez (2016) odmítají fixní pojetí kapacity především ve spojení s poznatky o ukládání informací do paměti o objektech, se kterými se můžeme setkat v reálném světě. Brady et al. (2016) tak poukazují na zjištění, že zatímco ukládání jednoduchých prvků (např. barva) do paměti je vysoce omezené a po uplynutí určitého času již další informace do paměti nepřibývají, informace o reálných objektech se s rostoucím časem dále rozšiřují a lidé si zároveň pamatují i více reálných objektů. Tento jev dávají do souvislosti se znalostmi existujícími ještě před započítáním ukládání nových informací do paměti (Brady et al., 2016). Podobně Brady a Oliva (2008) ukázali, že lidé jsou schopni generalizovat naučené zákonitosti v kategoriích scén, a to díky předchozím znalostem konceptu těchto scén.

Olson, Jiang a Moore (2005) ukazují, že významnou roli ve VSTM hraje především schopnost naučit se rozlišit, která informace by měla být uložena. Curby a Gauthier (2007) zjistily, že v případě lidských tváří byl kapacitní limit VSTM překonán a účastníci si pamatovali mnohem více tváří. Na základě tohoto zjištění Scolar, Vogel a Awh (2008) provedli experiment, při kterém chtěli prověřit, zda expertiza v určité oblasti může zvyšovat kapacitu VSTM. Výsledky ukázaly, že experti byli schopni zvýšit detailnost zapamatovaných reprezentací, nikoli však jejich počet (Scolari et al., 2008). Curby, Glazek a Gauthier (2009) poukázali na to, že stupeň tohoto navýšení odpovídal míře expertních dovedností v rámci zkoumané oblasti. Nicméně když Olson a Jiang (2004) zkoumali, zda lze pomocí

opakovaných stimulů zlepšit zapamatování těchto stimulů, zjistili, že ve VSTM nebyl významný rozdíl ve výkonu mezi opakovanými a neopakovanými stimuly.

d) Rozdílné úlohy

Bengson a Luck (2016) zároveň poukazují na důležitost instrukce při zadávání úlohy, která tak může podporovat odlišné strategie pro zpracování daných informací. Jedinci tak s ohledem na zadání jsou schopni určité volní kontroly nad vlastnostmi objektů, které mají být do paměti uloženy (Woodman & Vogel, 2008). Fougnie, Cormiea, Kanabar a Alvarez (2016) navíc na základě odlišných instrukcí pro zapamatování si stimulů zjistili, že v určitých podmínkách si účastníci byli schopni zapamatovat větší počet podnětů. Tento nárůst v počtu zapamatovaných podnětů byl však doprovázen mnohem menší přesností pamatovaného materiálu (Fougnie et al., 2016). Fougnie et al. (2016) tak poukazují na určitou flexibilitu kapacity VSTM.

1.2.3 Podoba reprezentací

Zatímco se mnohé studie neshodnou na tom, jaký přístup by měl být uplatňován na kapacitu VSTM, mnohé z nich přinášejí zajímavé poznatky ohledně samotných reprezentací. Brady, Konkle a Alvarez (2011) tak poukazují na důležitost zaměřit se právě na podobu, v jaké jsou tyto reprezentace uchovávány. Reprezentace ve VSTM mohou být udrženy ve vší své detailnosti až po dobu několika sekund (Zhang & Luck, 2009). Suchow, Fougnie a Alvarez (2017) ukazují, že lidé si jsou schopni uvědomovat, v jakém stavu se reprezentace ve VSTM nacházejí a tyto informace využít pro své chování. Zatímco Suchow et al. (2017) přistupují k reprezentacím jako k postupně mizejícím, Zhang a Luck (2009) se domnívají, že k jejich rozpadu může dojít zcela a velmi náhle.

Jednotlivé vlastnosti a jejich kombinace ve VSTM

Luck a Vogel (1997) zjistili, že lidé si jsou schopni udržet ve VSTM nejen informace o jednotlivých vlastnostech objektů, jakými jsou barva nebo orientace, ale také kombinaci těchto dvou informací najednou. To samé se potvrdilo i v případě, kdy bylo nutné si zapamatovat až čtyři vlastnosti u jednoho objektu (Vogel, Woodman, & Luck, 2001). V obou studiích si však účastníci byli schopni uložit do paměti informace zhruba o čtyřech objektech (Luck & Vogel, 1997; Vogel et al., 2001). Fougnie et al. (2010) se zabývali otázkou, zda je VSTM spíše založená na jednotlivých vlastnostech nebo na celistvých objektech. Ve svém výzkumu zjistili, že zvyšující se množství vlastností v jednom objektu sice neovlivnilo počet zapamatovaných objektů, ale mělo za následek snižující se přesnost reprezentací

uložených v paměti (Fougnie et al., 2010). Tím ukázali, že udržení ve VSTM více vlastností, které jsou součástí jednoho objektu, je pro paměť nákladnější (Fougnie et al., 2010).

Jiang, Olson a Chun (2000) navázali na tento výzkum svými zjištěními, že ve VSTM jsou informace o jednotlivých objektech uloženy na základě vzájemných prostorových vztahů mezi objekty. Tyto vzájemné prostorové vztahy mohou přispívat ke zvýšení množství informací uložených ve VSTM (Delvenne, Braithwaite, Riddoch, & Humphreys, 2002). Woodman, Vecera a Luck (2003) zjistili, že objekty, které jsou při vnímání seskupeny u sebe, bývají také ukládány do VSTM společně.

Wheeler a Treisman (2002) ukázaly, že mají-li být do VSTM zároveň uloženy vlastnosti stejného typu (např. dvě barvy), dochází mezi nimi k soupeření o omezené místo v paměti. V případě, že se jedná o vlastnosti různého typu (např. barva a orientace), jsou uloženy obě informace současně bez nutné kompetice o místo ve VSTM (Wheeler & Treisman, 2002). Podobně i Xu (2002b) zjistil, že pro účastníky bylo jednodušší sledovat pouze jednu vlastnost ze dvou náležejících ke stejnému objektu při úloze zaměřené na zaznamenání změny (*change detection paradigm*), pokud obě tyto vlastnosti byly stejného typu. Pokud však sledované vlastnosti u jednoho objektu byly různého typu, byly zapamatovány lépe (Xu, 2002b). V úloze pro zaznamenání změny (*change detection paradigm*) je účastníkům prezentován stejný stimul jednou v původní podobě a podruhé s uměle vloženou změnou v krátkém časovém odstupu od sebe a úkolem účastníků může být zmáčknout klávesu ve chvíli, kdy si změny všimnou či konkrétně identifikovat změnu. Zároveň se užívá i různých modifikací tohoto paradigmatu, např. problikávání původního a pozměněného stimulu do sebe (*change detection flicker paradigm*). Neschopnost změnu zaznamenat se označuje jako slepota ke změně (*change blindness*).

Tato zjištění byla podpořena také Delvennem a Bruyerem (2004), kteří zdůrazňují důležitost pochopení celistvosti objektu jakožto důležitého kritéria pro integraci jednotlivých vlastností do jednoho celku. S tím souvisí i zjištění Alvareze a Cavanaghe (2008), kteří ukázali, že kapacita pro vlastnosti, které se vázaly k ohraničení objektu, byla mnohem větší než v případě vlastností, které souvisely s vnitřním povrchem objektu. Pokud nejsou jednotlivé vlastnosti přiřazeny k jednomu objektu, dojde k jejich horšímu uložení ve VSTM (Delvenne & Bruyer, 2004). Delvenne a Bruyer (2006) na předešlé výzkumy navazují zjištěním, že přestože je důležité v úloze zaznamenání změny (*change detection paradigm*), zda sledované vlastnosti patří k jednomu objektu, může toto zapamatování být podpořeno také celkovým uspořádáním scény.

Tato zjištění podpořil svým výzkumem také Xu (2002a), který navíc ukázal, že nejlépe zapamatovatelné jsou různé vlastnosti náležející stejné části objektu, o něco hůře zapamatovatelné jsou pak ty, které sice nejsou ve stejné části, ale stále jsou součástí jednoho objektu a nejméně

zapamatovatelné jsou ty vlastnosti, které pocházejí z různých objektů vyskytujících se na různých místech v prostoru. Xu (2006) se dále zabýval tím, jaké faktory přispívají k tomuto lepšímu zapamatování si jednotlivých vlastností v případě, že náležejí k jednomu objektu a zjistil, že je to prostorová blízkost (*proximity*), která umožňuje shlukování různých vlastností na základě blízkého prostorového umístění, a hlubší propojenost mezi nimi (*connectedness*), která způsobuje, že jednotlivé části považujeme za součást jednoho objektu. Lamsweerde, Beck a Johnson (2016) se ve svém výzkumu zabývali dalšími principy pro shlukování více vlastností do jednoho celku, a kromě již zmíněné propojenosti (*connection*) se zajímali také o princip podobnosti (*similarity*). Zjistili, že oba tyto principy bylo možné flexibilně využívat s ohledem na zadání úkolu, třebaže princip propojenosti (*connection*) byl využíván častěji, pokud bylo nutné mezi těmito principy zvolit pouze jeden (Lamsweerde et al., 2016).

Spojování informací ve VSTM

Brady, Konkle a Alvarez (2009) upozorňují na to, že v žitém světě pracuje paměť často se vztahy, které se v něm vyskytují. Na základě těchto asociací jsou vytvářeny pravděpodobnostní predikce, které mohou propojovat informace mezi sebou a tím umožňovat zapamatování více informací (Brady, Konkle, & Alvarez, 2009). Logie et al. (2009) ve svém výzkumu pozorovali, že participanti si lépe pamatovali spojení dvou vlastností, pokud se tato konfigurace vyskytovala ve všech prezentovaných případech. Brady a Tenenbaum (2010, 2013) nejenže vyvracejí představu o tom, že by do paměti měly být ukládané pouze jednotlivé informace nezávisle na kontextu, ale také předkládají pravděpodobnostní model paměti, který zohledňuje vztahy mezi objekty, popřípadě mezi položkami, a vytváření souhrnných informací vyššího řádu. Takovým mechanismem je shlukování informací z více objektů do jedné společné charakteristiky (např. průměrná velikost celého prezentovaného setu), čímž do jisté míry dochází k vyrovnání kapacitních limitů u VSTM (Alvarez, 2011).

Například se ukázalo, že informace o velikosti prezentovaného objektu není ve VSTM uložena nezávisle na ostatních položkách, ale naopak je porovnávána ve vztahu k nim (Brady & Alvarez, 2011). Long, Konkle, Cohen a Alvarez (2016) zjistili, že objekty stejné velikosti musejí mezi sebou sdílet určité percepční vlastnosti, kterými se vzájemně od sebe odlišují, neboť participanti byli schopni rychleji najít požadovaný objekt, nacházel-li se mezi ostatními objekty jiné velikosti. Brady a Alvarez (2015a) tak navrhují možnou podobu reprezentací ve VSTM jako hierarchicky strukturovaných, ve kterých se zohledňují informace nejen z jedné samotné položky, ale také vzájemné vztahy mezi všemi prezentovanými položkami. Na základě tohoto předpokladu se Brady a Alvarez (2015a) domnívají,

že ani kapacita VSTM udávaná v řádu jednotek není informací o nezávislé kapacitě VSTM bez těchto vzájemných vztahů, neboť i tento poznatek byl vyabstrahován na základě více kontextuálních položek.

Ukazuje se tedy, že ne všechny typy reprezentací stojí na stejné úrovni a podléhají tak stejným vzájemným vztahům. Haberman, Brady a Alvarez (2015) zjistili, že soubor reprezentací nízkourovňových vlastností prezentovaného stimulu (např. barva) koreloval se souborem stejné úrovně (např. orientace), ale byl vzájemně nezávislý s vysokourovňovými vlastnostmi (např. výrazy obličeje).

Chekaf, Cowan a Mathy (2016) mluví o schopnosti paměti vytvářet kusy (*chunks*) informací na základě jejich stlačitelnosti, kompresibilitě. Ve svém výzkumu tak například zjistili, že čím větší uspořádanost byla mezi sérií stimulů, která napomáhala možnosti informace shlukovat, respektive stlačovat do větších celků, tím větší množství informací uložených v paměti se objevovalo (Chekaf et al., 2016).

1.3 Vizuelní dlouhodobá paměť

Schurgin (2018, 1035) definuje VLTM jako „*pasivní sklad vizuelních informací po delší časové úseky*“. VLTM je součástí dlouhodobé paměti, která se skládá z mnoha dalších částí. Z jiného úhlu pohledu, než je příjem informací podle smyslů, můžeme dlouhodobou paměť dělit na část deklarativní (explicitní), což jsou ve vědomí přístupné informace, a procedurální (implicitní), což jsou nevědomé poznatky (Plháková, 2004). Deklarativní paměť dále můžeme rozdělit na paměť sémantickou (fakta) a epizodickou (události) a procedurální paměť na dovednosti a návyky, priming, klasické podmiňování a neasociativní učení. Každá z těchto podčástí je spojena s jinou částí v mozku, jak ukazuje přiložené schéma níže (Squire, 2004).

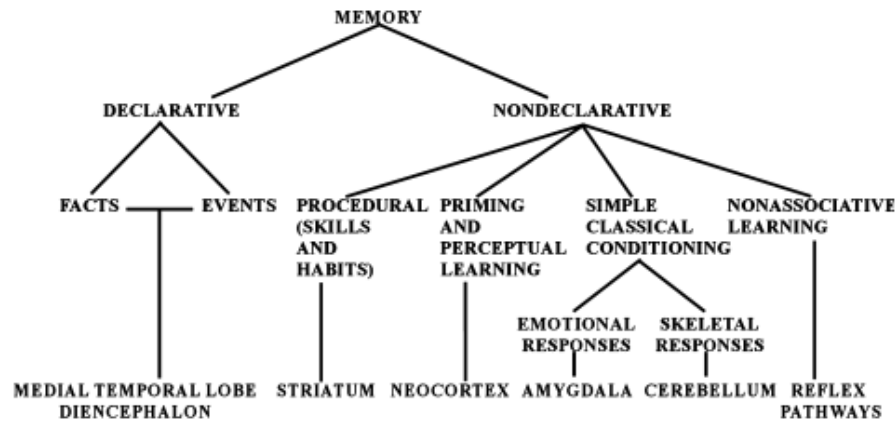


Fig. 1. A taxonomy of mammalian long-term memory systems. The taxonomy lists the brain structures thought to be especially important for each form of declarative and nondeclarative memory. In addition to its central role in emotional learning, the amygdala is able to modulate the strength of both declarative and nondeclarative memory.

Obrázek I. Dělení VLTM (Squire, 2004, 173)

Na základě společného srovnání čtyř vlastností mezi VSTM a VLTM, které provedli Luck a Hollingworth (2008) a se kterým jsme se blíže seznámili v předchozí podkapitole, jsme již měli možnost poznat podstatné znaky VLTM. Nyní si tyto poznatky prohloubíme dalšími zjištěními, nicméně pro komplexní pochopení toho, jak funguje ukládání a udržování vizuálních informací v paměti je důležité nazírat na VSTM a VLTM jako na dva úzce propojené a spolupracující systémy, spíše než na dvě oddělené části paměti. Jak ukazují Ranganath, Cohen a Brozinsky (2005), procesy probíhající v krátkodobé paměti při udržování informací významně ovlivňují ukládání informací do paměti dlouhodobé.

Brady et al. (2008) ukázali, že lidé si jsou schopni do VLTM uložit s velmi vysokou přesností (až 92%) tisíce objektů. Ve chvíli, kdy jsou do paměti ukládány objekty ze stejné kategorie, dochází ke snížení schopnosti udržet všechny konkrétní objekty v paměti (Konkle, Brady, Alvarez, & Oliva, 2010a). Tato zhoršená schopnost je nicméně podmíněna koncepční úrovní (např. co vše zařazují do konceptu „židle“), spíše než úrovní percepční (např. vizuálně mají všechny prezentované židle čtyři nohy, proto je náročnější udržet v paměti více objektů se čtyřma nohama) (Konkle et al., 2010a). Konkle, Brady, Alvarez a Oliva (2010b) dále provedli tento experiment nejen s objekty, ale také s celými scénami. Překvapivě zjistili, že výsledky se od sebe příliš nelišily, což by naznačovalo, že VLTM zachází s objekty a scénami podobně, přestože scény obsahují často obrovské množství objektů, a jsou tak považovány za vyšší a komplexnější celek (Konkle et al., 2010b). Lukavský a Děchtěrenko (2015) také zkoumali tzv. efekt příkladů (*exemplar effect*), neboli snížení výkonu paměti při prezentaci více příkladů z blízké kategorie scén, ale ve svém výzkumu nenalezli spojitost ani mezi výkonem paměti a

odlišností kategorie, ani mezi výkonem paměti a scénami z různých kategorií s podobnými globálními charakteristikami scény čili podstatou scény (*gist*), které si více přiblížíme v následující kapitole.

Schurgin a Flombaum (2015) se zaměřili na rozdíl v detailnosti VSTM a VLTM. Přestože v dřívějším výzkumu dospěli k závěru, že VLTM je schopná uložit stejně detailní informace jako VSTM (Brady, Konkle, Gill, Oliva, & Alvarez, 2013), Schurgin a Flombaum (2015) pozorovali horší výkon v neočekávaném testu zaměřeném na VLTM oproti výkonu v testech na VSTM, které však očekávané byly. Zároveň se tyto výkony lišily také v závislosti na typu úkolu, neboť zatímco VSTM nebyla narušena přidáním šumem do fotografie, výkon VLTM při zavedení šumu lineárně klesl (Schurgin & Flombaum, 2015). Zajímavé je také další zjištění, že přestože VLTM bývá považována za méně citlivou ke změnám v objektech pro jejich celkové rozpoznání, Schurgin a Flombaum (2018) ukázali, že při jejich testování byla naopak k těmto změnám méně citlivá VSTM a bylo v ní dosaženo lepších výkonů v rozpoznávání objektů i přes jejich variabilitu v zobrazení.

Brady, Konkle, Oliva a Alvarez (2009) poukazují na dostatek času pro uložení informací do VLTM nebo pozornost, kterou věnujeme každému z objektů, jako na důležité podmínky pro zaznamenání změn, které se s objekty v reálném prostředí mohou odehrávat. Jiang, Song a Rigas (2005) zkoumali, jakým způsobem dochází k zapamatování si opakovaně viděných scén, ke kterému v žitém světě běžně dochází. Opakované i nové scény složené z různě orientovaných písmen ve tvaru L a jednoho různě orientovaného T, které sloužilo jako hledaný objekt, byly participantům prezentovány po dobu pěti po sobě jdoucích dní a v každém z těchto dní měli participanté za úkol nalézt a určit orientaci T mezi ostatními písmeny L. Tento úkol byl participantům zadán ještě týden po posledním (pátém) dni učení. Jiang et al. (2005) zjistili, že ve všech případech byl hledaný objekt v opakovaných scénách vyhledán rychleji než ve scénách nových, čímž poukazují na schopnost paměti udržet velké množství opakovaných prostorových uspořádání.

Lidé se však velmi dobře učí i implicitně bez nutné volní snahy vše si co nejlépe zapamatovat. Castelano a Henderson (2005) ukázali, že participanté si nezáměrně vytvořili reprezentace vnímaných scén ve VLTM, i přestože netušili, že budou následně dotazováni na jejich zapamatování. Na schopnost ukládat informace i implicitně do dlouhodobé prostorové kontextuální paměti zase poukázali Chun a Jiang (2003).

2) Vnímání scény a její uspořádání

V této kapitole si vysvětlíme, co se přesně myslí vizuální scénou a na jakých podkladech funguje její rozpoznávání. Dále se podíváme na to, jaké tendence se obvykle u lidí při vnímání fotografií vyskytují. Následně se seznámíme s tím, jak lze scény kategorizovat a blíže prozkoumáme vnitřní vlastnost fotografie zvanou memorabilita. Přiblížíme si také jakou roli hrají ve scéně jednotlivé prvky, jako je barva, kontrast, okraje, umístění a velikost objektů, ale také komplexnější jevy, jakými jsou celé objekty, podstata scény (*gist*) a kontext. Nato prozkoumáme, jak může narušení vztahů ve scéně působit na ukládání a uchovávání informací v paměti a v závěru kapitoly si představíme termín extrapolace, který je pro naši výzkumnou otázku klíčový.

2.1 Vizuální scéna

Henderson a Hollingworth (1999, 244) vymezují *scénu* jako „*sémanticky koherentní (a často pojmenovatelné) zobrazení reálného prostředí zahrnující prvky pozadí a několik oddělených objektů uspořádaných prostorově umožněným způsobem*“. Oliva (2014) jako příklad uvádí kuchyň, ulici nebo les. Scéna je tedy jakýkoli výsek reality, který vidíme okolo sebe, když se rozhlédneme. O scéně mluvíme i tehdy, je-li zachycená na nějakém médiu, např. ve filmu nebo na namalovaném obraze. V našem případě pracujeme se scénami zachycenými na fotografiích, a proto následující poznatky budou vztahovány především ke vnímání fotografických scén. Tatler a Melcher (2007) ukázali, že oproti nefotografickým scénám (kresba, obraz) si účastníci lépe ukládají informace do paměti ze scén fotografických.

Přestože v žité realitě probíhá vnímání scén ve trojdimenzionálním prostoru, u fotografií dochází ke vnímání scén skrze dvoudimenzionální perspektivu. Cutting (2003) uvádí, že vnímání prostoru v obou těchto případech je založeno na stejných principech, přestože se oba typy scén v mnohém odlišují. Tatler (2008) upozorňuje na jejich rozdílnost a poukazuje na dynamickou sekvenci filmu jako na možný mezistupeň mezi reálnou scénou a scénami na fotografiích.

Vnímání a rozpoznávání scény se děje také z různých úhlů pohledu (Ehinger, 2013). Preferované úhly pohledu se označují jako *kanonické* a patří mezi ně pohled zřepředu (v některých případech pohled zezadu), pohled ze strany a pohled shora (Hall & Owen, 2005).

Vnímání scén na fotografiích a v trojdimenzionálním prostoru se také odlišuje v podobě reprezentací v paměti. Zatímco u vnímání statických scén jsou tyto reprezentace spíše abstraktní, nicméně informačně bohaté, v případě aktivních interakcí uvnitř scény jsou tyto reprezentace spíše konkrétní, zaměřené na úkol a informačně ne tolik obsáhlé (Tatler & Land, 2011).

2.2 Vnímání scény na základě vlastností fotografie

Vlastnosti technického zařízení také mohou ovlivňovat průběh vnímání. Jak například uvádí Peirce (2018), při obnovovacím kmitočtu obrazu u plochého počítačového monitoru dochází k obnovování obrazu odshora monitoru postupně k jeho spodní části, neboť se jedná o vertikální frekvenci. Rozdíl mezi vrchní a spodní částí může činit až 10 ms, což znamená, že vrchní část prezentovaného stimulu může být prezentována až o 10 ms dříve než část spodní (Peirce, 2018).

Dále se ukazuje, že existuje centrální tendence, tedy vnímání středu scény častěji než jejích vnějších částí (Tatler, 2007), a to i bez ohledu na vnímaný obsah (Clarke & Tatler, 2014). U jevu centrální tendence není zcela jasné, z jakého důvodu vzniká, nicméně zaměření se na střed scény se jeví jako optimální pozice pro její vnímání (Tatler, 2007). Sheth a Shimojo (2001) zjistili, že v souvislosti s centrální tendencí existuje také tzv. jev komprese prostoru (*compression of space*). Pozorovatelé mají sklon přisuzovat pozici vnímaného objektu blíže ke středu pohledu, oproti skutečné pozici objektu (Sheth & Shimojo, 2001).

Stejný sklon k častějšímu vnímání středu scény se objevuje i u scén s nízkým rozlišením obrazu (Judd, Durand, & Torralba, 2011). Judd et al. (2011) poukazují na to, že u velmi nízkého rozlišení (např. při výšce 4 pixely) je tento jev ještě silnější, přestože jej obecně lze nalézt všude. Torralba (2009) navíc zjistil, že teprve při rozlišení 32x32 pixelů je pozorovateli poskytnuto dostatečné množství informací k rozpoznání sémantické kategorie scény, což je slovně vyjádřené shrnutí scény. Sémantické informace o scéně však bývají participanty popisovány až následně po informacích sensorické úrovně (např. jas) a po informacích na úrovni vlastností (např. tvar) (Fei-Fei, Iyer, Koch, & Perona, 2007).

Dále existuje i levostranná tendence, tedy oční pohyby směřující k levé straně vnímané fotografie (Foulsham, Gray, Nasiopoulos, & Kingstone, 2013). Tato tendence bývá někdy spojována s převažujícím směrem čtení textu (zleva doprava) a dominantnosti ruky. Chokron a De Agostini (1995) našli významné rozdíly u této tendence mezi vzorkem Francouzů zvyklých psát zleva doprava a vzorkem Izraelitů zvyklých psát zprava doleva, a to již u dětí před zahájením školní docházky.

V neposlední řadě se Foulsham a Kingstone (2010) zaměřili na tendenci horizontální a vertikální a zjistili, že zatímco u fraktálů převažovala tendence horizontální, u přírodních scén prvotní oční pohyby směřovaly vertikálně, a to vždy kolmo na orientaci scény. Na bližší kategorizaci scén se zaměříme v následující podkapitole.

2.3 Kategorizace scén

Lidskou přirozeností je třídění a kategorizace světa. Do jedné kategorie zařazujeme různé objekty podle určitých společných prvků a vlastností. Uvnitř dané kategorie potom můžeme rozlišovat, zda vybraný objekt je více či méně typický v rámci vybrané kategorie. To lze použít i u scén. Scény lze kategorizovat do sémantických a funkčních skupin (Ehinger, Xiao, Torralba, & Oliva, 2011). Do sémantických skupin můžeme rozřazovat scény na základě objektů, které obsahují (Li, Su, Lim, & Fei-Fei, 2012). Greene, Baldassano, Esteva, Beck a Fei-Fei (2016) ukazují, že kategorizace scén probíhá sice na základě objektů a vnitřních vlastností scén, nicméně ne kvůli objektům a vlastnostem samotným, nýbrž proto, že odkazují k možným jednáním uvnitř scény a tím k samotné funkčnosti scény. Tyto vlastnosti odkazující k funkčnosti scény jsou pro kategorizaci scén důležité (Greene & Hansen, 2019).

Greene a Oliva (2009b) zjistily, že participanti dokázali se 75% správností kategorizovat prezentované stimuly scén, a to již při pouhých 19 – 67 ms expozice. Participanti však dokázali rychleji zařadit scénu na základě globálních vlastností (*global-property*), než ji zařadit do její základní pojmové kategorie (*basic-level*) (Greene & Oliva, 2009b). To konkrétně znamená, že participanti byli schopni rychleji rozlišit, že se jedná např. o rozlehlé prostředí (*global-property*), nežli o hory (*basic-level*) (Greene & Oliva, 2009b).

Joubert, Rousselet, Fize a Fabre-Thorpe (2007) ukázali, že při 26 ms prezentaci stimulů byli participanti schopni s 96% správností rozlišit, zda se jedná o přírodní (*natural*) či umělé prostředí (*man-made*), a to již při průměrném reakčním čase okolo 390 ms. Pokud měli participanti za úkol při stejné době expozice stimulů rozlišit, zda se jedná o prostředí nejen přírodní (*natural*) či umělé (*artificial*), ale také o prostředí vnitřní (*indoor*) či vnější (*urban*), průměrný reakční čas se zvýšil na 400-460 ms (Rousselet, Joubert & Fabre-Thorpe, 2005). Rousselet et al. (2005) zároveň zjistili, že přírodní scény byly participanty zpracovány rychleji než scény umělé.

Torralba a Oliva (1999) ukazují, že scény lze kategorizovat podle hlavní sémantické osy poskytující informaci o přirozenosti (*naturalness*) scény, a to právě na scény přírodní (*natural*) a scény umělé (*urban*). U přírodních scén rozlišují stupeň otevřenosti (*openness*), zatímco u umělých scén stupeň vertikality (*verticalness*) (Torralba & Oliva, 1999). Loschky a Larson (2008) zjistili, že přítomnost či nepřítomnost ostrých hran ve scéně napomáhá tomu, zda bude scéna kategorizována jako scéna přírodní (nepřítomnost ostrých hran) nebo scéna umělá (přítomnost ostrých hran). Fei-Fei et al. (2007) ukazují, že participanti mají větší tendenci vnímat přírodní scény jako vnější spíše než vnitřní.

Podobné třídění jako u Torralby a Olivy (1999) nacházíme také u Olivy, Torralby, Guérin-Duguého a Heralta (1999) s tím rozdílem, že u umělých scén namísto stupně vertikality rozlišují uzavřenost scény (*expanded/enclosed*). Guérin-Dugué a Oliva (2000) na základě globální distribuce

lokálně dominantních orientací (*local dominant orientations*) dospívají také k těmto čtyřem sémantickým skupinám, do kterých třídí scény: vnitřní/vnější a otevřené/zavřené.

Ukazuje se, že významné vlastnosti celkového prostorového uspořádání (*global spatial layout properties*), jakými jsou hlavní hloubka (*dominant depth*), otevřenost (*openness*) a perspektiva (*perspective*) (Ross & Oliva, 2010), jsou lidé schopni vnímat velmi rychle a na základě nich správně kategorizovat přírodní scény (Greene & Oliva, 2006, 2009a). Z hlediska rozlišení obrazu je otevřenost nejlépe rozpoznána při vysokém rozlišení, hloubka při středním a perspektiva při nízkém rozlišení (Ross & Oliva, 2010). Tato kategorizace se dále využívá také ve výpočetních a statistických modelech založených na nízkourovňových reprezentacích (Oliva & Torralba, 2001; Torralba & Oliva, 2003), které mohou sloužit např. k rozpoznávání vnitřních scén, u kterých není jednoznačné, zda jsou rozpoznávány spíše podle globálních (podstata) či lokálních (objekty) vodítek (Quattoni & Torralba, 2009; Torralba & Sinha, 2001a).

Jsme-li vystaveni příliš mnoha podnětům ze stejné kategorie, může dojít k fenoménu adaptace (Oliva, 2013). Greene a Oliva (2010) zjistily, že při přílišné prezentaci scén s podobnými vlastnostmi může dojít skrze adaptaci ke změně kategorizace těchto scén. Ukázalo se, že jestliže si účastníci zvykli na dimenzi uzavřenosti, která jim byla prezentována, hodnotili následně dvojznačnou scénu jako otevřenou a naopak (Greene & Oliva, 2010).

Významnou roli při kategorizaci hraje také typičnost prezentované scény vzhledem k naší reprezentaci, kterou o dané kategorii máme uloženou v paměti. Caddigan, Choo, Fei-Fei a Beck (2017) zjistili u přírodních scén, že právě lepší typičnost scény přispívá k jejich jednoduššímu zaznamenání oproti méně reprezentativním položkám.

2.4 Memorabilita fotografie

Při vnímání fotografií je také důležité zohledňovat memorabilitu, která se u jednotlivých fotografií liší. Memorabilita neboli zapamatovatelnost, respektive nezapomenutelnost fotografie, je přímo vlastností fotografie jako takové (Oliva, Isola, Khosla, & Bainbridge, 2013). Jedná se o stabilní vlastnost, která nijak nepodléhá změnám napříč různými pozorovateli (Isola, Xiao, Torralba, & Oliva, 2011). Různí lidé tak mají tendenci pamatovat a zapomínat stejné fotografie (Oliva et al., 2013). Navíc je tato vlastnost stabilní i v čase (Isola, Xiao, Parikh, Torralba, & Oliva, 2014). Goetschalckx, Moors a Wagemans (2018) zjistili, že tuto stálost v čase lze pozorovat jak u krátkých intervalů v rozmezí několika desítek minut, tak i u delších časových rozestupů až v délce trvání jednoho týdne. Goetschalckx et al.

(2018) zároveň ukazují, že fotografie s vyšší memorabilitou se z paměti vytrácejí pomaleji než ty, které jsou hůře zapamatovatelné.

Otázkou tak zůstává, jaké aspekty přispívají k lepší zapamatovatelnosti fotografie. Isola, Xiao et al. (2011) zjistili, že nesémantické kategorie jako barvy, počty objektů ve fotografii nebo počet pixelů zaplněných určitým objektem nehrály významnou roli ve schopnosti předpovědět memorabilitu fotografie, zatímco sémantické kategorie memorabilitu zvyšovaly. Pojmenování objektů ve scéně a pojmenování kategorie celé scény významně vysvětlovalo danou memorabilitu (Isola, Xiao et al., 2011). K podobným závěrům dospěli i Dubey, Peterson, Khosla, Yang a Ghanem (2015) se zaměřením se na objekty, přičemž objekty s největší memorabilitou významně ovlivňovaly celkovou memorabilitu fotografie. Isola, Xiao et al. (2011) tak zjistili, že nejvíce zapamatovatelné jsou takové fotografie, na kterých se vyskytují lidé, dále pak vnitřní scény a scény s velkými objekty a nejméně zapamatovatelné pak přírodní scény s krajinou. Isola, Parikh, Torralba a Oliva (2011) rozšířili předchozí poznání o zjištění, že nejzapamatovatelnější fotografie jsou sice s lidmi, ale s těmi, kteří mají viditelné tváře. Dále se zajímali o to, jaký vliv na memorabilitu má nezvyklost či estetičnost prezentovaných scén a ukázalo se, že ani jeden aspekt nepřispívá k lepší memorabilitě scény (Isola, Parikh et al., 2011). Gygli, Grabner, Riemenschneider, Nater a Van Gool (2013) se kromě zkoumání vlivu estetičnosti a nezvyklosti zaměřili také i na zajímavost (*interestingness*) fotografie. Přestože ani ona nijak nekoreluje se skutečnou memorabilitou fotografie (Squalli-Houssaini, Duong, Gwenaëlle, & Demarty, 2018), koreluje překvapivě s tím, co si lidé myslí, že si budou pamatovat (Gygli et al., 2013).

Bylinskii, Isola, Bainbridge, Torralba a Oliva (2015) se kromě vnitřních faktorů fotografií zabývali také vnějšími faktory, které mohou ovlivňovat memorabilitu: kontext, ze kterého fotografie pochází a chování pozorovatele. Za kontext autoři považují kontext ostatních fotografií, ve kterém je vybraná fotografie prezentována. Důležitost kontextu zdůrazňují především proto, že dřívější výzkumy (např. Konkle et al., 2010a; Vogt & Magnussen, 2007) ukazují, že fotografie, které se odlišují od ostatních prezentovaných fotografií, jsou lépe pamatovány v dlouhodobé paměti. Podobně i Bylinskii, Isola, Bainbridge et al. (2015) a Bylinskii, Isola, Torralba a Oliva (2015b) svými výsledky podpořili toto zjištění, neboť nejodlišnější fotografie z hlediska kontextu byly zároveň ty nejvíce pamatované. Další vnější faktor, na který se Bylinskii, Isola, Bainbridge et al. (2015) a Bylinskii, Isola, Torralba a Oliva (2015a) zaměřili, bylo chování pozorovatele skrze oční pohyby. Autoři ukázali, že to, jakým způsobem si pozorovatel fotografií prohlíží může být u mnoha fotografií nápomocné vodítko pro následnou predikci, jak dobře si danou fotografii bude pozorovatel pamatovat. Jednoznačnost těchto predikcí je však zkomplikována v případech, kdy se veškeré důležité informace nacházejí pouze ve středu fotografie a není tak zapotřebí mnoha očních pohybů k zaznamenání a uložení zásadních informací (Bylinskii et al., 2015a; Bylinskii, Isola, Bainbridge et al., 2015).

Výzkumy využívajících očních pohybů pro zkoumání memorability tak poukazují na důležitost pozornosti pozorovatele (Mancas & Le Meur, 2013). Bainbridge (2017) se proto zaměřila na prozkoumání tří typů jevů spojených s pozorností, jmenovitě na pozornost zdola nahoru (*bottom-up attention*), pozornost shora dolů (*top-down attention*) a priming, s otázkou, zda jsou tyto jevy více s memorabilitou provázané a mohly by ji blíže vysvětlit. Skrze pět tematicky zaměřených experimentů došla k závěru, že ani jeden z těchto procesů nenarušil memorabilitu prezentovaných fotografií, a proto lze memorabilitu považovat za vnitřní charakteristiku samotné fotografie nezávislou na zmíněných procesech (Bainbridge, 2017).

Nezávislost memorability na dalších procesech byla potvrzena i výzkumy na neurologické úrovni. Bainbridge, Dilks a Oliva (2017) ukazují, že memorabilita se tak odlišuje jak od nízkourovňového vnímání, tak od procesů spjatých s individuální pamětí. Jednak u memorability dochází k zapojování odlišných mozkových center než u vnímání a paměti (Bainbridge et al., 2017), a dále pak jejich nervové signály dosahují odlišné rychlosti (Khaligh-Razavi, Bainbridge, Pantazis, & Oliva, 2016). K ranému vnímání dochází již zhruba při 100 ms u obličejů (Cauchoix, Barragan-Jason, Serre, & Barbeau, 2014) a při 200 ms u scén (Rivolta, Palermo, Schmalzl, & Williams, 2012). U memorability lze nervovou aktivitu pozorovat již při 243 ms u obličejů a při 445 ms u scén a během kódování do paměti u obličejů při 715 ms a u scén při 668 ms (Khaligh-Razavi et al., 2016). Zdá se, že procesy percepce a paměti nejsou oddělené systémy, ale lze na ně pohlížet jako na určité kontinuum (Bussey & Saksida, 2007), ve kterém právě memorabilita vytváří důležité spojení mezi vnímáním a pamětí (Bainbridge et al., 2017; Bainbridge & Rissman, 2018).

Z dalších výzkumů se mimo jiné také ukázalo, že podle rychlosti nervového signálu můžeme usuzovat na to, zda bude daný stimulus spíše zapamatován či nikoli. Mohsenzadeh, Oliva a Pantazis (2017) zjistili, že více zapamatovatelné fotografie byly rychleji zpracovávány než méně zapamatovatelné fotografie, a to již mezi 149-228 ms.

Někteří výzkumníci si položili otázku, zda by bylo možné zasahovat do fotografií, a tak u nich měnit memorabilitu. Khosla, Xiao, Isola, Torralba a Oliva (2012) navrhli možnou úpravu jednotlivých částí fotografie na základě jejich jednotlivých segmentů s předpokladem, že každá část fotografie obsahuje jiné množství informací, z nichž je memorabilita vyabstrahována. Narušením jednotlivých částí fotografie by tak mohlo dojít k úbytku informací ve fotografii obsažených, a tím i k poklesu memorability (Khosla et al., 2012). Siarohin et al. (2017) se oproti tomu uchýlili k praktickému pokusu za využití automatického editování fotografií. V procesu editace fotografie dochází k nalezení nejvhodnějšího filtru na zvýšení memorability pro danou fotografii, čímž se jim skutečně podařilo upravovat fotografie tak, aby byly co nejvíce zapamatovatelné (Siarohin et al., 2017).

2.5 Vnímání scény na základě jejích jednotlivých prvků

U scén zaměřujeme svůj pohled nejen na základě zadání a úkolu (Bengson & Luck, 2016), se kterým ke vnímání scény přistupujeme, ale také na základě prvků, které scénu tvoří (Tatler, Baddeley & Gilchrist, 2005). Na základě nich vzniká i tzv. *saliency map*, která se skládá z barevnosti, intenzity, kontrastu, orientace okrajů a mnoha dalších prvků (Henderson, 2003).

Základní informace o vnímaných stimulech přináší do receptorů na sítnici světlo (Šíkl, 2012). Světlo je elektromagnetické záření, které lze charakterizovat jako vlnu a zároveň i jako proud částic. V podobě vlny u něj rozlišujeme vlnovou délku a intenzitu. Vlnová délka je vzdálenost dvou nejbližších vrcholů vlny. Pozorovaný barevný tón odpovídá určité vlnové délce ve viditelném spektru elektromagnetického záření. Intenzitou je rozuměna amplituda vlny čili její maximální hodnota (Šíkl, 2012).

Na základě intenzity můžeme mluvit o jasu (průměr intenzity) či o kontrastu (směrodatná odchylka intenzity vzhledem k průměru) (Mante, Frazor, Bonin, Geisler, & Carandini, 2005). Jas ani kontrast nejsou na sobě závislé (Mante et al., 2005). V neposlední řadě rozlišujeme u scény jako základní vlastnost také okraje, ze kterých lze vyabstrahovat vlastnosti komplexnější, a to jednotlivé tvary nacházející se ve scéně (Laaksonen, Oja, Koskela, & Brandt, 2000).

2.5.1 Okraje a kontrast

Tatler, Baddeley et al. (2005) zjistili, že pro místa, na která se participant dávali častěji, byla zásadnější přítomnost kontrastu a okrajů, nežli jas a barevnost. Navíc se rozdíl mezi participanty s přibývajícím časem vnímání scény snižovaly (Tatler, Baddeley et al., 2005). Baddeley a Tatler (2006) v navazujícím výzkumu ukázali, že zatímco okraje ve scéně byly pro vnímání významným prvkem, kontrast byl důležitý již méně.

2.5.2 Umístění a barevnost

Aginsky a Tarr (2000) zjistili, že zatímco přítomnost a umístění objektu ve scéně jsou automaticky zpracovány při vnímání, neboť se jedná o prvky důležité pro uspořádání scény, tak povrchové vlastnosti (jako např. barva) vyžadují speciální pozornost, aby byly zakódovány do paměti. Tatler, Gilchrist a Land (2005) oproti tomu ukázali, že jak umístění, tak i identita a barva objektu byly zakódovány do paměti, ale informace o pozici byla shromažďována z víceexpozic, zatímco informace o identitě a barvě akumulovány nebyly.

V experimentech Hollingwortha a Rasmussena (2010) byla zjištěna souvislost mezi umístěním a barevností objektů. Ukázalo se, že objekt, který se přesunul z původní pozice do pozice nové, byl v krátkodobé paměti participantů provázán s barevností, kterou vykazoval v pozici původní. Nicméně barevnost v pozici původní byla zvýhodněna při pamatování oproti nové pozici nebo nové barvě (Hollingworth & Rasmussen, 2010).

Hollingworth se ve svém výzkumu dlouhodobě zaměřuje na provázanost umístění objektu ve scéně a vliv na vizuální paměť. Pomocí metody *2 alternative forced-choice*, která spočívá ve výběru mezi dvěma odpověďmi (stimuly), přičemž pouze jedna z nich je správná (jen jeden stimul je původně prezentovaný), Hollingworth (2006) ukázal, že lépe se objekty pamatují v původním kontextu nežli vytržené v izolaci a na původním místě nežli přesunuté na místo nové. Následně skrze metodu *change detection* ověřil, že participantů dosahovali lepších výkonů, pokud původní pozice objektu byla zachována a zároveň nebylo narušeno kontextuální uspořádání scény (Hollingworth, 2007). Dále při užití metody *visual search*, v níž měli participantů za úkol nalézt požadovaný objekt a rozhodnout, zda se orientace prezentovaného objektu shoduje s orientací objektu v následně prezentované scéně, jejíž byl objekt součástí, Hollingworth (2009) zjistil, že předchozí znalost scény usnadňuje následné hledání objektů v ní.

Budeme-li zvažovat samotný význam barevnosti scény, Spence, Wong, Rusan a Rastegar (2006) zjistili, že je-li scéna barevná, podstatně se zvyšuje rozpoznatelnost scény v paměti oproti stejné scéně nebarevné. Ukazuje se, že barevnost hraje důležitou roli již v raných fázích zpracování scény (Spence et al., 2006). Jiné studie zase poukazují na důležitost normálnosti barev u diagnostických vodítek ve scéně, jakými jsou např. objekty nebo jiné části scény vedoucí k jejímu rozpoznání (Oliva & Schyns, 2000; Tanaka & Presnell, 1999). Podobně i Goffaux et al. (2005) zjistili, že normálnost barevnosti scény napomáhá ke schopnosti kategorizovat přírodní scény na jejich základní úrovni (*basic-level*). Ehinger a Brockmole (2008) oproti tomu zjistili, že normálnost barevnosti scény nebyla nijak důležitá při úkolu vizuálního hledání ve scéně.

2.5.3 Velikost objektů

V žitém světě můžeme pozorovat objekty z různých úhlů pohledu. Samotná velikost objektů se však nemění. Konkle a Oliva (2007) se zabývaly otázkou, zda existuje preferovaná velikost objektu při vnímání a zjistily, že existuje korelace preferované velikosti objektu (tzv. *normative size*) na obrazovce se skutečnou velikostí objektu v realitě. Také se ukázalo, že dlouhodobá i krátkodobá paměť je citlivá vůči této normativní velikosti (Konkle & Oliva, 2007). Oproti tomu Spotorno, Malcolm a Tatler (2015)

zjišťovali vztah mezi velikostí objektu a zaměřením očních pohybů při neobvyklém umístění objektu a poukázali na to, že čím větší objekt byl, tím dříve byl pohledem fixován.

2.6 Vnímání scény na komplexnější úrovni

Každá scéna se obvykle vyznačuje smysluplným prostorovým uspořádáním. Vnímání scény tak s sebou nese nutnost zpracování velkého množství vizuálních informací, a to často ve velmi krátkém čase (Oliva, 2014). K tomuto zpracovávání informací dochází na základě předchozích zkušeností, které se scénami máme a které jsou uloženy v paměti. Určitá místa tak máme spojená s výskytem určitých objektů v nich (Hidalgo-Sotelo & Oliva, 2010). Tak můžeme například očekávat, že objekt lednice se bude spíše nacházet ve scéně kuchyně než např. na ulici nebo dokonce v lese. Nebo je pravděpodobnější, že obloha bude spíše ve vrchní části scény než v části spodní.

2.6.1 Podstata scény (gist)

Možnosti lidského vnímání se ve velké míře odvíjejí od biologických vlastností a procesů lidského organismu. Ukazuje se, že kam v rámci scén zaměřujeme svůj pohled, je ovlivněno zadáním úkolu a tudíž cílem, kterého chceme dosáhnout (Tatler, Hayhoe, Land, & Ballard, 2011) a také délkou předchozích akcí (Tatler, Baddeley, & Vincent, 2006). Jinam zaměřujeme svůj pohled v případě, že si chceme co nejlépe zapamatovat celou scénu, jinam v případě, že se zaměřujeme na vyhledávání konkrétních objektů a jinam, když chceme co nejrychleji vystihnout podstatu scény.

Rozpoznání podstaty scény znamená být schopen ve velmi krátkém časovém intervalu (méně než 100 ms) pochopit smysl scény a rozpoznat důležité informace a objekty ve scéně (Oliva, 2005). Oliva (2005) dále poukazuje na to, že bez ohledu na komplexnost scény rozpoznávání probíhá na všech třech procesuálních úrovních vizuálních informací, a to nízkourovňové (např. barva), střední úrovně (např. povrch) a vysokoúrovňové (objekty). Podle toho můžeme rozlišit buďto percepční podstatu scény (*perceptual gist*), která je založena na počátečních fázích vizuálního zpracování informací a odkazuje ke strukturálním reprezentacím, anebo konceptuální podstatu (*conceptual gist*) čili sémantický rozměr scény, který vyžaduje prezentaci o něco delší než 100ms k tomu, aby mohl být konsolidován v paměti (Oliva, 2005).

2.6.2 Přístup zaměřený na objekty a na celou scénu

Obecně můžeme rozlišit dva přístupy ke vnímání scén. Zprv se jedná o přístup zaměřený na scénu či prostor (*scene or space-centered approach*), a za druhé jde o přístup zaměřený na objekty (*object-centered approach*) (Oliva, 2014). Přístup zaměřený na objekty je hierarchický přístup postupující od zpracovávání nízkourovňových informací, přes objekty až po celkový vjem scény, zatímco přístup zaměřený na scénu předpokládá prvotní celkový vjem scény a až následné rozčlenění na jednotlivé objekty a podčásti (Zhou & Zhang, 2010). Nuthmann a Henderson (2010) zkoumali, kam účastníci směřují svůj zrak a pozornost při sledování scény a zjistili, že účastníci obvykle sledovali místo blízké středu objektů. Na základě těchto zjištění Nuthmann a Henderson (2010) tvrdí, že pozornost je přitahována nepřímo skrze korelaci s důležitými objekty ve scéně. Oliva a Torralba (2002, 2006) oproti tomu zakládají svůj přístup spíše na zaměření se na scénu jako celek a tvrdí, že k rychlému rozpoznání scény není potřeba zpracování jednotlivých objektů.

2.6.3 Centrální a periferní vidění

Další výraznou roli tvoří centrální a periferní vidění. Loschky, Nuthmann, Fortenbaugh a Levi (2017) udávají, že centrální vidění, které nám poskytuje nejostřejší obraz, dosahuje 5° okolo fixovaného místa. To je místo, kam je směřována naše zaměřená pozornost. Většina našeho zorného pole je však pokryta periferním viděním (Loschky et al., 2017), které se ukazuje být zásadní pro prostorové uspořádání a navigaci v prostoru (Ehinger, 2013). U periferního vidění je nicméně snížené prostorové rozlišení (Levi & Carney, 2009) a barevná citlivost (Newton & Eskew, 2003), přestože si to lidé mnohdy neuvědomují a domnívají se tak, že vnímají více vizuálních informací, než ve skutečnosti vidí (Lau & Rosenthal, 2011). Larson a Loschky (2009) ukázali, že pro rozpoznání podstaty scény (*gist*) je účinnější periferní vidění, zatímco centrální vidění je v tomto ohledu efektivnější, neboť pro dosažení stejné úrovně rozpoznání podstaty scény vystačí s méně informacemi.

2.6.4 Důležitost kontextu ve scéně

Scéna je ve své přirozenosti nesmírně komplexní a žádný jev ani objekt se v ní nevyskytuje izolovaně. Jednotlivé objekty jsou danými zákonitostmi provázány mezi sebou a prostředím, ve kterém je můžeme nalézt (Oliva & Torralba, 2007). Protože jsou informace o vizuálním kontextu zpracovávány již v raných fázích vnímání, jsou právě kontextuální asociace důležité pro zaměření pozornosti na důležitá místa ve scéně (Torralba, 2005).

Ukazuje se, že do paměti jsou ukládány informace o prostorové provázanosti jednotlivých objektů, spíše než jejich absolutní umístění v prostoru (Treisman & Zhang, 2006). Hollingworth (2005a) ve svém výzkumu poukázal na to, že participanti si byli schopni zapamatovat jak určité pozice konkrétních objektů, tak i jejich prostorový kontext. Vztahy mezi objekty nám také pomáhají k rychlému zorientování se ve scéně a např. při potřebě nalezení určitého objektu nám mohou pomáhat zúžit prohledávaný prostor (Torralba, Murphy & Freeman, 2010). Toho se také využívá i ve výpočetních vědách (Murphy, Torralba, & Freeman, 2004; Torralba 2003b; Torralba, Oliva, Castelhana, & Henderson, 2006). Kromě pozice objektu kontext napomáhá také k rozpoznání identity objektu a určení jeho měřítka (Torralba, 2003a). Zároveň se však ukazuje, že odhadovaná četnost výskytu objektů v určitém prostředí je participanty v průměru o 32% přeceňována, což poukazuje na využívání určitých schematických představ scén, které však nemusejí plně odpovídat realitě (Greene, 2016).

Kontext scény však nemusí nutně znamenat pouze to, co je přímo na fotografii viditelné. Kontextem se také může myslet ta část obrazu, která viditelnou část scény obklopuje, nicméně není na fotografii již zachycena (Khosla, An, Lim, & Torralba, 2014). Khosla et al. (2014) zjišťovali, zda je možné na základě prezentovaných fotografií scén odhadnout, jak daleko je konkrétní firma od dané scény nebo zda z ní lze odhadnout kriminalitu v dané oblasti, a to bez přidání vodítek. Ukázalo se, že obojí možné je, přestože počítače si v těchto úlohách vedli o něco lépe než lidé (Khosla et al., 2014). Kontext tedy může být jak uvnitř scény samotné mezi jejími jednotlivými prvky, tak i vně scény. Khosla et al. (2014) ukázali, že i tento vnější kontext obemykající zobrazenou scénu je velice bohatý na informace, které o scéně můžeme získat (např. potenciální míra kriminality v dané oblasti nebo její ekonomické klima).

2.7 Narušení vztahů ve scéně a její fragmentace

Jak jsme si dosud ukázali, scény nejsou pouhým součtem jednotlivých částí, ale všechny detaily ve scéně jsou vzájemně propojené podle určitých pravidel a dohromady tvoří komplexní celek. Tato pravidla mohou být však různými způsoby narušena. Například může dojít k neobvyklému prostorovému umístění objektu ve scéně či k různým druhům fragmentace scény. Otázkou tak zůstává, jaké změny v zapamatování si scén toto narušení způsobuje.

U scény můžeme rozlišovat dva typy struktury, sémantickou a prostorovou (Stainer, Scott-Brown, & Tatler, 2013). Při narušení sémantické struktury dochází k narušení obsahové smysluplnosti scény (např. Davenport, 2007; Davenport & Potter, 2004). To znamená, že přestože všechny části scény zůstávají zachovány, mohou být různě přeuspořádané a tím je obsahová kontinuita scény narušená.

Fragmentace a přeuspořádání scény ovlivňuje zároveň i její prostorovou strukturu. Ta se zrcadlí právě v prostorové smysluplnosti scény. Při jejím narušení může dojít např. k diskontinuitě objektů, jejichž jednotlivé části mohou být rozdělené a oddělené od sebe v prostoru, anebo přerušené oddělovacími čárami (Stainer et al., 2013).

2.7.1 Neobvyklé prostorové umístění objektu

Nejprve se blíže podíváme na narušení vztahů mezi neobvyklým umístěním objektu a jeho přirozeným kontextem scény. Spotorno et al. (2015) zjišťovali, zda při hledání objektů ve scéně upřednostňujeme nejprve prozkoumávat správně umístěné objekty, anebo objekty, které se na základě zkušeností z reálných scén vyskytovaly na místech neočekávaných. Z jejich výsledků vyplývá, že nesprávně umístěné objekty byly s větší pravděpodobností fixované již v prvotních fázích prozkoumávání scény. Pro prvotní fixaci však byla ještě podstatně větší velikost pozorovaného objektu, nežli jeho správné či nesprávné umístění. Čím větší objekt byl, tím spíše byl fixován nejdříve (Spotorno et al., 2015).

Na tento výzkum dále navázali Spotorno a Tatler (2017). Z jejich zjištění vyplývá, že dříve než na neočekávané objekty ve scéně (sémanticky nekonzistentní) se pohled participantů zaměřil na objekty (tzv. diagnostické objekty), které jsou významné pro rozpoznání scény a k jejímu porozumění. Spotorno a Tatler (2017) zároveň zjistili, že objekty, na které participanté zaměřovali svůj pohled nejdříve, se lišily také v závislosti na zadaném úkolu. Pokud je úkolem si co nejlépe zapamatovat prezentovanou scénu, je velmi důležitý sémantický kontext, zatímco při potřebě identifikovat změnu je důležitá významnost objektu (Spotorno & Tatler, 2017).

Podobně Choi, Torralba a Willsky (2012) za pomoci počítačových rozpoznávacích systémů zjišťovali, k jakým změnám dojde, pokud se naruší vztah mezi objektem a jeho přirozeným kontextem ve scéně (např. objekt se nacházel v prostoru bez spodní fyzické podpory, na nepravděpodobném místě, popřípadě jeho velikost byla vzhledem ke kontextu neobvyklá). Z jejich závěrů vyplývá, že především chybějící fyzická podpora byla důležitým vodítkem pro narušené kontextuální vztahy ve scéně (Choi et al., 2012). Již dříve za pomoci výpočetního přístupu Torralba a Sinha (2001b) vytvořili fotografie scén s narušeným kontextem a zajímali se, jak důležitý je kontext scény pro předpověď místa a velikosti objektu pro jejich systém. Toto testování však neproběhlo na lidských pozorovatelích.

2.7.2 Fragmentace scény

Dříve než se hlouběji ponoříme do poznatků o rozčlenění scény, je důležité si uvědomit, že existuje mnoho způsobů, jak lze scénu dělit a jak ji prezentovat. Například Hutchings, Czerwinski, Meyers a

Stasko (2004) ukazují, jakým způsobem je možné vytvořit jednu scénu fyzicky rozčleněnou na devíti samostatných obrazovkách (3x3) uspořádaných tak, že tvoří jednu velikou obrazovku nebo jak lze na jednu velikou obrazovku včlenit mnoho menších samostatných scén. Ve svém výzkumu využili zobrazení více samostatných scén na jedné obrazovce také VanRullen, Reddy a Fei-Fei (2005) a zjistili, že vzdálenost mezi dvěma scénami byla klíčová pro schopnost scény rozpoznat. Pokud byly dvě scény příliš blízko u sebe, byla míra rozpoznání horší nežli pro scény daleko od sebe. Důležitost blízkosti stimulů zkoumali i Tan a Czerwinski (2003). V tomto případě však využili dvou částí psaného textu prezentovaných samostatně na dvou obrazovkách. Při daných podmínkách zjistili, že zatímco fyzická oddělenost stimulů nebyla sama o sobě nikterak významná, celková vzdálenost stimulů ve vizuálním poli byla důležitá pro rozdělování pozornosti (Tan & Czerwinski, 2003). Foulsham a Kingstone (2013) se zaměřili na studování jednotlivých trajektorií očních pohybů při rozpoznávání fotografie. Nejprve prezentovali vybrané části jedné scény na jedné obrazovce v sekvencích za sebou a následně zjišťovali, jakým způsobem si budou účastníci lépe pamatovat fotografie celé. Ukázalo se, že právě tyto jednotlivé části byly pro rozpoznání celé scény klíčové (Foulsham & Kingstone, 2013).

Dalším možným způsobem fragmentace scény je rozčlenění jedné scény na jednotlivé dílky, tzv. *patches*, a jejich přeuspořádání v rámci jedné scény. Protože fragmentací scény dochází k narušení běžné struktury scény, a tím i k narušení celistvosti objektů a ploch v ní, zaměřili se Sanocki, Michelet, Sellers a Reynolds (2006) ve svém výzkumu na vztah mezi percepcí scény a její reprezentací. Ve svých experimentech zkoumali jak narušení pouze strukturální (rozdělení scény podle vertikální osy uprostřed a přehození dvou polovin stejné scény), tak i narušení strukturální a zároveň sémantické (rozdělení scény podle vertikální osy uprostřed a namíchání jedné poloviny z původní scény a jedné poloviny ze scény jiné). Dále také oba případy rozšiřovali o dělení na více jednotlivých *patches*. Sanocki et al. (2006) na základě těchto experimentů zjistili, že jednotlivé dílky tvořily funkční subsystémy a nijak nenarušovaly celkovou reprezentaci scén. Podobně i Stainer et al. (2013) potvrdili, že při fragmentaci jedné scény a při promíchání jejích jednotlivých dílků dochází k vnímání těchto dílků spíše jako jednotlivých subscén nežli jako více částí jednoho celku. To lze pozorovat i na rozložení očních pohybů při vnímání takovýchto scén, které se z prvotní centrální tendence stimulu jakožto celku rychle přesouvají k jednotlivým kvadrantům a centrální tendenci vykazují vůči každému zvlášť (Stainer et al., 2013).

Foulsham, Alan a Kingstone (2011) se ve svém výzkumu zaměřili na to, jak promíchání jednotlivých dílků ovlivňuje oční pohyby při vnímání scény s ohledem na odlišné úlohy a na odlišný počet dílků, na které je scéna rozdělena. Přestože v jednom případě zůstala scéna v původním rozložení a ve druhém byly jednotlivé dílky promíchány, dílek s dotazovaným objektem vždy zůstal na původní pozici a v obou případech se tak vyskytoval na stejném místě. Foulsham et al. (2011) zjistili,

že promíchání dílků prodlužuje reakční čas participantů, zhoršuje přesnost odpovědí a vede ke kratším očním pohybům, které jsou více rozložené po celé fotografii. Tyto tendence se však ukázaly pouze u fotografií rozdělených na větší počet dílků. Navíc se ukázalo, že při hledání předem známých objektů nepromíchanost scény významně napomáhá k rychlému nasměrování na místo, kde daný objekt očekáváme. K tomu dochází jednak na základě běžné zkušenosti s uspořádáním prostoru (nebe a strop nahoře, podlaha a zem dole) a jednak díky vztahovým vodítkům obsažených v konkrétní fotografii (hodiny spíše na zdi vedle tabule nežli na zemi blízko židlí) (Foulsham et al., 2011). K závěru, že pro rychlé rozpoznání scény je podstatný sémantický obsah fotografie přispívající k informacím o scéně jakožto celku, dospěli na základě svých experimentů i Velisavljević a Elder (2002). Ti v jednom ze svých experimentů využili také inverze v kontrastu a orientaci fotografie, což znamená, že fotografii převrátili vzhůru nohama a zvýšili stupeň kontrastu mezi barvami na fotografii. Následně i tyto fotografie prezentovali buďto s původním uspořádáním dílků, anebo s dílky promíchanými. V tomto případě nebyl nalezen rozdíl mezi fotografiemi s promíchanými a nepromíchanými dílky, což autoři přisuzují právě důležitosti sémantického uspořádání fotografie a nikoli uspořádání prostorového (Velisavljević, & Elder, 2002) a dále také vysvětlují, že VSTM je připravená na vnímání přirozeně barevných a normálně orientovaných scén, s jakými se setkáváme v žitém světě (Velisavljević & Elder, 2008b).

Velisavljević a Elder (2008a) se ve svých dalších experimentech zaměřili na to, jakou roli hraje konkrétní umístění dílků ve vnímání fragmentované scény. Zajímali se především o souvislost s centralitou a periferií fotografie. Participantům byly prezentovány fotografie, po jejichž expozici měli účastníci rozhodnout ze dvou vybraných dílcích (jeden vzatý z prezentované fotografie a druhý nikoli), který z nich byl součástí předchozí fotografie. Velisavljević a Elder (2008a) zjistili, že zatímco u normálně uspořádaných scén se výkon postupně snižoval od středu scény k jejím okrajovým částem, u scén s promíchaným uspořádáním dílků se tato tendence neobjevila. V následujícím experimentu Velisavljević a Elder (2008a) ukázali, že tento jev nelze vysvětlit pouze na základě nízkourovňových vlastností fotografie, neboť ty ve fotografii jako celku zůstávají konstantní i při promíchání jejích jednotlivých dílků. Velisavljević a Elder (2008a) tak poukázali na skutečnost, že velká část tohoto jevu souvisí spíše s faktory na straně pozorovatele nežli na straně samotné fotografie. V dalších experimentech se Velisavljević a Elder (2008a) zaměřili na jednotlivé vlastnosti fotografií a zjistili, že oproti černobílým scénám dosahovali participanté lepších výsledků u scén barevných, a to jak v případě původního uspořádání, tak i u promíchání dílků.

Varakin a Levin (2008) se ve svých experimentech zaměřili na schopnost detekovat změnu ve fragmentované scéně a zjistili, že promíchanost dílků tuto schopnost narušila. V případě pouhé inverze scén se tato schopnost nijak nezměnila, což ukazuje na provázanost tohoto efektu

s promícháním scény (Varakin & Levin, 2008). Varakin, Klemes a Porter (2013) se dále věnovali tomu, jaký vliv má fragmentace scény na subjektivní vnímání doby trvání prezentovaného stimulu a zjistili, že fragmentované scény participanti vnímali jako prezentované po kratší čas než scény s normálním uspořádáním. To autoři dávají do souvislosti s jevem, který představili Tse, Intriligator, Rivest a Cavanagh (2004) jako časově subjektivní expanzi (*time's subjective expansion*). Při tomto jevu dochází k subjektivní změně vnímání času podle nutnosti zapojení pozornosti a množství informací potřebných ke zpracování (Tse et al., 2004), což je právě při vnímání fragmentovaných scén zesíleno oproti vnímání scén normálně uspořádaných (Varakin et al., 2013). Tento jev vymizel v případě, kdy byla do obou uspořádání přidána tzv. mřížka, o které více pojednám v následující části. Při přidání mřížky do prezentované fotografie došlo k vizuálnímu oddělení jejích jednotlivých segmentů. V případě, kdy byli participanti nuceni se přímo soustředit na strukturu scény opět došlo k subjektivní změně vnímání času bez ohledu na přítomnost mřížky, což je dááno do souvislosti s potřebnou kvalitou zpracování informací (Varakin et al., 2013).

V našich experimentech byla stimulem fragmentovaná fotografie složená z dílků pocházejících ze tří různých fotografií a podobně jako většina výzkumů prezentovaných v této podkapitole i my jsme do našeho výzkumu začlenili mřížku, přičemž nás zajímalo, jakým způsobem může právě mřížka zasahovat do vnímání fragmentované fotografie.

2.7.3 Mřížka

Za mřížku (*grid*) lze označit takové horizontální a vertikální linky, které zvolenou scénu rozdělují na libovolný počet dílků. S jejím využitím se setkáváme zejména u fragmentovaných fotografií, kde buďto slouží přímo jako oddělovací prvek jednotlivých dílků, anebo pouze jejich roztržitost zdůrazňuje. Zásadní výhoda při včlenění mřížky do fragmentovaného stimulu je, že přidané linky vyrovnávají množství ukončených linek mezi fragmentovanou a nefragmentovanou fotografií (Stainer et al., 2013).

Koenderink, Richards a Doorn (2012) rozlišují mřížky, které část scény překrývají a jiné, které scénu nepřekrývají. V případě překrývajících mřížek jsou čáry vedeny přímo přes část fotografie. Tím dochází k zakrytí určitých informací a vizuálnímu systému je tak zabráněno zpracování těchto informací v samotných počátcích. V takovém případě dochází k jevu tzv. amodální okluze (*amodal occlusion*) (Koenderink et al., 2012). Přestože se ve skutečnosti část informací nedostává do našeho vizuálního systému, mozek si tyto informace dotváří na základě sousedících informací a spojuje je v jeden celek. Máme tak dojem, že i části zakryté mřížkou vnímáme, přestože tyto informace nejsou nesené na smyslové úrovni (Calabi, 2013). Jako příklad lze uvést psa, který stojí za plaňkovým plotem.

Ačkoli ve skutečnosti z něho vnímáme pouze ty části těla, které nejsou zakryté plaňkami, máme dojem, jako bychom viděli celého psa a skrze amodální okluzi si nepercipované části nevědomě dotváříme v mysli. Tento případ mřížek použili například Varakin a Levin (2008). Ve druhém případě mřížka žádnou část fotografie nepřekrývá. Fotografie je rozdělena na jednotlivé dílky, které jsou od sebe mřížkou pouze odděleny. Tím nedochází ke ztrátě žádné informace, která má být percipována. Konkrétní případ si můžeme představit jako obrázek složený z puzzle, který rozdělíme na jednotlivé dílky a necháme mezi nimi prázdné mezery (Koenderink et al., 2012). Takový design využili ve svém výzkumu například Stainer et al. (2013).

V případě nepromíchaných scén působí mřížka jako narušující faktor, neboť se jedná o další možný objekt, který lze v rámci stimulu sledovat a který tak s ostatními prvky na fotografii bojuje o pozornost (Foulsham et al., 2011). Foulsham et al. (2011) také zjistili, že v případě mřížky, která dělí scénu na více dílků, je toto narušení větší nežli v případě mřížky dělící scénu na méně dílků. Zároveň v případě překrývající mřížky u nepromíchaných scén dochází při zrakovém vyhledávání (*search task*) ke změně ve způsobu prohledávání scény, a to na postupné procházení jednotlivých dílků a v případě rozdělenosti objektů mřížkou i k větší míře chybovosti (Bi, Bae, & Balakrishnan, 2010). Také Varakin a Levin (2008) zkoumali vliv překrývající mřížky na schopnost zaznamenat změnu (*change detection task*) a přítomnost mřížky působila větší chybovost i zde. Varakin a Levin (2008) poukazují na důležitost struktury ve scéně při tomto typu úloh. Varakin et al. (2013) ve svém experimentu primárně nekladli důležitost na strukturu scény, ale na změny v subjektivním vnímání času a v tomto případě nebyl rozdíl u stimulů s mřížkou a stimulů bez mřížky. Mřížka se tak zdá být důležitým prvkem při vnímání ve spojitosti se strukturou scény.

2.8 Extrapolace

Extrapolace je proces domýšlení, při kterém na základě skutečně vnímaných informací dochází i k odhadu takových informací, které sahají za hranice pozorovaných částí. Mozek si tak ze známých prvků fotografie může vytvořit dojem známosti i u těch částí scén, které nebyly přímo součástí stimulu, a to i přesto, že tyto části nikdy neprošly senzorkým systémem daného člověka.

Proces extrapolace se v algoritmickém provedení používá pro odhadnutí a doplnění chybějících částí fotografií či videí (Seiler & Kaup, 2008; Seiler, Meisinger, & Kaup, 2008). V takovém případě se jedná o signální extrapolaci, která je součástí vizuální komunikace a která se tak snaží o nápravu chyb, ke kterým mohlo dojít v například v průběhu přenosu dat (Seiler et al., 2008; Seiler & Kaup, 2008). Signální extrapolace tak probíhá na základě informací ze sousedních částí chybějících vnitřních dílků (Lakshman, Ndjiki-Nya, Köppel, Doshkov, & Wiegand, 2009). Jiné výzkumy (např. Wang, Lai, Liang,

Martin, & Hu, 2014; Zhang, Xiao, Hays, & Tan, 2013) se oproti snahám o extrapolaci chybějících vnitřních částí scény pokusili o extrapolaci z původní velikosti fotografie na fotografii panoramatickou. Originální scénu tak naopak použili jako známý vnitřní díl k chybějícímu okolí.

V případě našeho výzkumu nás zajímalo, zda při využití stimulu složeného ze tří různých scén také dochází k extrapolaci, a to konkrétně u neviděných vnitřních částí fotografií a zda jsou participantů schopni rozlišit skutečně percipované části fotografií od těch neviděných.

Empirická část

3) Cíl výzkumu a výzkumné otázky

Každý den jsme vystaveni obrovskému množství zrakových podnětů (scén). Není však stále jasné, s jakou přesností si tyto scény ukládáme a vybavujeme z paměti a také, zda si do paměti ukládáme scény celé, nebo jen její dílčí části. Konkrétně není jasné, do jaké míry jsou zapamatovány detaily a do jaké míry dochází k extrapolaci neviděného obsahu z prezentovaných dílčích částí.

K prozkoumání této otázky jsme využili fragmentovaných scén (fotografie rozčleněné na 4x4 dílků), které se vždy skládaly ze tří různých kategorií (*indoor*, *natural* a *man-made*) a pokaždé se lišily počtem prezentovaných dílků v jednotlivých kategoriích (3, 5 a 8 z 16 celkem) u jedné fragmentované scény.

Naše hlavní výzkumná otázka je, jestli existuje vztah mezi počtem prezentovaných dílků a schopností rozpoznat, které dílky byly prezentované a které nebyly (přestože i tyto dílky pocházejí z prezentovaných fotografií, než byly fragmentovány). Především nás zajímá citlivost paměti (*memory sensitivity*, d') a *bias* vycházející z teorie detekce signálu (*signal detection theory*, SDT).

Očekáváme, že výkon paměti bude ovlivněn dvěma mechanismy: extrapolací a seskupováním (*grouping*). Očekáváme, že může dojít k jednomu z těchto tří možných výsledků:

(1) Extrapolace je silnější než seskupování.

Čím méně dílků z jedné kategorie je prezentováno, tím vyšší je d' a nižší *bias*. Čím méně dílků z jedné kategorie je prezentováno, tím nižší je referenční skupina vizuálních informací v paměti, se kterou je potřeba porovnat dotazovaný dílek z prezentované fragmentované scény.

(2) Mezi extrapolací a seskupováním není žádný vztah.

Neexistuje žádný lineární trend ani pro d' ani pro *bias*. Počet prezentovaných dílků z jedné kategorie není nijak spojen s naší schopností určit, zda dotazovaný patch byl, nebo nebyl součástí prezentované fragmentované scény.

(3) Seskupování je silnější než extrapolace.

Čím méně dílků z jedné kategorie je prezentováno, tím nižší je d' a vyšší bias. Čím více dílků z jedné kategorie je prezentováno, tím dříve a více zaměřena je naše pozornost na tyto dílky. Schopnost určit, zda dotazovaný patch byl, nebo nebyl součástí prezentované fragmentované scény je úzce spojené s extrapolací.

4) Experiment 1

Cílem Experimentu 1 bylo ověřit výše představené výzkumné otázky. Sběr dat proběhl v únoru a v březnu 2018 a předcházela mu pilotní výzkum, který proběhl v únoru 2018. Studie byla schválena Etickou komisí Psychologického ústavu AV ČR.

4.1 Metody

4.1.1 Participanti

Exp.1 se zúčastnilo celkem 30 participantů ve věkovém rozmezí 18-48 let (průměr = 23,2, SD = 5,7, medián = 22). Dva participanté byli studenti SŠ, ostatní byli studenti VŠ. Účast na experimentu byla odměněna, participanté mohli volit mezi splněním podmínky k atestaci do předmětu (9 participantů), nebo finanční odměnou ve výši 150,- Kč (21 participantů).

4.1.2 Design a stimuly

Při tvorbě podnětových stimulů bylo potřeba zvážit různé prvky fragmentovaných fotografií. Jmenovitě se jednalo o druhy kategorií, podkategorií a konkrétních fotografií, ze kterých budou stimuly vytvářeny. Dále bylo třeba vybrat vhodnou kompozici čili počet kategorií, ze kterých bude stimul složen a v jakém poměru. Následně bylo potřeba zvážit, jakým způsobem kontrolovat vnitřní uspořádání fragmentované fotografie, k čemuž nám posloužila tzv. maska. A v neposlední řadě jsme do podnětových fotografií pro Exp.1 včlenili mřížku.

Databáze SUN397

Všechny fotografie, které byly využity na tvorbu stimulů, byly vybrány z databáze SUN397. Jedná se o databázi vycházející z databáze SUN (= *Scene UNderstanding*) a je volně dostupná na těchto oficiálních stránkách: <https://vision.princeton.edu/projects/2010/SUN/>. Zatímco databáze SUN obsahuje 908 kategorií a 131 072 fotografií scén větších než 200 x 200 pixelů (Xiao, Ehinger, Hays, Torralba, & Oliva, 2016), SUN397 sestává z 397 kategorií a v každé z nich je k dispozici nejméně 100 tematických fotografií (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010).

Kategorie a podkategorie

Xiao et al. (2013) definují kategorii scén jako funkční a sémantickou samostatnou jednotku, která většinou člověku slouží k vykonávání nějaké činnosti. Běžně se v rámci scén rozlišují kategorie

vnitřní (*indoor*) a vnější (*outdoor*) a dále pak přírodní (*natural*) a umělé (*man-made*). Kombinací těchto charakteristik nám vznikají čtyři možné kategorie: vnitřní-přírodní, vnitřní-umělé, vnější-přírodní a vnější-umělé. Ke každé z nich si můžeme uvést konkrétní příklad podkategorie: vnitřní-přírodní (jeskyně), vnitřní-umělé (kuchyň), vnější-přírodní (kaňon) a vnější-umělé (chata). Z důvodu nedostatku dalších podkategorií v kategorii vnitřní-přírodní, jsme byli nuceni pracovat pouze se třemi ostatními kategoriemi, které jsme označili následovně: vnitřní-umělé (*indoor* = I), vnější-přírodní (*natural* = N) a vnější-umělé (*man-made* = M). Tyto kombinace jsou znázorněny na Obrázku II.

TYPY SCÉN	vnitřní (<i>indoor</i>)	vnější (<i>outdoor</i>)
přírodní (<i>natural</i>)	X (jeskyně)	<i>natural</i> (kaňon)
umělá (<i>man-made</i>)	<i>indoor</i> (kuchyň)	<i>man-made</i> (chata)

Obrázek II. Typy scén kategorií a podkategorií

Celkem byly vybrány tři základní kategorie (I, N, M) a v každé z nich 10 podkategorií (např. kuchyň, kaňon, chata) po 40 fotografiích každá. Celkem se jednalo o $3 \times 10 \times 40 = 1200$ základních fotografií, které sloužily jako materiál pro výrobu prezentovaných podnětových stimulů, fragmentovaných fotografií.

Všechny fotografie prošly důkladným výběrem. Nejprve byly z databáze SUN397 ponechány pouze fotografie s minimálními rozměry 600x600 pixelů. Následně byly vybrány kategorie s dostatečným počtem fotografií, tedy alespoň se 40 fotografiemi. Dále byly tyto kategorie ručně procházeny a ponechány pouze ty, na nichž se nevyskytovaly rušivě výrazné prvky, jako například lidské či zvířecí obličejové kvůli jejich odlišné memorabilitě (Isola, Xiao et al., 2011), výrazné nápisy, různé defekty při pořizování scén (rozmazanost fotografií, čas pořízení fotografie, aj.). Také se dbalo na to, aby se jednalo o scény reálné, čímž bylo vyřazeno i několik počítačových návrhů scén a také,

aby zároveň nebyly fotografie příliš jednotvárné, a tudíž informace v jednotlivých částech scény splývající (např. příliš velká část jednoduté oblohy, trávy, vody).

Kompozice

Při tvorbě fragmentovaných fotografií byly vždy náhodně vybrány celkem tři fotografie, jedna fotografie z každé ze tří základních kategorií (I, N, M).



Obrázek III. Příklad konkrétních fotografií se zařazením do podkategorie a kategorie

Každá fotografie byla oříznuta na 512x512 pixelů a automaticky rozčleněna na $4 \times 4 = 16$ dílků, tzv. *patchů* (každý o velikosti 128x128 pixelů). Na základě určitých kritérií vnitřního uspořádání, která jsou detailněji rozepsána v další sekci, byla následně vytvořena fragmentovaná fotografie.

Rozmýšlení samotné kompozice fragmentované fotografie prošlo vlastním vývojem. Jednalo se o ustálení počtu kategorií, ze kterých bude výsledná fotografie složena a v jakém poměru jednotlivých dílků to bude. Než se ukázalo, že na základě SUN397 bude možné pracovat pouze se třemi základními kategoriemi (I, N, M), uvažovali jsme o dalších různých možnostech, např. s počtem dvou nebo čtyř kategorií, či případně pěti podkategoriemi. Po protřídění databáze nám přišlo nejvhodnější použít tyto tři základní kategorie také pro tvorbu samotných fragmentovaných fotografií. Z celkových 16 dílků jsme uvažovali konkrétní poměr dílků mezi těmito třemi kategoriemi a zvolili jsme poměr 8 ku 5 ku 3, aby se případné rozdíly mezi počtem zobrazovaných dílků na zapamatování jednotlivých patchů mohli zanést na lineární křivku.



Obrázek IV. Konečná podoba fragmentované fotografie vytvořené z předchozích příkladů na Obrázku III.

Maska

Každý dílek má v rámci fragmentované fotografie své možné fixní místo, odpovídající konkrétní pozici v původní fotografii. To znamená, že dílek v levém horním rohu fotografie se může ve vytvořené fragmentované fotografii nacházet opět pouze v levém horním rohu fragmentované fotografie. Nicméně na toto místo může být vybrán dílek jen z jedné základní kategorie ze tří možných celkem. Toto vnitřní uspořádání fotografie označujeme jako *maska*. Při výběru vhodných vnitřních uspořádání pro výsledné fotografie bylo důležité stanovení kritérií, které by zamezily výskytu nežádoucích jevů ve stimulech.

Celkem bylo automaticky vygenerováno přes 200 tisíc možných vnitřních kombinací při rozdělení 16 dílků na kategorie po 3, 5 a 8. Důležité bylo, aby se v souboru podnětů nevyskytoval žádný s velkou souvislou plochou nebo souvislým znakem. Proto jsme zvolili, aby plocha viděné oblasti z fotografie se pohybovala v rozmezí 4-5 dílků (25-50 %). Dále jsme vyloučili, aby se na podnětech vyskytovaly souvislé kvadranty z jedné fotografie. Také jsme stanovili, aby se v souboru podnětů nevyskytl žádný takový dílek, u něž by byl alespoň jeden již viděný dílek, který by sousedil se čtyřmi neviděnými dílky. Ex post jsme vyloučili takové podněty, u kterých se vyskytla souvislá řada či sloupec ze stejné fotografie. Pro tvorbu podnětových fotografií bylo celkem vybráno 60 *masek*.

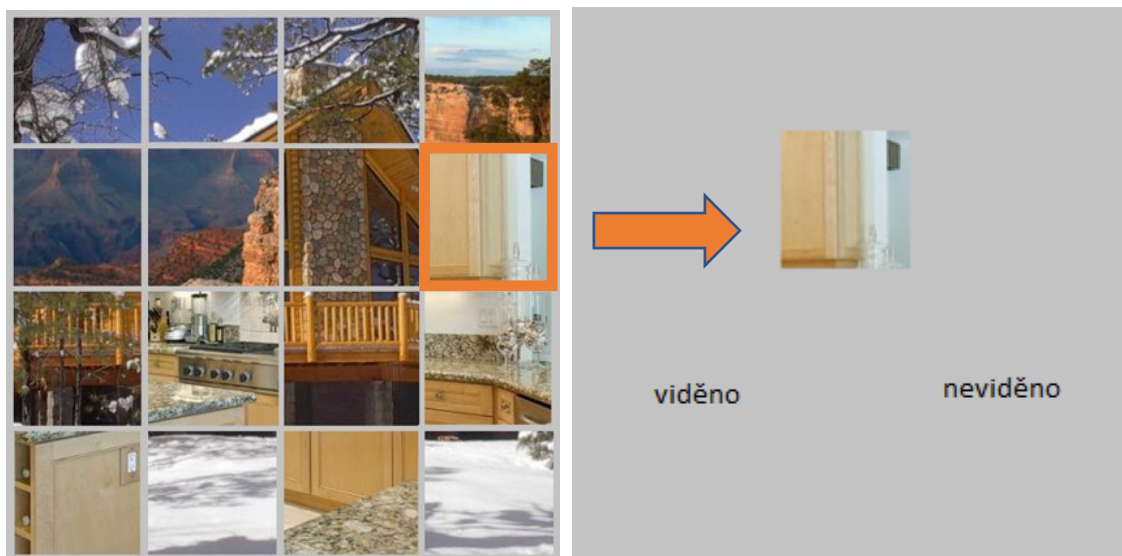
Mřížka

V našem experimentu jsme zvolili překrývající typ mřížky o velikosti 6 pixelů, jejíž linky jsou vedeny přímo přes část vizuálního podnětu.

Patch

Každý jeden dílek ze šestnácti, na které byla každá fotografie rozčlenitelná, měl rozměry 128x128 pixelů. Pravděpodobnost, že jeden konkrétní dílek bude následně i součástí fragmentované fotografie, byla 1 ku 3, neboť fragmentovaná fotografie vždy sestávala právě ze tří fotografií.

Jako patch označujeme jeden dílek z těchto tří fotografií, na který se po prezentaci fragmentované fotografie dotazujeme, zda byl, nebo nebyl součástí této prezentované fragmentované fotografie. Takový patch buďto skutečně byl součástí fragmentované fotografie (v takovém případě jej označujeme jako „viděný“, *seen* = S), anebo její součástí nebyl („neviděný“, *unseen* = U). Ve všech případech však každý patch pocházel z jedné z původních tří fotografií.



Obrázek V. Příklad vybraného patche z fragmentované fotografie, na který byl následně participant dotazován. Oranžové ohraničení a šipka jsou pouze názornou pomůckou a nebyly součástí experimentálních stimulů

PsychoPy

PsychoPy neboli *Psychology software in Python*, je open-source software volně dostupný na těchto oficiálních stránkách: <http://www.psychopy.org>. Tento program se skládá ze dvou rozhraní, ve kterých

je možné vytvořit experiment. V prvním případě se jedná o kódovací editor, tzv. „Coder“, který je určen pro programování. Druhá složka je grafické rozhraní, tzv. „Builder“, který obsahuje jednotlivé komponenty, ze kterých lze postupně sestavit celý experiment (Peirce, 2018).

PsychoPy byl vytvořen Jonem Peircem v roce 2002, a to pouze jako knihovna pro programovací jazyk Python. Postupně za přispění mnoha nadšenců se rozrostl v technicky nenáročný software pro programování experimentů (Peirce, 2018).

Pc

Experiment byl proveden na počítačích s operačním systémem Windows 7 a velikostí monitoru 23,6 palců s rozlišením 1920x1080.

4.1.3 Procedura

Průběh sběru dat

Experiment proběhl v LABELS. Participantů bylo zvaných přes databázi ORSEE a každý participant se přihlásil na konkrétní termín a čas experimentu. Na jednom testovacím termínu mohlo být přítomno maximálně 5 participantů (dané počtem počítačů v laboratoři). Délka jednoho testovacího termínu byla 60 minut, přičemž část v PsychoPy trvala participantům přibližně 35-45 minut.

Participantů přicházeli do laboratoře najednou a sami si mohli vybrat, ke kterému z pěti počítačů se posadí. V ojedinělých případech pozdních příchodů byl participant instruován ve vedlejší místnosti a následně posazen k volnému počítači v laboratoři. Všechny počítače byly od sebe oddělené zástěnou, takže se participantů v průběhu experimentu navzájem nerušili a neviděli na obrazovky ostatních. Na počátku byl participantům předložen k pročetí a k podepsání Informovaný souhlas LABELS. Následně byly každému rozdány vytisknuté instrukce s obrázkovým doprovodem k průběhu experimentu v PsychoPy a experimentátor jednotlivé kroky hromadně vysvětlil. Participantů se před podepsáním Informovaného souhlasu i poté při seznámení se s instrukcemi k průběhu experimentu mohli dotazovat na vše, co jim nebylo jasné. Dále byli také upozorněni na to, že když se na obrazovce nachází fixační kříž, mohou si sami podle uvážení na chvíli odpočinout a pak zase pokračovat dál stiskem mezerníku. Průběžné přestávky byly doporučovány jak před zahájením experimentu, tak ještě dvakrát či třikrát v průběhu experimentu. Při pilotní studii nebylo zjištěno, že by tato oznámení v průběhu experimentu vyrušovala. Většina participantů po experimentu sdílela, že si přestávky dělali a že se po nich dokázali lépe soustředit. Participantů byli také instruováni, že důležitější je přesnost odpovědi, než její rychlost a že na odpověď mají kolik času chtějí. Tím, že si participantů spouštěli každý

trial sami mezerníkem, mohl každý participant postupovat svým preferovaným tempem. Všichni participanté absolvovali experiment v PsychoPy v rozmezí 35-45 minut a po jeho skončení se odebrali do vedlejší místnosti za experimentátorem. Zde byli dotázáni, zda vše proběhlo v pořádku a nenastal nějaký technický či jiný problém. Na závěr participanté zvolili jednu z nabízených možností odměny (finanční, potvrzení o účasti ke splnění atestace z předmětu) a bylo jim poděkováno za účast.

Struktura experimentu v PsychoPy

Celou strukturu experimentu v PsychoPy můžeme rozdělit na následující části: instrukce, zácvičná část, experimentální část a závěr.

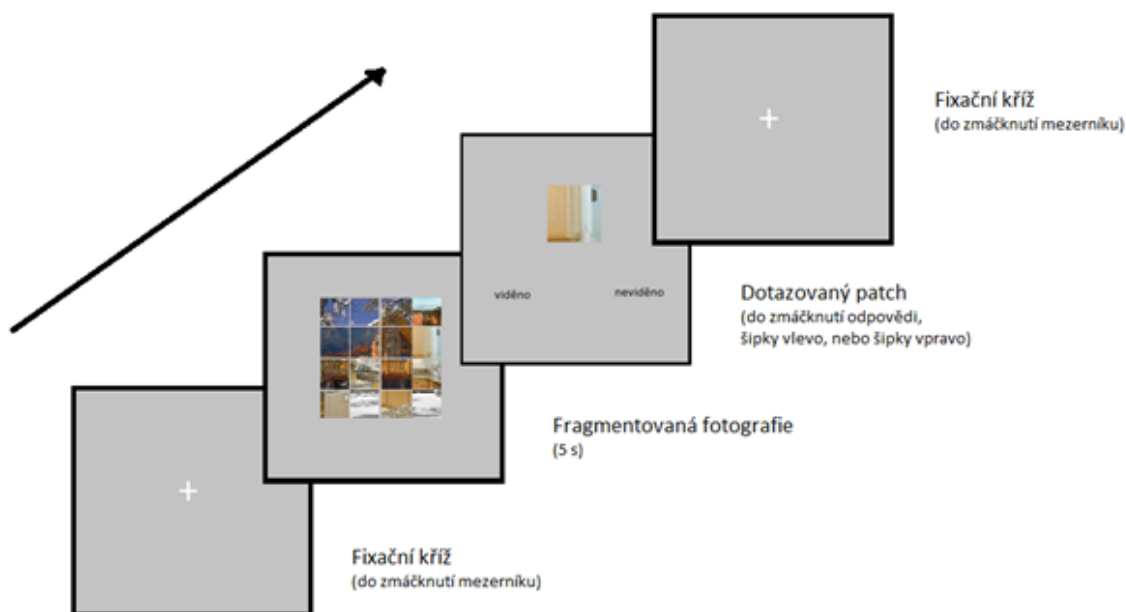
Úvodní instrukce byla následující:

„Pozorně si, prosím, prohlédněte následující fotografie a rozhodněte, zda jste danou část fotografie již viděli, nebo neviděli. Jednotlivé úlohy si spouštíte stiskem mezerníku. Odpovídejte pomocí šipek:

šipka vlevo značí zobrazený podnět

šipka vpravo nový podnět“

Po zmáčknutí mezerníku se participant přesunul do části zácvičné, o čemž byl informován na obrazovce. Ta sloužila k tomu, aby se participant seznámil s průběhem experimentu a ověřil si, zda porozuměl instrukci. Po celou dobu experimentu bylo vše prezentováno na šedém pozadí (RGB: 128, 128, 128) a všechny texty byly napsány bílou barvou (RGB: 255, 255, 255). Zácvičná část sestávala z 24 opakování, tzv. *trialů*. Každý trial měl následující podobu. Nejprve se na obrazovce objevil bílý fixační kříž a zůstal tam až do chvíle, dokud participant nezmáčkne mezerník. Po zmáčknutí mezerníku fixační kříž zmizel a byl nahrazen prezentací zácvičné fragmentované fotografie uprostřed obrazovky. Prezentace fragmentované fotografie trvala 5 s. Po uplynutí této doby fragmentovaná fotografie zmizela a byla nahrazena prezentací jednoho patche, který příslušel k prezentované fragmentované fotografii. Patch byl zobrazen uprostřed ve vrchní části obrazovky a nikoli na své pozici v původní fotografii, ze které pocházel. Tento patch byl prezentován s dotazem, zda se jedná o dílek, který byl prezentován v rámci fragmentované fotografie, anebo o dílek, který prezentován nebyl. Patch byl prezentován po neomezený časový limit, dokud nebyla zmáčkuta jedna ze dvou šipek. Pomocí levé („*viděno*“) a pravé („*neviděno*“) šipky na klávesnici participant odpovídal, zda patch byl, nebo nebyl, součástí předchozí fragmentované fotografie. Po zaznamenané odpovědi se na obrazovce opět objevil fixační kříž a zůstal zobrazen, dokud participant opět nezmáčkne mezerník. Po zmáčknutí mezerníku následovalo spuštění dalšího trialu.



Obrázek VI. Struktura experimentu v PsychoPy

Po části zácvičné byl participant informován, že následuje experimentální část. Její forma byla stejná, jako v části zácvičné. Namísto 24 trialů však tato část obsahovala 216 trialů. Ve všem ostatním se nelišila od předchozí části. Všem participantům byly prezentovány stejné podněty podle stejného protokolu.

V závěru byl participant na obrazovce informován o konci experimentu a bylo mu poděkováno za účast.

Sekvence a doba zobrazení fragmentované fotografie a patche

V počátečních úvahách a prvotních pokusech předcházejících pilotní studii (listopad 2016 – listopad 2017) jsme rozmýšleli jednotlivé prvky experimentu. Kromě již výše popsané kompozice fragmentované fotografie a včlenění mřížky, jsme také důkladně rozvažovali sekvenci mezi zobrazením fragmentované fotografie a patche a po jakou dobu budou zobrazeni. Ukázalo se, že sekvence vytvořená z pěti po sobě jdoucích fragmentovaných fotografií následovaných otázkou na jeden patch se jeví jako příliš náročná. Stejně tak náročná se jevila i původně zamýšlená doba zobrazení u fragmentované fotografie 3 s a u patche 5 s. Z tohoto důvodu jsme nakonec zvolili nejjednodušší sekvenci zobrazení, a to prezentaci fragmentované fotografie rovnou následovanou prezentací patche s touto fotografií souvisejícím. Zároveň jsme se rozhodli navýšit i dobu zobrazení, a to u fragmentované fotografie na 5 s a u patche jsme nechali časový limit neomezený.

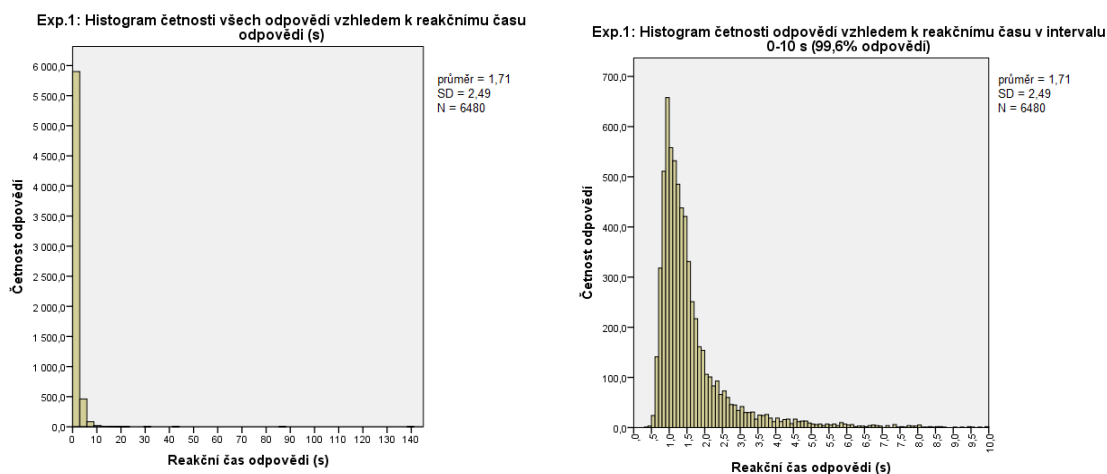
4.2 Výsledky

K analýze dat bylo užito programů Microsoft Office Excel, SPSS a JASP. Pro všechny analýzy byla zvolená hladina významnosti $p = ,05$.

4.2.1 Reakční čas

Pro ověření toho, zda participanti v průběhu celého experimentu skutečně věnovali pozornost experimentální úloze, bylo předem zvoleno kritérium pro vyřazení těch participantů, jejichž reakční čas by byl méně než 0,5 s u více než 10 % trialů. Žádný z participantů toto kritérium nenaplnil, a tak žádná data nemusela být z následné analýzy vyřazena. Pod 0,5 s se celkem vyskytly pouze čtyři odpovědi z celkových 6480 (30 participantů x 216 odpovědí každý = 6480).

Celkový průměrný reakční čas (RT) byl 1,71 s (SD = 2,49). U viděných patchů byl průměrný RT 1,73 s (SD = 3,27) a u neviděných patchů byl průměrný RT 1,68 s (SD = 1,30). RT ani v jednom z případů nemá normální rozdělení, pro viděné patche $D(3240) = 0,36$, $p < ,001$, a pro neviděné $D(3240) = 0,22$, $p < ,001$. Mezi viděnými a neviděnými dílky nebyla zjištěna rozdílnost rozptylů RT $F(1, 6478) = ,004$, $p = ,947$.



Obrázek VII. Exp.1: Histogramy četnosti odpovědí vzhledem k reakčním časům

4.2.2 Signal detection theory (SDT)

Hit, false alarm, miss, correct rejection

Signal detection theory (SDT), teorie detekce signálu, je metoda vhodná pro případy, ve kterých je potřeba rozlišit mezi dvěma stimuly (Stanislaw & Todorov, 1999). V našem případě se jedná o rozlišení dotazovaných patchů, které buďto byly součástí prezentované fragmentované fotografie (S) a patchů, které prezentované nebyly (U). Zároveň participant u těchto dvou možných stimulů volí jednu z odpovědí ano/ne (*yes/no task*), respektive viděno/neviděno v našem výzkumu, podle toho, zda se domnívají, že dotazovaný patch byl, nebo nebyl součástí prezentované fragmentované fotografie. Při zkombinování všech možností (stimulus: S x U, odpověď: viděno x neviděno) dostáváme čtyři možné charakteristiky: *hit*, *false alarm* (FA), *miss* a *correct rejection* (CR).

Hit udává počet případů, při kterých stimulus byl přítomen (patch S) a zároveň participant správně odpověděl, že tento patch viděl (Stanislaw & Todorov, 1999). Tento případ označujeme jako *true positive* (TP).

False alarm (FA) udává počet případů, při kterých stimulus nebyl přítomen (patch U) a zároveň participant nesprávně odpověděl, že tento patch viděl (Stanislaw & Todorov, 1999). Tento případ označujeme jako *false positive* (FP).

Miss udává počet případů, při kterých stimulus byl přítomen (patch S) a zároveň participant nesprávně odpověděl, že tento patch neviděl (Stanislaw & Todorov, 1999). Tento případ označujeme jako *false negative* (FN).

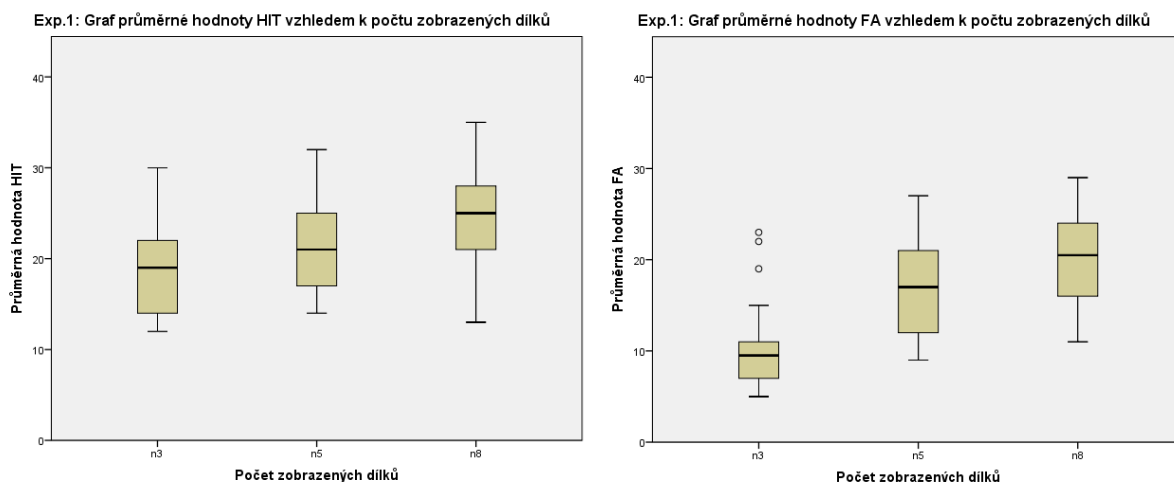
Correct rejection (CR) udává počet případů, při kterých stimulus nebyl přítomen (patch U) a zároveň participant správně odpověděl, že tento patch neviděl (Stanislaw & Todorov, 1999). Tento případ označujeme jako *true negative* (TN).

		Byl patch skutečně na předchozí fotografii?	
		ano	ne
Odpověď participanta:	viděno	HIT <i>(true positive)</i>	FA <i>(false positive)</i>
	neviděno	MISS <i>(false negative)</i>	CR <i>(true negative)</i>

Obrázek VIII. Confusion matrix: rozložení HIT, FA, MISS a CR

V Exp.1 se celková hodnota HIT pohybovala v rozmezí 41-97 (průměr = 64,87, SD = 13,86), přičemž u patchů pocházejících ze skupiny tří dílků HIT(n3) byla celková hodnota v rozmezí 12-30 (průměr = 18,83, SD = 5,05), u patchů pocházejících ze skupiny pěti dílků HIT(n5) byla celková hodnota v rozmezí 14-32 (průměr = 21,67, SD = 4,98) a u patchů pocházejících ze skupiny osmi dílků HIT(n8) se celková hodnota pohybovala mezi 13-35 (průměr = 24,37, SD = 4,97). Tyto hodnoty ukazují, že z čím většího počtu prezentovaných dílků (S) dotazovaný patch pocházel, tím více participantů správně odpovídali, že jej viděli.

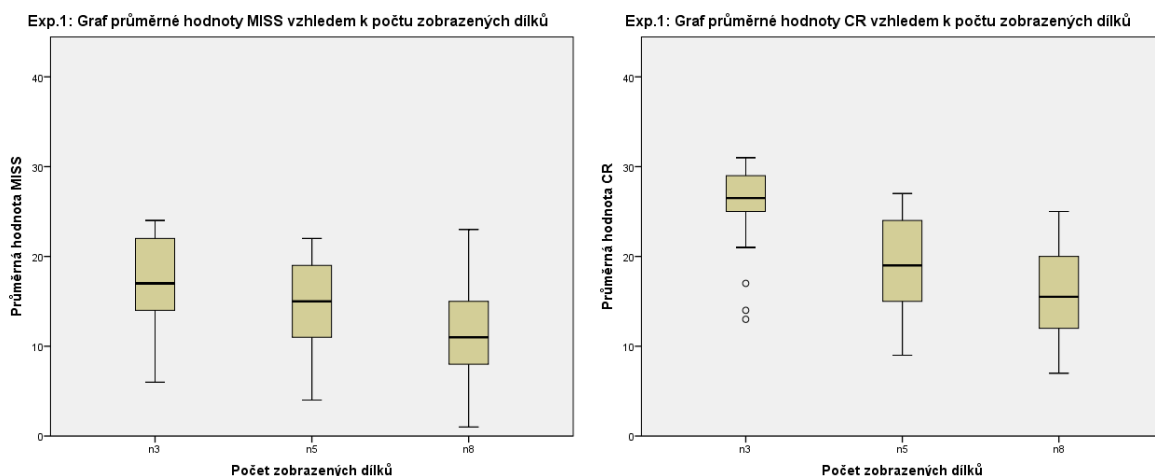
Celková hodnota FA se pohybovala v rozmezí 26-77 (průměr = 47,37, SD = 13,95). Pro FA(n3) se celková hodnota pohybovala v rozmezí 5-23 (průměr = 10,17, SD = 4,62), pro FA(n5) v rozmezí 9-27 (průměr = 17,00, SD = 5,32) a pro FA(n8) v rozmezí 11-29 (průměr = 20,20, SD = 5,42). Podobně jako u HIT i u FA můžeme pozorovat stoupající tendenci s vyšším počtem dílků, což znamená, že z čím většího počtu pocházel patch, který prezentován nebyl (U), tím spíše docházelo k nesprávným odpovědím, že jej participantů viděli. Výsledky pro HIT a FA jsou graficky znázorněny na Obrázku IX.



Obrázek IX. Exp.1: Grafy průměrných hodnot HIT a FA vzhledem k počtu zobrazených dílků

Celková hodnota MISS se pohybovala v rozmezí 11-67 (průměr = 43,13, SD = 13,86). Pro MISS(n3) byla celková hodnota v rozmezí 6-24 (průměr = 17,17, SD = 5,05), pro MISS(n5) v rozmezí 4-22 (průměr = 14,33, SD = 4,98) a pro MISS(n8) v rozmezí 1-23 (průměr = 11,63, SD = 4,99). Zde můžeme pozorovat tendenci klesající čili z čím méně dílků pocházel patch, který prezentován byl (S), tím více participantů nesprávně odpovídali, že jej viděli.

Celková hodnota CR se pohybovala v rozmezí 31-82 (průměr = 60,63, SD = 13,95). Pro CR(n3) se celková hodnota pohybovala v rozmezí 13-31 (průměr = 25,83, SD = 4,62), pro CR(n5) v rozmezí 9-27 (průměr = 19,00, SD = 5,32) a pro CR(n8) v rozmezí 7-25 (průměr = 15,80, SD = 5,42). V případě CR z čím méně dílků pocházel patch, který nebyl prezentován (U), tím více participantů správně odpovídali, že jej neviděli. Výsledky pro MISS a CR jsou graficky znázorněny na Obrázku X.



Obrázek X. Exp.1: Grafy průměrných hodnot MISS a CR vzhledem k počtu zobrazených dílků

Pravděpodobnost a přesnost odpovědí

Díky průměrnému skóru pro hit, FA, miss a CR máme nyní hrubou představu o tom, jak si účastníci v experimentu vedli. Tyto hodnoty nám sami o sobě nicméně neříkají nic o pravděpodobnostech odpovědí na určitý typ podnětu nebo jak přesní účastníci ve svém odpovídání jsou. K tomu nám slouží další charakteristiky, které lze vyvodit a vypočítat z právě jichž získaných základních informací o počtu hit, FA, miss a CR. Patří mezi ně např. *hit rate (sensitivity)*, *false alarm rate* a *accuracy*.

Hit rate je pravděpodobnost kladných odpovědí v případě přítomného stimulu (Stanislaw & Todorov, 1999). *Hit rate* je také někdy označováno jako citlivost (*sensitivity*) nebo znovuvybavení (*recall*). Jedná se o poměr případů, ve kterých stimulus byl přítomen a účastník správně odpověděl, že jej viděl, vůči všem případům, ve kterých byl stimulus přítomen (Wickens, 2002). S ohledem na pomocnou tabulku na Obrázku VIII., jej můžeme vypočítat jako:

$$H = \frac{TP}{TP+FN} \quad \text{nebo také jako} \quad H = \frac{HIT}{HIT+MISS}$$

False alarm rate je pravděpodobnost kladných odpovědí při nepřítomnosti stimulu (Stanislaw & Todorov, 1999). Jedná se o poměr případů, ve kterých stimulus přítomen nebyl a účastník nesprávně odpověděl, že jej viděl, vůči všem případům, ve kterých stimulus nebyl přítomen (Wickens, 2002). Výpočet lze zapsat jako:

$$F = \frac{FP}{FP+TN} \quad \text{nebo také jako} \quad F = \frac{FA}{FA+CR}$$

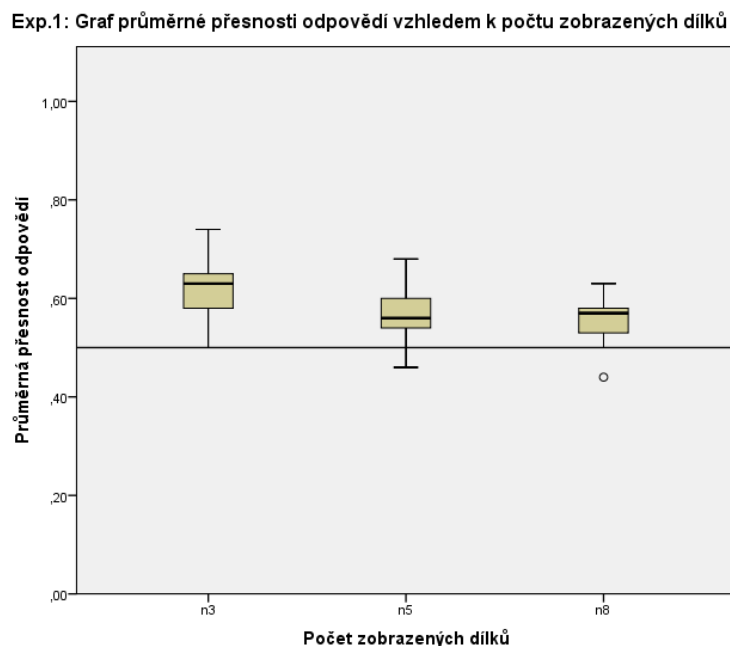
Accuracy (proportion of correct responses) neboli přesnost odpovědi je poměr správných odpovědí vůči všem odpovědím (Wickens, 2002). To lze vypočítat jako:

$$p(c)^* = \frac{(TP+TN)}{TP+TN+FP+FN} \quad \text{nebo také jako} \quad p(c)^* = \frac{HIT+CR}{HIT+CR+FA+MISS}$$

Celkový *hit rate (H)* za všechny účastníky v Exp.1 se pohyboval v rozmezí 0,38-0,90 (průměr = 0,60, SD = 0,13). To znamená, že v průměru byla pravděpodobnost kladné odpovědi v případě přítomného stimulu 0,60 (60 %). $H(n3)$, tedy v případě patche ze tří dílků, se pohyboval v rozmezí 0,33-0,83 (průměr = 0,52, SD = 0,14), $H(n5)$ se pohyboval v rozmezí 0,39-0,89 (průměr = 0,60, SD = 0,14) a $H(n8)$ v rozmezí 0,36-0,97 (průměr = 0,68, SD = 0,14). Z vypočtených hodnot tak můžeme pozorovat stoupající tendenci s narůstajícím počtem dílků.

Celkový *false alarm rate* (F) za všechny participanty v Exp.1 se pohyboval v rozmezí 0,24-0,71 (průměr = 0,44, SD = 0,13). To znamená, že v průměru byla pravděpodobnost kladné odpovědi při nepřítomnosti stimulu 0,44 (44 %). $F(n3)$, tedy v případě patche ze tří dílků, se pohyboval v rozmezí 0,14-0,64 (průměr = 0,28, SD = 0,13), $F(n5)$ se pohyboval v rozmezí 0,25-0,75 (průměr = 0,47, SD = 0,15) a $F(n8)$ v rozmezí 0,31-0,81 (průměr = 0,56, SD = 0,15). Z vypočtených hodnot tak můžeme pozorovat stoupající tendenci s narůstajícím počtem dílků.

Celková *accuracy* ($p(c)^*$) za všechny participanty v Exp.1 se pohybovala v rozmezí 0,51-0,64 (průměr = 0,58, SD = 0,03). To znamená, že přesnost odpovědí byla v průměru 0,58 (58 %). U $p(c)^*$ pro 3 dílky se přesnost odpovědi pohybovala v rozmezí 0,50-0,74 (průměr = 0,62, SD = 0,05), pro 5 dílků v rozmezí 0,46-0,68 (průměr = 0,57, SD = 0,05) a pro 8 dílků v rozmezí 0,44-0,63 (průměr = 0,56, SD = 0,04). Z vypočtených hodnot tak můžeme pozorovat (viz Obrázek XI.) klesající tendenci přesnosti odpovědí s narůstajícím počtem dílků.



Obrázek XI. Exp.1: Graf průměrné přesnosti odpovědi vzhledem k počtu zobrazených dílků

Bias a d'

Další významné charakteristiky popisující výkon participantů v úloze, jsou *bias* a d' . Obě charakteristiky využívají výpočtu na základě standardizovaného skóru z -skór z *hit rate* [$z(H)$] a z *false alarm rate* [$z(F)$]. Tím jsou původní pravděpodobnosti odpovědí transformovány na rozdělení s průměrem = 0 a SD = 1.

Bias je tendence odpovídat jedním, nebo druhým směrem čili tendence odpovídat více ano, nebo ne (Stanislaw & Todorov, 1999). K určení vychýlenosti odpovědí určitým směrem používáme vzorec:

$$c = -0,5[z(H) + z(F)]$$

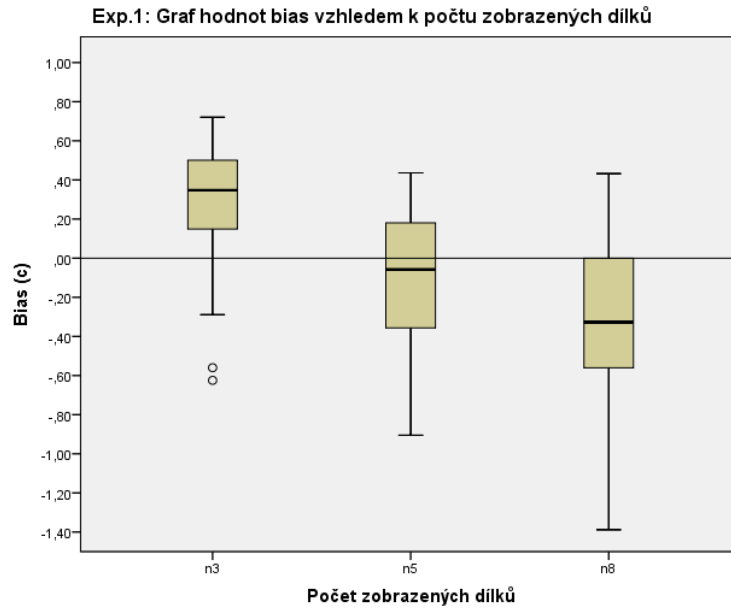
Bias se také označuje jako *c(criterion)* čili kritérium. V případě, že $c < 0$, znamená to, že hodnota (*H*) a hodnota (*F*) je vysoká a participantů mají tudíž tendenci odpovídat *ano* jak v případě, kdy podnět prezentován byl, tak i v případě, kdy prezentován nebyl. Toto kritérium označujeme jako *liberální*. V případě, že $c > 0$, znamená to, že hodnota *H* a hodnota *F* je nízká a participantů mají tudíž tendenci odpovídat *ne* jak v případě, kdy podnět prezentován byl, tak i v případě, kdy prezentován nebyl. Toto kritérium označujeme jako *konzervativní*. *Bias* tedy může nabývat jak pozitivních, tak negativních hodnot, přičemž extrémní případy nastávají ve chvíli, kdy obě hodnoty *H* i *F* jsou obě vysoké, nebo obě nízké. Za velmi extrémní hodnoty můžeme považovat $c = \pm 2,33$ (Wickens, 2002).

Index citlivosti (d') je způsob, jakým lze měřit, jak moc člověk určuje hodnotu správně a lze jej vypočítat jako:

$$d' = z(H) - z(F)$$

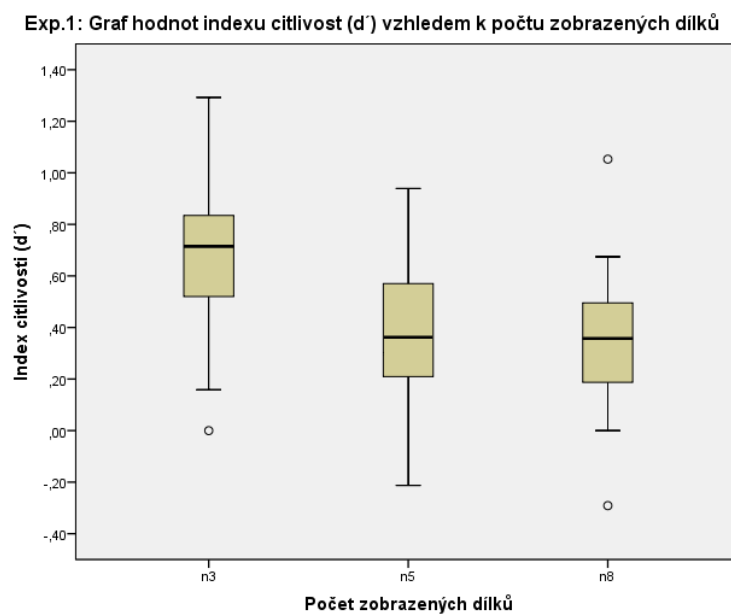
Výsledné hodnoty d' se pohybují zhruba od 0 až po 4,65, což bývá považováno za velmi extrémní hodnotu (Wickens, 2002).

V našem případě se celkový *bias* v Exp.1 pohyboval v rozmezí -0,92 až 0,51 (průměr = -0,06, SD = 0,34). To znamená, že celková tendence participantů odpovídat určitým způsobem byla velmi blízká nule, a tudíž vyvážená, s lehkým náklonem k *liberálnosti* v odpovědích čili k tendenci odpovídat *ano*. V případě tří dílků *c*(n3) bylo v rozmezí -0,62 až 0,72 (průměr = 0,27, SD = 0,34), v případě pěti dílků *c*(n5) bylo v rozmezí -0,91 až 0,44 (průměr = -0,11, SD = 0,36) a v případě osmi dílků *c*(n8) se pohybovalo v rozmezí -1,39 až 0,43 (průměr = -0,34, SD = 0,41). Z výsledných hodnot můžeme pozorovat, že z čím více dílků dotazovaný patch pocházel, tím *liberálnější* participantů ve svých odpovědích byli a tím spíše odpovídali *ano*, že podnět viděli (viz Obrázek XII.).



Obrázek XII. Exp.1: Graf průměrných hodnot bias vzhledem k počtu zobrazených dílků

V případě d' se celková hodnota pohybovala v rozmezí 0,07-0,74 (průměr = 0,44, SD = 0,16). V případě tří dílků $d'(n3)$ bylo v rozmezí 0-1,29 (průměr = 0,67, SD = 0,29), v případě pěti dílků $d'(n5)$ bylo v rozmezí -0,21 až 0,94 (průměr = 0,35, SD = 0,28) a v případě osmi dílků $d'(n8)$ se pohybovalo v rozmezí -0,29 až 1,05 (průměr = 0,34, SD = 0,25). Z výsledných hodnot lze usuzovat, že účastníci lépe určovali patche v případě tří dílků nežli v případě pěti či osmi dílků (viz Obrázek XII.).



Obrázek XIII. Exp.1: Graf hodnot d' vzhledem k počtu zobrazených dílků

4.2.3 Regresní analýza

Zajímá nás pravděpodobnost, s jakou můžeme předpovídat správnost odpovědi (závisle proměnná) s ohledem na počet zobrazených dílků (nezávisle proměnná). Protože se v obou případech jedná o kategorická či ordinální data, je nutno k analýze použít binární logistickou regresi. V takovém případě dochází ke srovnání nulového modelu bez jakýchkoli fixních efektů s modely alternativními, které by co nejlépe predikovaly pravděpodobnost odpovědi na základě pozorovaných dat (Field, 2009).

Logistický regresní model se ukázal být statisticky významný, $\chi^2(2) = 20,931$, $p < ,001$, přičemž velikost vztahu mezi oběma proměnnými je 0,4% (Nagelkerke $R^2 = 0,004$). Správnost odpovědi v modelu je 58,1 %. Jak je vidět v Tabulce I. u počtu tří zobrazených dílků (n3) byla 1,6krát vyšší pravděpodobnost správné odpovědi nežli odpovědi nesprávné. V případě pěti dílků (n5) byla šance na správnou odpověď nižší ($\text{Exp}(B) = 0,79$) a ještě o něco nižší v případě osmi dílků (n8) ($\text{Exp}(B) = 0,77$).

Tabulka I. Exp.1: koeficienty logistické regrese

Koeficienty

	B	Std. chyba	Poměr šancí [Exp(B)]	p	95% Interval spolehlivosti	
					Dolní mez	Horní mez
N3 (referenční skupina)	0,491	0,044	1,634	< ,001	0,404	0,578
N5	-0,230	0,062	0,794	< ,001	-0,352	-0,109
N8	-0,259	0,062	0,772	< ,001	-0,380	-0,137

Poznámka. Správnost odpovědi „1“ kódováno jako skupina 1.

4.3 Diskuze

V Exp.1 nás zajímalo, jaká je role detailu a extrapolace při zapamatování si fragmentovaných scén. K zodpovězení této otázky jsme zvolili charakteristiky SDT, u kterých nás zajímalo, jak přesně a správně dokáží participanti rozlišit části fragmentovaných fotografií, které právě viděli, anebo neviděli.

Zatímco průměrná hodnota měla u HIT a FA s větším počtem prezentovaných dílků stoupající tendenci, u MISS a CR měla s větším počtem prezentovaných dílků tendenci klesající. Tento jev nám

pomáhá pochopit zjištěný průměrný bias, který nás informuje o tom, že s větším počtem dílků jsou participantů ve svých odpovědích liberálnější a více odpovídají, že podnět viděli, a to i v případě, kdy jej skutečně viděli, ale i kdy jej neviděli.

Zároveň s tím můžeme pozorovat, že u přesnosti odpovědí se objevuje trend, že čím méně dílků z kategorie dotazovaného patche bylo zobrazeno, tím přesnější a správnější odpovědi participantů byli. To jsme mohli pozorovat jak u hodnot pro přesnost (accuracy), tak i u hodnot d' a zároveň i z modelu logistické regrese. Nicméně tato vyšší přesnost a správnost se vyskytovala především u skupiny 3 dílků, v případě dílků 5 a 8 se hodnota lišila jen velmi málo.

To znamená, že při malém počtu 3 dílků jsou lidé ve svých odpovědích o něco více konzervativní, ale zároveň jsou v těchto odpovědích přesnější, méně chybují, takže dokáží lépe rozlišit, zda dotazovaný patch skutečně viděli, nebo neviděli na prezentované fragmentované fotografii. Lze tedy předpokládat, že u tří dílků hraje významnější roli detail nežli extrapolace.

V případě pěti dílků dochází ke zhoršení přesnosti v odpovědích a zároveň se vyrovnává tendence odpovídat určitým směrem k nezájatosti. Je zajímavé, že tato tendence se pro kategorii osmi dílků následně stáčí k zaujatosti odpovídat ano bez ohledu na předchozí přítomnost stimulu, ale přesnost se již rapidně nezhoršuje. To poukazuje na možnost, že zde by mohlo docházet spíše k extrapolaci čili při větším počtu dílků participant může nabývat dojmu, že viděl i ty podněty, které ve skutečnosti neviděl. Nutno také podotknout, že přesnost u osmi dílků ($p(c) = 0,56$) je také mnohem blíže samotné hranici hádání (0,5) než například přesnost u tří dílků ($p(c) = 0,62$).

5) Experiment 2

Cílem Experimentu 2 bylo zjistit, zda se objevuje odlišný vztah mezi počtem prezentovaných dílků a schopností rozpoznat, které dílky byly prezentované a které nebyly, v případě, pokud fragmentované fotografie obsahují či neobsahují mřížku. Sběr dat proběhl v červnu 2018.

5.1 Metody

5.1.1 Participanti

Exp.2 se zúčastnilo celkem 29 participantů, 1 byl dodatečně vyřazen kvůli předchozí účasti v Exp.1. Zbýlých 28 participantů se pohybovalo v rozmezí 20-37 let (průměr = 24, SD = 3,1, medián = 24). Dva účastníci v době testování již nebyly studenty VŠ, zbylí participantů studovali VŠ. Účast na experimentu byla odměněna stejně jako u Exp.1, splnění podmínky k atestaci do předmětu zvolili 3 participantů, finanční odměnu ve výši 150,- Kč zvolilo 25 participantů.

5.1.2 Design a stimuly

Pro Exp.2 bylo zvoleno stejných stimulů fragmentovaných fotografií jako v případě Exp.1, pouze do nich nebyla začleněna mřížka. Výsledný stimul měl podobu jako Obrázek IV.

5.1.3 Procedura

Procedura Exp.2 byla shodná s procedurou v Exp.1.

5.2 Výsledky

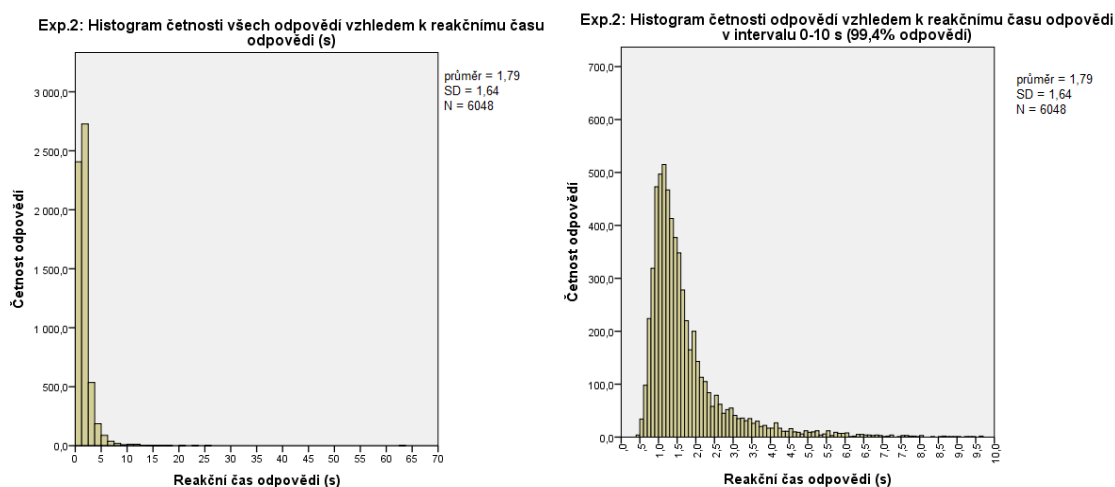
K analýze dat bylo užito programů Microsoft Office Excel, SPSS a JASP. Pro všechny analýzy byla zvolená hladina významnosti $p = ,05$.

5.2.1 Reakční čas

Pro ověření toho, zda participantů v průběhu celého experimentu skutečně věnovali pozornost experimentální úloze, bylo předem zvoleno kritérium pro vyřazení těch participantů, jejichž reakční čas by byl méně než 0,5 s u více než 10 % trialů. Žádný z participantů toto kritérium nenaplnil, a tak žádná

data nemusela být z následné analýzy vyřazena. Pod 0,5 s se celkem vyskytly pouze čtyři odpovědi z celkových 6048 (28 participantů x 216 odpovědí každý = 6048).

Celkový průměrný reakční čas (RT) byl 1,79 s (SD = 1,64). U viděných patchů byl průměrný RT 1,77 s (SD = 1,83) a u neviděných patchů byl průměrný RT 1,82 s (SD = 1,43). RT ani v jednom z případů nemá normální rozdělení, pro viděné patche $D(3024) = 0,26, p < ,001$, a pro neviděné $D(3024) = 0,21, p < ,001$. Rozdílnost rozptylů mezi viděnými a neviděnými dílky se za pomoci Levenova testu ukazuje jako statisticky významná $F(1, 6046) = 6,62, p = ,01$.

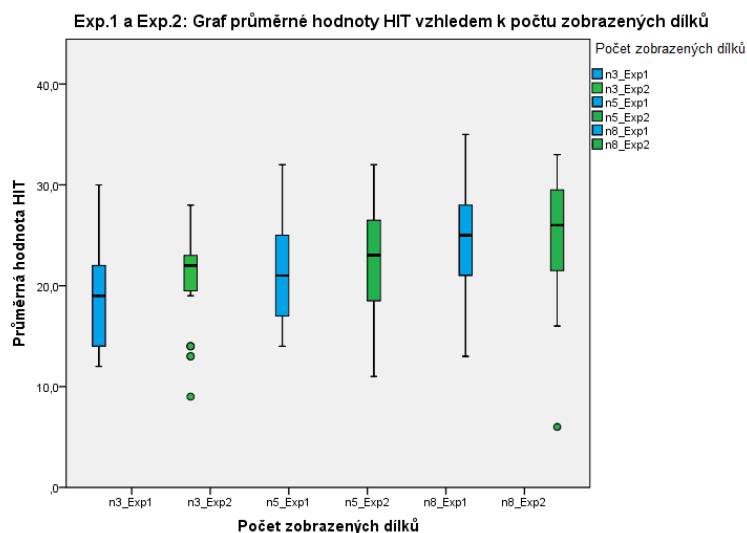


Obrázek XIV. Exp.2: Histogramy četnosti všech odpovědí vzhledem k reakčnímu času

5.2.2 Signal detection theory

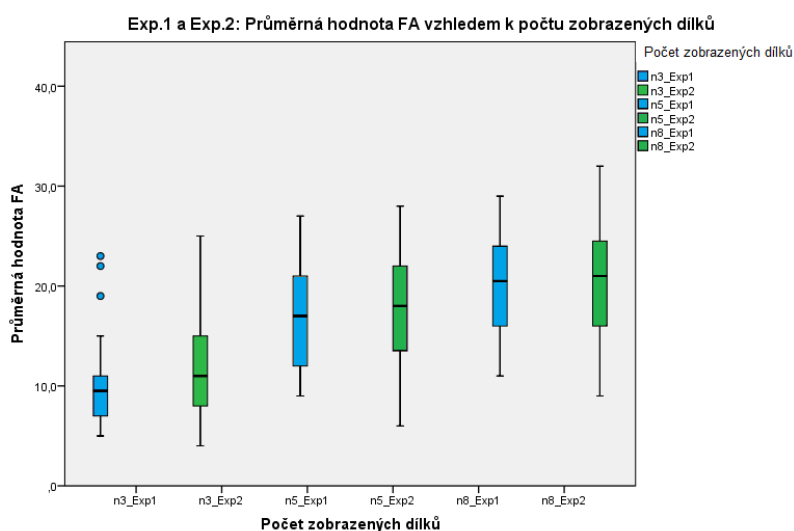
Hit, false alarm, miss, correct rejection

V Exp.2 se celková hodnota HIT pohybovala v rozmezí 39-88 (průměr = 68,04, SD = 14,40), přičemž u patchů pocházejících ze skupiny tří dílků HIT(n3) byla celková hodnota v rozmezí 9-28 (průměr = 20,39, SD = 4,51), u patchů pocházejících ze skupiny pěti dílků HIT(n5) byla celková hodnota v rozmezí 11-32 (průměr = 22,57, SD = 5,80) a u patchů pocházejících ze skupiny osmi dílků HIT(n8) se celková hodnota pohybovala mezi 6-33 (průměr = 25,07, SD = 6,39). Tyto hodnoty ukazují, že z čím většího počtu prezentovaných dílků (S) dotazovaný patch pocházel, tím více participantů správně odpovídali, že jej viděli. Stejný trend ve velmi blízkých hodnotách jsme měli možnost pozorovat i v Exp.1, jejich srovnání je graficky znázorněno na Obrázku XV.



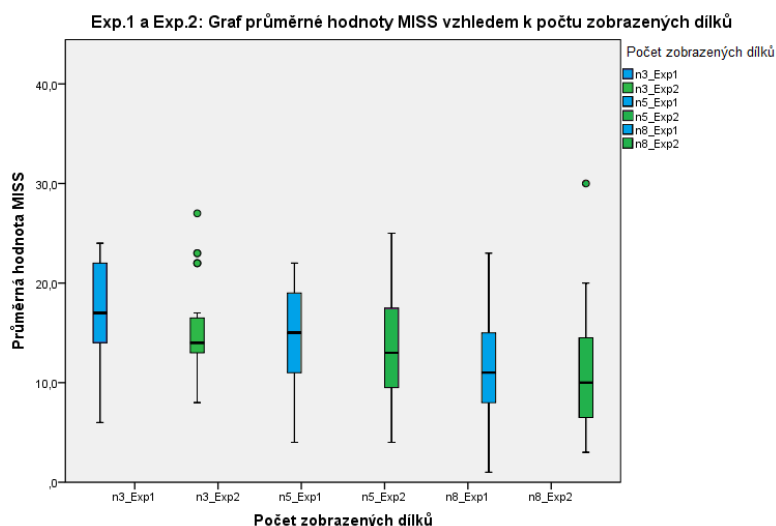
Obrázek XV. Exp.1 a Exp.2: Graf průměrné hodnoty HIT vzhledem k počtu zobrazených dílků

Celková hodnota FA se pohybovala v rozmezí 22-72 (průměr = 49,43, SD = 14,33). Pro FA(n3) se celková hodnota pohybovala v rozmezí 4-25 (průměr = 11,68, SD = 5,06), pro FA(n5) v rozmezí 6-28 (průměr = 17,61, SD = 5,67) a pro FA(n8) v rozmezí 9-32 (průměr = 20,14, SD = 5,78). Z dat můžeme pozorovat stoupající tendenci s vyšším počtem dílků, což znamená, že z čím většího počtu pocházel patch, který prezentován nebyl (U), tím spíše docházelo k nesprávným odpovědím, že jej účastníci viděli.



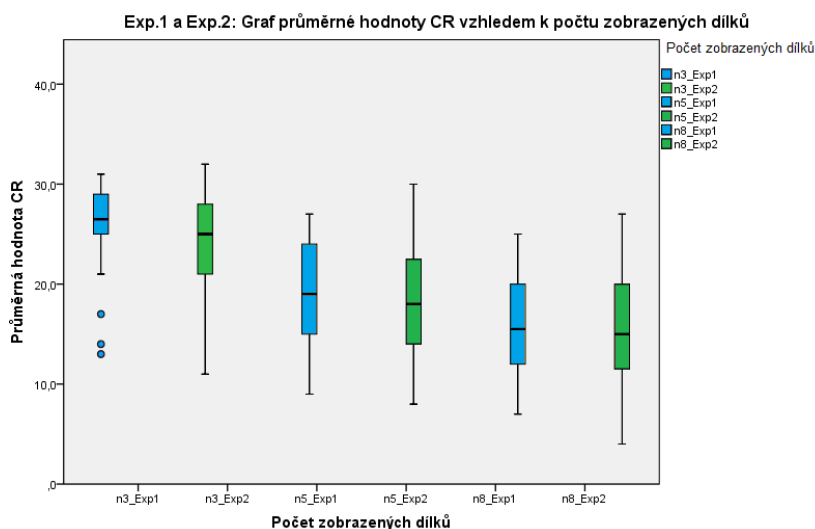
Obrázek XVI. Exp.1 a Exp.2: Graf průměrné hodnoty FA vzhledem k počtu zobrazených dílků

Celková hodnota MISS se pohybovala v rozmezí 20-69 (průměr = 39,96, SD = 14,40). Pro MISS(n3) byla celková hodnota v rozmezí 8-27 (průměr = 15,61, SD = 4,51), pro MISS(n5) v rozmezí 4-25 (průměr = 13,43, SD = 5,80) a pro MISS(n8) v rozmezí 3-30 (průměr = 10,93, SD = 6,39). Zde můžeme pozorovat tendenci klesající čili z čím méně dílků pocházel patch, který prezentován byl (S), tím více participantů nesprávně odpovídali, že jej viděli.



Obrázek XVII. Exp.1 a Exp.2: Graf průměrné hodnoty MISS vzhledem k počtu zobrazených dílků

Celková hodnota CR se pohybovala v rozmezí 36-86 (průměr = 58,57, SD = 14,33). Pro CR(n3) se celková hodnota pohybovala v rozmezí 11-32 (průměr = 24,32, SD = 5,06), pro CR(n5) v rozmezí 8-30 (průměr = 18,39, SD = 5,67) a pro CR(n8) v rozmezí 4-27 (průměr = 15,86, SD = 5,78). V případě CR z čím méně dílků pocházel patch, který nebyl prezentován (U), tím více participantů správně odpovídali, že jej neviděli.



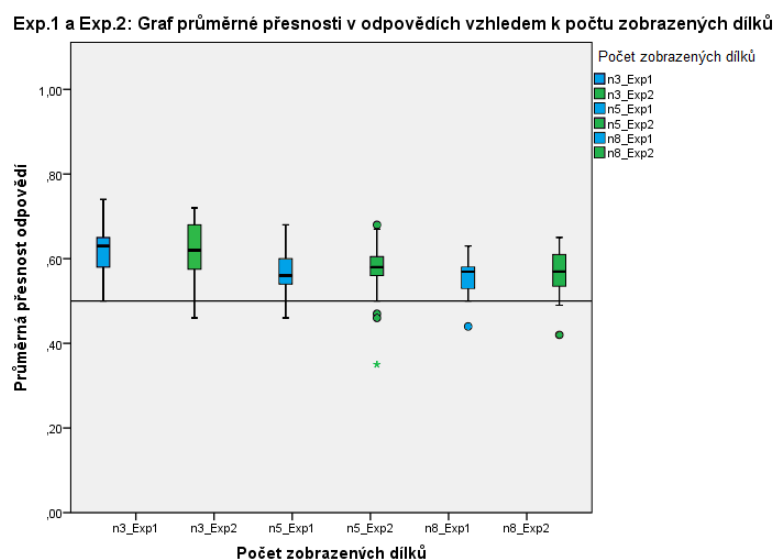
Obrázek XVIII. Exp.1 a Exp.2: Graf průměrné hodnoty CR vzhledem k počtu zobrazených dílků

Pravděpodobnost a přesnost odpovědi

Celkový *hit rate* (H) za všechny participanty v Exp.2 se pohyboval v rozmezí 0,36-0,81 (průměr = 0,63, $SD = 0,13$). To znamená, že v průměru byla pravděpodobnost kladné odpovědi v případě přítomného stimulu 0,63 (63 %). $H(n3)$, tedy v případě patche ze tří dílků, se pohyboval v rozmezí 0,25-0,78 (průměr = 0,57, $SD = 0,13$), $H(n5)$ se pohyboval v rozmezí 0,31-0,89 (průměr = 0,63, $SD = 0,16$) a $H(n8)$ v rozmezí 0,17-0,92 (průměr = 0,70, $SD = 0,18$). Z vypočtených hodnot tak můžeme pozorovat stoupající tendenci s narůstajícím počtem dílků.

Celkový *false alarm rate* (F) za všechny participanty v Exp.2 se pohyboval v rozmezí 0,20-0,67 (průměr = 0,46, $SD = 0,13$). To znamená, že v průměru byla pravděpodobnost kladné odpovědi při nepřítomnosti stimulu 0,46 (46 %). $F(n3)$, tedy v případě patche ze tří dílků, se pohyboval v rozmezí 0,11-0,69 (průměr = 0,32, $SD = 0,14$), $F(n5)$ se pohyboval v rozmezí 0,17-0,78 (průměr = 0,49, $SD = 0,16$) a $F(n8)$ v rozmezí 0,25-0,89 (průměr = 0,56, $SD = 0,16$). Z vypočtených hodnot tak můžeme pozorovat stoupající tendenci s narůstajícím počtem dílků.

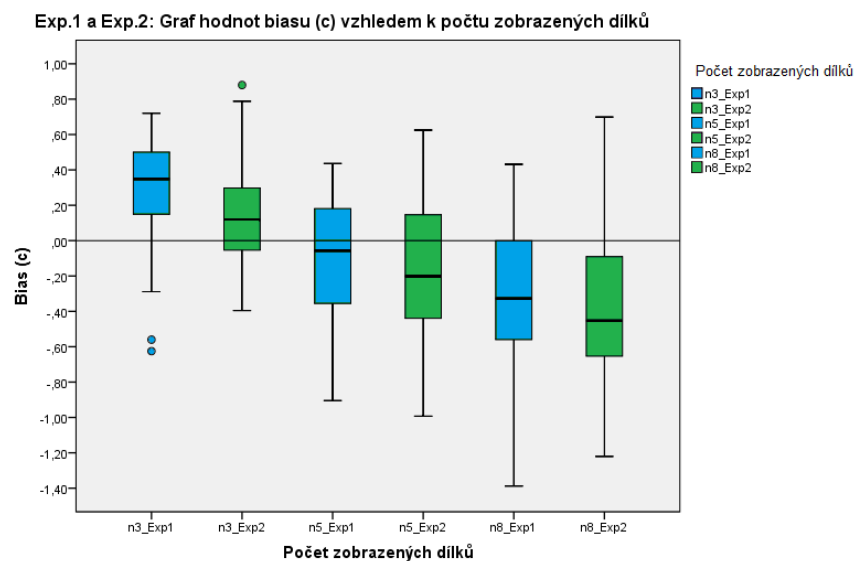
Celková *accuracy* ($p(c)^*$) za všechny participanty v Exp.2 se pohybovala v rozmezí 0,41-0,67 (průměr = 0,59, $SD = 0,05$). To znamená, že přesnost odpovědí byla v průměru 0,59 (59 %). U ($p(c)^*$) pro 3 dílky se přesnost odpovědi pohybovala v rozmezí 0,46-0,72 (průměr = 0,62, $SD = 0,07$), pro 5 dílků v rozmezí 0,35-0,68 (průměr = 0,57, $SD = 0,07$) a pro 8 dílků v rozmezí 0,42-0,65 (průměr = 0,57, $SD = 0,06$). Grafické znázornění na Obrázku XIX. nám umožňuje optické srovnání přesnosti odpovědí pro různý počet dílků u Exp.1 a Exp.2.



Obrázek XIX. Exp.1 a Exp.2: Graf průměrné přesnosti v odpovědích vzhledem k počtu zobrazených dílků

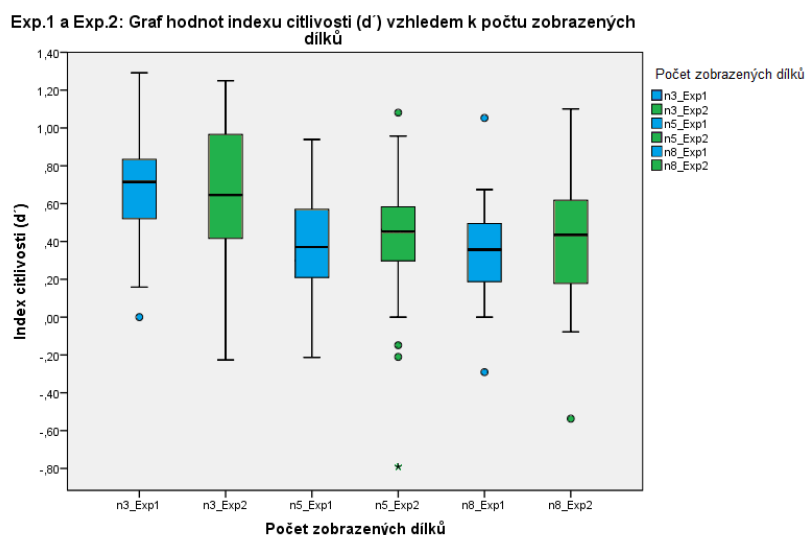
Bias a d'

V Exp.2 se celkový *bias* pohyboval v rozmezí -0,66 až 0,58 (průměr = -0,12, SD = 0,33). To znamená, že celková tendence participantů odpovídat určitým způsobem byla lehce nakloněna k *liberálnosti* v odpovědích čili k tendenci odpovídat *ano*. V případě tří dílků $c(n3)$ bylo v rozmezí -0,40 až 0,88 (průměr = 0,16, SD = 0,32), v případě pěti dílků $c(n5)$ bylo v rozmezí -0,99 až 0,63 (průměr = -0,16, SD = 0,40) a v případě osmi dílků $c(n8)$ se pohybovalo v rozmezí -1,22 až 0,70 (průměr = -0,37, SD = 0,47). Z výsledných hodnot můžeme pozorovat, že z čím více dílků dotazovaný patch pocházel, tím *liberálnější* participantů ve svých odpovědích byli a tím spíše odpovídali *ano*, že podnět viděli.



Obrázek XX. Exp.1 a Exp.2: Graf průměrného biasu vzhledem k počtu zobrazených dílků

V případě d' se celková hodnota pohybovala v rozmezí -0,47 až 0,91 (průměr = 0,46, SD = 0,25). V případě tří dílků $d'(n3)$ bylo v rozmezí -0,23 až 1,25 (průměr = 0,66, SD = 0,37), v případě pěti dílků $d'(n5)$ bylo v rozmezí -0,79 až 1,08 (průměr = 0,39, SD = 0,37) a v případě osmi dílků $d'(n8)$ se pohybovalo v rozmezí -0,54 až 1,10 (průměr = 0,41, SD = 0,34). Z výsledných hodnot lze usuzovat, že participantů nejlépe určovali patche v případě tří dílků a dále pak velmi podobně v případě pěti a osmi dílků. Grafické srovnání hodnot d' z Exp.1 a Exp.2 je na Obrázku XXI.



Obrázek XXI. Exp.1 a Exp.2: Graf průměrného d' vzhledem k počtu zobrazených dílků

5.2.3 Regresní analýza

Podobně jako v Exp.1 nás zajímá pravděpodobnost, s jakou můžeme předpovídat správnost odpovědi (závisle proměnná) s ohledem na počet zobrazených dílků (nezávisle proměnná). K analýze je použita binární logistická regrese.

Logistický regresní model se ukázal být statisticky významný, $\chi^2(2) = 15,248$, $p < ,001$, přičemž velikost vztahu mezi oběma proměnnými je 0,3% (Nagelkerke $R^2 = 0,003$). Správnost odpovědi v modelu je 58,6 %. Jak je vidět v Tabulce II., u počtu tří zobrazených dílků (n3) byla 1,6krát vyšší pravděpodobnost správné odpovědi nežli odpovědi nesprávné. V případě pěti dílků (n5) byla šance na správnou odpověď nižší ($\text{Exp}(B) = 0,81$) a v případě osmi dílků (n8) ($\text{Exp}(B) = 0,80$).

Tabulka II. Exp.2: koeficienty logistické regrese

	B	Std. chyba	Poměr šancí [Exp(B)]	p	95% Interval spolehlivosti	
					Dolní mez	Horní mez
N3 (referenční skupina)	0,494	0,046	1,639	< ,001	0,404	0,584
N5	-0,216	0,064	0,805	< ,001	-0,342	-0,090
N8	-0,218	0,064	0,804	< ,001	-0,344	-0,092

Poznámka. Správnost odpovědi „1“ kódováno jako skupina 1.

5.2.4 ANCOVA

V Exp.1 oproti Exp.2 byla do podnětových fragmentovaných fotografií začleněna mřížka. Zajímalo nás, jestli se mřížka podílí na tom, jak správně budou participanti odpovídat v souvislosti s určitým počtem zobrazených dílků. Pro prozkoumání této otázky je využito metody dvoucestná ANCOVA.

Pro zjištění rozdílů v d' byly do analýzy začleněny proměnné „experiment“ (1; 2) a „počet zobrazených dílků“ (3; 5; 8). Pro analýzu byly splněny podmínky homogenity rozptylu $F(5,168) = 1,393$, $p = 0,229$. Interakce mezi experimentem a počtem zobrazených dílků nebyla statisticky významná $F(2) = 0,248$, $p = 0,780$ a ani rozdíl v samotném experimentu nebyl statisticky významný $F(1) = 0,408$, $p = 0,524$. Jako statisticky významný se ukázal být pouze počet zobrazených dílků $F(2) = 16,335$, $p < ,001$.

Tabulka III. Exp.1 a Exp.2: ANCOVA, párové srovnání d'

Párové srovnání

Závisle proměnná: d'

Počet zobrazených dílků	Počet zobrazených dílků	Rozdíl průměrů	Std. chyba	Sig. ^b	95% Interval spolehlivosti pro rozdíly ^b	
					Dolní mez	Horní mez
n3	n5	,295*	,059	,000	,151	,438
	n8	,293*	,059	,000	,149	,437
n5	n3	-,295*	,059	,000	-,438	-,151
	n8	-,002	,059	1,000	-,145	,142
n8	n3	-,293*	,059	,000	-,437	-,149
	n5	,002	,059	1,000	-,142	,145

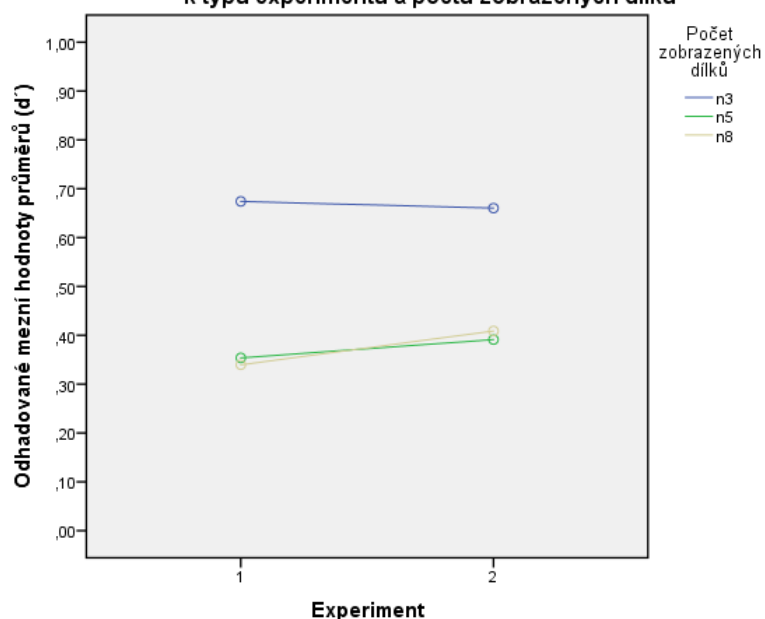
Založené na odhadovaných mezních hodnotách průměrů.

*. Rozdíl průměrů je signifikantní na úrovni ,05.

b. Upraveno pro vícenásobná srovnání: Bonferroni.

Z Tabulky III. můžeme vypořádat, že pouze v případě srovnání n3 s n5 nebo n3 s n8 (bez ohledu na pořadí) byl rozdíl v průměrech statisticky významný ($p < ,001$). Grafické znázornění mezních hodnot průměrů pro d' s ohledem na typ experimentu a počet zobrazených dílků je na Obrázku XXII.

Exp.1 a Exp.2: ANCOVA odhadovaných mezních hodnot průměrů (d') vzhledem k typu experimentu a počtu zobrazených dílků



Obrázek XXII. Exp.1 a Exp.2: ANCOVA odhadovaných mezních hodnot průměru d' vzhledem k typu experimentu a počtu zobrazených dílků

Pro zjištění rozdílů v *bias* byly do analýzy začleněny proměnné „experiment“ (1; 2) a „počet zobrazených dílků“ (3; 5; 8). Pro analýzu byly splněny podmínky homogenity rozptylu $F(5,168) = 1,330$, $p = 0,254$. Interakce mezi experimentem a počtem zobrazených dílků nebyla statisticky významná $F(2) = 0,160$, $p = 0,853$ a ani rozdíl v samotném experimentu nebyl statisticky významný $F(1) = 1,254$, $p = 0,264$. Jako statisticky významný se ukázal být pouze počet zobrazených dílků $F(2) = 31,626$, $p < ,001$.

Tabulka IV. Exp.1 a Exp.2: ANCOVA, párové srovnání bias

Párové srovnání

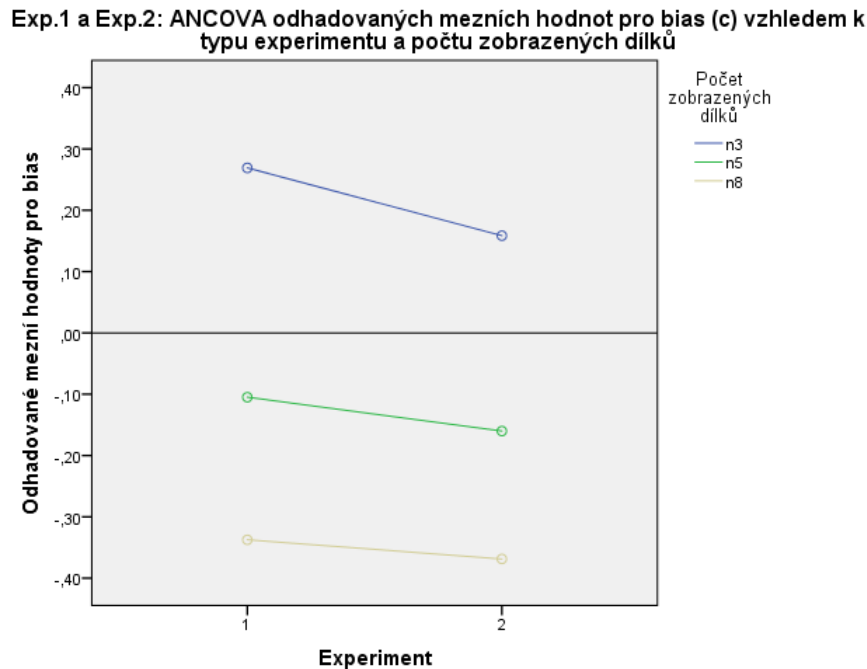
Závisle proměnná: bias

Počet zobrazených dílků	Počet zobrazených dílků	Rozdíl průměrů	Std. chyba	Sig. ^b	95% Interval spolehlivosti pro rozdíly ^b	
					Dolní mez	Horní mez
n3	n5	,346*	,072	,000	,173	,520
	n8	,567*	,072	,000	,393	,741
n5	n3	-,346*	,072	,000	-,520	-,173
	n8	,221*	,072	,008	,047	,394
n8	n3	-,567*	,072	,000	-,741	-,393
	n5	-,221*	,072	,008	-,394	-,047

Založené na odhadovaných mezních hodnotách průměrů.

*. Rozdíl průměrů je signifikantní na úrovni ,05.

Z Tabulky IV. můžeme vypočítat, že všechny hodnoty se od sebe signifikantně liší ($p < ,001$; $p < ,01$). Všechny hodnoty pro *bias* jsou graficky znázorněny na Obrázku XXIII.



Obrázek XXIII. Exp.1 a Exp.2: ANCOVA odhadovaných mezních hodnot průměru bias vzhledem k typu experimentu a počtu zobrazených dílků

5.3 Diskuze

V Exp.1 oproti Exp.2 byla do podnětových fragmentovaných fotografií začleněna mřížka. Zajímalo nás, jestli se mřížka podílí na tom, jak správně budou participanti odpovídat v souvislosti s určitým počtem zobrazených dílků čili jestli mřížka může napomáhat k lepšímu zapamatování si při určitém počtu dílků, anebo naopak přesnějšímu zapamatování zabraňovat. Ze srovnání výsledků Exp.1 a Exp.2 na všech úrovních (základní skóry pro HIT, FA, MISS a CR, accuracy, bias, d' , logistická regrese i ANCOVA) vyplývá, že výsledné hodnoty se mezi sebou významně nelišily a docházelo ke stejným trendům, které jsem již popsala v diskuzi k Exp.1. Z toho vyplývá, že přítomnost nebo nepřítomnost mřížky významně neovlivnila ani přesnost odpovědí (d'), ani tendenci odpovídat určitým způsobem (bias), a tudíž mřížka nemá zásadní vliv na VSTM.

Do designu stávajících experimentů by bylo vhodné začlenit více kritérií pro kontrolu pozornosti participantů. Mezi testové stimuly by příště měly být vřazeny kontrolní podněty, které by nebyly náročné pro výkon VSTM, ale zjišťovaly by míru úspěšnosti odpovědi na lehké úrovni, čímž by

se prověřila aktuální pozornost participantů v průběhu celého testování. Zároveň by z analýzy dat měly být vyřazeny odpovědi pod 0,5 s, což je určitým nedostatkem této práce, neboť by pak došlo ke zpřesnění hodnot. V každém z experimentů se takové případy objevily čtyři. Dále by bylo vhodné znáhodnit pořadí prezentovaných stimulů pro každého participanta zvlášť, což jsem v těchto experimentech opomenula.

Design experimentu s fragmentovanými fotografiemi je prozatím ne příliš probádané pole, které by mohlo přinést mnoho zajímavých poznatků. Náš design byl zvolen záměrně na co nejjednodušší jednotky, neboť pro složitější designy zatím chybělo jakékoli srovnání se základními proměnnými. Možnými cestami dalších výzkumů by se tak mohlo stát, po kolika fragmentovaných fotografiích je člověk stále schopen správně odpovídat na dotazovaný patch jedné z nich? Jak se změní správnost odpovědí dojde-li k prodloužení či zkrácení prezentace fragmentované fotografie? Byli by participantů schopni po několika minutách správně rozlišit pouze na základě dotazovaných patchů, ze kterých celistvých fotografií patche viděli? Jakou roli by v tom hrál počet prezentovaných dílků?

Závěr

V této práci jsem se zabývala otázkou role detailů a extrapolace u fragmentovaných scén. Byly provedeny dva experimenty a u obou byly zjištěny největší hodnoty správnosti v odpovědích u 3 zobrazených dílků (Exp.1: $d'(n3) = 0,67$; Exp.2: $d'(n3) = 0,66$), u kterých převažovala tendence biasu k záporným odpovědím (Exp.1: $c(n3) = 0,27$; Exp.2: $c(n3) = 0,16$). U počtu 5 a 8 dílků se správnost odpovědí snížila (Exp.1: $d'(n5) = 0,35$; $d'(n8) = 0,34$; Exp.2: $d'(n5) = 0,39$; $d'(n8) = 0,41$) a zároveň se bias přiklonil ke kladným odpovědím (Exp.1: $c(n5) = -0,11$, $c(n8) = -0,34$; Exp.2: $c(n5) = -0,16$, $c(n8) = -0,37$). V neposlední řadě nás také zajímal možný vliv mřížky na zapamatování, který se však neprokázal.

Seznam použité literatury

- Aginsky, V., & Tarr, M. J. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition*, 7(1/2/3), 147-162.
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122-131.
- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15(2), 106-111.
- Alvarez, G. A., & Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than on surface features. *Perception & Psychophysics*, 70(2), 346-364.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, 18(7), 622-628.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Eds.), *Psychology of learning and motivation*, (Vol. 2, pp. 89-195). New York: Academic Press.
- Baddeley, R. J., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Eds.), *Psychology of learning and motivation*, (Vol. 8, pp. 47-90). New York: Academic Press.
- Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research*, 46, 2824-2833.
- Bainbridge, W. A. (2017). The resiliency of memorability: A predictor of memory separate from attention and priming. *ArXiv preprint arXiv:1703.07738*, 1-37.
- Bainbridge, W. A., Dilks, D. D., & Oliva, A. (2017). Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *NeuroImage*, 149, 141-152.
- Bainbridge, W. A., & Rissman, J. (2018). Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Scientific Reports*, 8(1), 8679, 1-11.
- Bengson, J. J., & Luck, S. J. (2016). Effects of strategy on visual working memory capacity. *Psychonomic Bulletin & Review*, 23, 265-270.
- Bi, X., Bae, S., & Balakrishnan, R. (2010). Effects of interior bezels of tiled-monitor large displays on visual search, tunnel steering, and target selection. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, 65-74.
- Bradley, C., & Pearson, J. (2012). The sensory components of high-capacity iconic memory and visual working memory. *Frontiers in Psychology*, 3, 1-8.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, 22(3), 384-392.
- Brady, T. F., & Alvarez, G. A. (2015a). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, 15(15):6, 1-24.

- Brady, T. F., & Alvarez, G. A. (2015b). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 921-929.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*(4), 487-502.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5):4, 1-34.
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325-14329.
- Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, *24*(6), 981-990.
- Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. *Communicative & Integrative Biology*, *2*(1), 1-3.
- Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, *19*(7), 678-685.
- Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences*, *113*(27), 7459-7464.
- Brady, T. F., & Tenenbaum, J. B. (2010). Encoding higher-order structure in visual working memory: A probabilistic model. In *32nd Annual Meeting of the Cognitive Science Society*.
- Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*(1), 85-109.
- Bussey, T. J., & Saksida, L. M. (2007). Memory, perception, and the ventral visual-perirhinal-hippocampal stream: Thinking outside of the boxes. *Hippocampus*, *17*(9), 898-908.
- Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*, 165-178.
- Bylinskii, Z., Isola, P., Torralba, A., & Oliva, A. (2015a). How you look at a picture determines if you will remember it. *Eye*, *65*(70), 75-76.
- Bylinskii, Z., Isola, P., Torralba, A., & Oliva, A. (2015b). Modeling context effects on image memorability. 1-2.
- Caddigan, E., Choo, H., Fei-Fei, L., & Beck, D. M. (2017). Categorization influences detection: A perceptual advantage for representative exemplars of natural scene categories. *Journal of Vision*, *17*(1):21, 1-11.
- Calabi, C. (2013). Amodal completion, perception and visual imagery. *Phenomenology and Mind*, *4*, 233-243.

- Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, 12(6), 1017-1040.
- Cauchoix, M., Barragan-Jason, G., Serre, T., & Barbeau, E. J. (2014). The neural dynamics of face detection in the wild revealed by MVPA. *Journal of Neuroscience*, 34(3), 846-854.
- Clarke, A. D. F., & Tatler, B. (2014). Deriving an appropriate baseline for describing fixation behaviour. *Vision Research*, 102, 41-51.
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19(1), 51-57.
- Cowan, N. (2017). Mental objects in working memory: Development of basic capacity or of cognitive completion? *Advances in Child Development and Behavior*, 52, 81-104.
- Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review*, 14(4), 620-628.
- Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 94-107.
- Cutting, J. E. (2003). Reconceiving perceptual space. In H. Hecht, M. Atherton, & R. Schwartz (Eds.), *Perceiving pictures: An interdisciplinary approach to pictorial space* (pp. 215-238). Cambridge, Massachusetts: MIT Press.
- Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, 35(3), 393-401.
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15(8), 559-564.
- Deco, G., Rolls, E. T., & Romo, R. (2009). Stochastic dynamics as a principle of brain function. *Progress in Neurobiology*, 88(1), 1-16.
- Delvenne, J.-F., Braithwaite, J. J., Riddoch, M. J., & Humphreys, G. W. (2002). Capacity limits in visual short-term memory for local orientations. *Current Psychology of Cognition*, 21(6), 681-690.
- Delvenne, J.-F., & Bruyer, R. (2004). Does visual short-term memory store bound features? *Visual Cognition*, 11(1), 1-27.
- Delvenne, J.-F., & Bruyer, R. (2006). A configural effect in visual short-term memory for features from different parts of an object. *The Quarterly Journal of Experimental Psychology*, 59(9), 1567-1580.
- Dubey, R., Peterson, J., Khosla, A., Yang, M.-H., & Ghanem, B. (2015). What makes an object memorable? *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1089-1097.
- Ehinger, K. A. (2013). *Visual features for scene recognition and reorientation* (Disertační práce, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA). Staženo z <http://hdl.handle.net/1721.1/81578>
- Ehinger, K. A., & Brockmole, J. R. (2008). The role of color in visual search in real-world scenes: Evidence from contextual cuing. *Perception & Psychophysics*, 70(7), 1366-1378.
- Ehinger, K. A., Xiao, J., Torralba, A., & Oliva, A. (2011). Estimating scene typicality from human ratings and image features. In *33rd Annual Conference of the Cognitive Science Society (COGSCI)*.

- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, 7(1):10, 1-29.
- Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE Publications.
- Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, 10(12):27, 1-11.
- Fougnie, D., Cormiea, S. M., Kanabar, A., & Alvarez, G. A. (2016). Strategic trade-offs between quality and quantity in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 42(8), 1231-1240.
- Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3, 1-8.
- Foulsham, T., Alan, R., Kingstone, A. (2011). Scrambled eyes? Disrupting scene structure impedes focal processing and increases bottom-up guidance. *Attention, Perception, & Psychophysics*, 73(7), 2008-2025.
- Foulsham, T., Gray, A., Nasiopoulos, E., & Kingstone, A. (2013). Leftward biases in picture scanning and line bisection: A gaze-contingent window study. *Vision Research*, 78, 14–25.
- Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features. *Vision Research*, 50(8), 779-795.
- Foulsham, T., & Kingstone, A. (2013). Fixation-dependent memory for natural scenes: An experimental test of scanpath theory. *Journal of Experimental Psychology: General*, 142(1), 41-56.
- Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: competitive content maps for attention and memory. *Trends in Cognitive Sciences*, 17(3), 134-141.
- Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, 17(5), 673-679.
- Goetschalckx, L., Moors, P., & Wagemans, J. (2018). Image memorability across longer time intervals. *Memory*, 26(5), 581-588.
- Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, 12(6), 878-892.
- Greene, M. R. (2016). Estimations of object frequency are frequently overestimated. *Cognition*, 149, 6-10.
- Greene, M. R., Baldassano, C., Esteva, A., Beck, D. M., & Fei-Fei, L. (2016). Visual scenes are categorized by function. *Journal of Experimental Psychology: General*, 145(1), 82-94.
- Greene, M. R., & Hansen, B. C. (2019). From pixels to scene categories: Unique and early contributions of functional and visual features. *Computational Cognitive Neuroscience*, in print.
- Greene, M. R., & Oliva, A. (2006). Natural scene categorization from conjunctions of ecological global properties. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 291-296). Norwood, New Jersey: Erlbaum.
- Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, 58(2), 137-176.

- Greene, M. R., & Oliva, A. (2009b). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20*(4), 464-472.
- Greene, M. R., & Oliva, A. (2010). High-level aftereffects to global scene properties. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1430-1442.
- Groen, I. I. A., Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philosophical Transactions of the Royal Society B, 372*(1714), 1-11.
- Guérin-Dugué, A., & Oliva, A. (2000). Classification of scene photographs from local orientations features. *Pattern Recognition Letters, 21*, 1135-1140.
- Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., & Van Gool, L. (2013). The interestingness of images. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1633-1640.
- Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple independent levels of ensemble representation. *Journal of Experimental Psychology: General, 144*(2), 432-446.
- Hall, P., & Owen, M. (2005). Simple canonical views. *The British Machine Vision Conference (BMVC)*, 839-848.
- Hayes, T. R., & Henderson, J. M. (2017). Scan patterns during real-world scene viewing predict individual differences in cognitive capacity. *Journal of Vision, 17*(5):23, 1-17.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*(11), 498-504.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243-271.
- Hidalgo-Sotelo, B., & Oliva, A. (2010). Person, place, and past influence eye movements during visual search. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 820-825). Austin, Texas: Cognitive Science Society.
- Hollingworth, A. (2005a). Memory for object position in natural scenes. *Visual Cognition, 12*(6), 1003-1016.
- Hollingworth, A. (2005b). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(3), 396-411.
- Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Human Perception and Performance, 32*(1), 58-69.
- Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance, 33*(1), 31-47.
- Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for binding of target object to scene location. *Visual Cognition, 17*(1/2), 273-291.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 113-136.

- Hollingworth, A., & Rasmussen, I. P. (2010). Binding objects to locations: The relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 36(3), 543-564.
- Hutchings, D. R., Czerwinski, M., Meyers, B., & Stasko, J. (2004). Exploring the use and affordances of multiple display environments. *Workshop on Ubiquitous Display Environments at UbiComp*, 1–6.
- Chekaf, M., Cowan, N., & Mathy, F. (2016). Chunk formation in immediate memory and how it relates to data compression. *Cognition*, 155, 96-107.
- Choi, M. J., Torralba, A., & Willsky, A. S. (2012). Context models and out-of-context objects. *Pattern Recognition Letters*, 33(7), 853-862.
- Chokron, S., & De Agostini, M. (1995). Reading habits and line bisection: A developmental approach. *Cognitive Brain Research*, 3(1), 51–58.
- Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(2), 224-234.
- Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011). Understanding the intrinsic memorability of images. *NIPS*, 1-9.
- Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 1469-1482.
- Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? *Cognition Vision and Pattern Recognition (CVPR)*, 145-152.
- Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 683-702.
- Jiang, Y., Song, J.-H., & Rigas, A. (2005). High-capacity spatial contextual memory. *Psychonomic Bulletin & Review*, 12(3), 524-529.
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, 47, 3286-3297.
- Judd, T., Durand, F., & Torralba, A. (2011). Fixations on low-resolution images. *Journal of Vision*, 11(4):14, 1-20.
- Khaligh-Razavi, S.-M., Bainbridge, W. A., Pantazis, D., & Oliva, A. (2016). From what we perceive to what we remember: Characterizing representational dynamics of visual memorability. *bioRxiv*, 049700.
- Khosla, A., An, B., Lim, J. J., & Torralba, A. (2014). Looking beyond the visible scene. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Khosla, A., Xiao, J., Isola, P., Torralba, A., & Oliva, A. (2012). Image memorability and visual inception. *SIGGRAPH Asia 2012 Technical Briefs*, 35.
- Koenderink, J., Richards, W., & Doorn, A. van. (2012). Blow-up: A free lunch? *i-Perception*, 3(2), 141-145.
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, 139(3), 558-578.

- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551-1556.
- Konkle, T., & Oliva, A. (2007). Normative representation of objects: Evidence for an ecological bias in object perception and memory. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 407–413). Austin, Texas: Cognitive Science Society.
- Laaksonen, J., Oja, E., Koskela, M., & Brandt, S. (2000). Analyzing low-level visual features using content-based image retrieval. *Proceedings of ICONIP 2000*.
- Lakshman, H., Ndjiki-Nya, P., Köppel, M., Doshkov, D., & Wiegand, T. (2009). An automatic structure-aware image extrapolation applied to error concealment. *Proceedings of International Conference on Image Processing (ICIP)*, 913-916.
- Lamsweerde, A. E. van, Beck, M. R., & Johnson, J. S. (2016). Visual working memory organization is subject to top-down control. *Psychonomic Bulletin & Review*, *23*, 1181-1189.
- Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9*(10):6, 1-16.
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365-373.
- Levi, D. M., & Carney, T. (2009). Crowding in peripheral vision: Why bigger is better. *Current Biology*, *19*(23), 1988-1993.
- Li, L.-J., Su, H., Lim, Y., & Fei-Fei, L. (2012). Objects as attributes for scene classification. In K. N. Kutulakos (Eds.), *Trends and Topics in Computer Vision. Lecture Notes in Computer Science*, *6553* (pp. 1-13). Berlin, Heidelberg: Springer.
- Liu, K., & Jiang, Y. (2005). Visual working memory for briefly presented scenes. *Journal of Vision*, *5*, 650-658.
- Logie, R. H., Brockmole, J. R., & Vandembroucke, A. R. E. (2009). Bound feature combinations in visual short-term memory are fragile but influence long-term learning. *Visual Cognition*, *17*(1/2), 160-179.
- Long, B., Konkle, T., Cohen, M. A., & Alvarez, G. A. (2016). Mid-level perceptual features distinguish objects of different real-world sizes. *Journal of Experimental Psychology: General*, *145*(1), 95-109.
- Loschky, L. C., & Larson, A. M. (2008). Localized information is necessary for scene categorization, including the natural/man-made distinction. *Journal of Vision*, *8*(1):4, 1-9.
- Loschky, L. C., Nuthmann, A., Fortenbaugh, F. C., & Levi, D. M. (2017). Scene perception from central to peripheral vision. *Journal of Vision*, *17*(1):6, 1-5.
- Luck, S. J., & Hollingworth, A. (2008). *Visual memory*. New York: Oxford University Press.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279-281.
- Lukavský, J., & Děchtěrenko, F. (2015). The effect of scene category distinctiveness on memory performance. *Journal of Vision*, *15*(12), 357.
- Machizawa, M. G., Goh, C. C. W., & Driver, J. (2012). Human visual short-term memory precision can be varied at will when the number of retained items is low. *Psychological Science*, *20*(10), 1-6.

- Mancas, M., & Le Meur, O. (2013). Memorability of natural scenes: The role of attention. In *2013 IEEE International Conference on Image Processing*, 196-200.
- Mante, V., Frazor, R. A., Bonin, V., Geisler, W. S., & Carandini, M. (2005). Independence of luminance and contrast in natural scenes and in the early visual system. *Nature Neuroscience*, *8*(12), 1690-1697.
- Mohsenzadeh, Y., Oliva, A., & Pantazis, D. (2017). Unconscious perception of scenes reveals a perceptual neural signature of memorability. *Journal of Vision*, *17*(10), 570-571.
- Murphy, K., Torralba, A., & Freeman, W. T. (2004). Using the forest to see the trees: A graphical model relating features, objects, and scenes. In S. Thrun, L. K. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems*, *16* (pp. 1499-1506). Cambridge, Massachusetts: MIT Press.
- Nakonečný, M. (2016). *Obečná psychologie*. Praha: Triton.
- Neisser, U. (2014). *Cognitive psychology: Classic edition*. New York & London: Psychology Press.
- Newton, J. R., & Eskew, R. T., Jr. (2003). Chromatic detection and discrimination in the periphery: A postreceptoral loss of color sensitivity. *Visual Neuroscience*, *20*, 511-521.
- Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8):20, 1-19.
- Oakes, L. M., Ross-Sheehy, S., & Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychological Science*, *17*(9), 781-787.
- Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 251-256). Cambridge, Massachusetts: Academic Press.
- Oliva, A. (2014). Scene perception. In J. S. Werner, & M. L. Chalupa (Eds.), *The new visual neurosciences* (pp. 725-732). Cambridge, Massachusetts: MIT Press.
- Oliva, A., Isola, P., Khosla, A., & Bainbridge, B. A. (2013). What makes a picture memorable? *SPIE*, 2-4.
- Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*, 176-210.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145-175.
- Oliva, A., & Torralba, A. (2002). Scene-centered description from spatial envelope properties. In H. H. Bülthoff, S.-W. Lee, T. Poggio, & C. Wallraven (Eds.), *Lecture notes in computer science*, *2525. Biologically motivated computer vision* (pp. 263-272). New York: Springer-Verlag.
- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23-36.
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520-527.
- Oliva, A., Torralba, A., Guérin-Dugué, A., & Hérault, J. (1999). Global semantic classification of scenes using power spectrum templates. In *Proceedings of the Challenge of Image Retrieval (CIR)*, Electronic Workshops in Computing Series. Newcastle: Springer-Verlag.
- Olson, I. R., & Jiang, Y. (2004). Visual short-term memory is not improved by training. *Memory & Cognition*, *32*(8), 1326-1332.

- Olson, I. R., Jiang, Y., & Moore, K. S. (2005). Associative learning improves visual working memory performance. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 889-900.
- Orel, M., & Procházka, R. (2017). *Vyšetření a výzkum mozku: pro psychology, pedagogy a další nelékařské obory*. Praha: Grada.
- Peirce, J. (2018). *Building experiments in PsychoPy*. Thousand Oaks, CA: SAGE Publications.
- Pinto, Y., Sligte, I. G., Shapiro, K. L., & Lamme, V. A. F. (2013). Fragile visual short-term memory is an object-based and location-specific store. *Psychonomic Bulletin & Review*, *20*, 732-739.
- Plháková, A. (2004). *Učebnice obecné psychologie*. Praha: Academia.
- Quattoni, A., & Torralba, A. (2009). Recognizing indoor scenes. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Quinlan, P. T., & Dyson, B. (2008). *Cognitive psychology*. New York: Pearson/Prentice Hall.
- Ranganath, C., Cohen, M. X., & Brozinsky, C. J. (2005). Working memory maintenance contributes to long-term memory formation: neural and behavioral evidence. *Journal of Cognitive Neuroscience*, *17*(7), 994-1010.
- Rivolta, D., Palermo, R., Schmalzl, L., & Williams, M. A. (2012). An early category-specific neural response for the perception of both places and faces. *Cognitive Neuroscience*, *3*(1), 45-51.
- Ross, M. G., & Oliva, A. (2010). Estimating perception of scene layout properties from global image features. *Journal of Vision*, *10*(1):2, 1-25.
- Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*(16), 5975-5979.
- Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the „gist“ of real-world natural scenes? *Visual Cognition*, *12*(6), 852-877.
- Sanocki, T., Michelet, K., Sellers, E., & Reynolds, J. (2006). Representations of scene layout can consist of independent, functional pieces. *Perception & Psychophysics*, *68*(3), 415-427.
- Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review*, *15*(1), 215-222.
- Seiler, J., & Kaup, A. (2008). Fast orthogonality deficiency compensation for improved frequency selective image extrapolation. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 781-784.
- Seiler, J., Meisinger, K., & Kaup, A. (2008). Orthogonality deficiency compensation for improved frequency selective image extrapolation. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1-4.
- Sheth, B. R., & Shimojo, S. (2001). Compression of space in visual memory. *Vision Research*, *41*, 329-341.
- Schurigin, M. W. (2018). Visual memory, the long and the short of it: a review of visual working memory and long-term memory. *Attention, Perception, & Psychophysics*, *80*, 1035-1056.

- Schurgin, M. W., & Flombaum, J. I. (2015). Visual long-term memory has weaker fidelity than working memory. *Visual Cognition*, 23(7), 859-862.
- Schurgin, M. W., & Flombaum, J. I. (2018). Visual working memory is more tolerant than visual long-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 44(8), 1216-1227.
- Siarohin, A., Zen, G., Majtanovic, C., Alameda-Pineda, X., Ricci, E., & Sebe, N. (2017). How to make an image more memorable? A deep style transfer approach. *ACM ICMR*.
- Sligte, I. G., Vandenbroucke, A. R. E., Scholte, H. S., & Lamme, V. A. F. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in Psychology*, 1, 1-10.
- Spence, I., Wong, P., Rusan, M., & Rastegar, N. (2006). How color enhances visual memory for natural scenes. *Psychological Science*, 17(1), 1-6.
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2015). Disentangling the effects of spatial inconsistency of targets and distractors when searching in realistic scenes. *Journal of Vision*, 15(2):12, 1-21.
- Spotorno, S., & Tatler, B. W. (2017). The elephant in the room: Inconsistency in scene viewing and representation. *Journal of Experimental Psychology: Human Perception and Performance*, 43(10), 1717-1743.
- Squalli-Houssaini, H., Duong, N. Q. K., Gwenaëlle, M., & Demarty, C.-H. (2018). Deep learning for predicting image memorability. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
- Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82(3), 171-177.
- Stainer, M. J., Scott-Brown, K. C., & Tatler, B. W. (2013). Behavioral biases when viewing multiplexed scenes: Scene structure and frames of reference for inspection. *Frontiers in Psychology*, 4(624), 1-12.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137-149.
- Suchow, J. W., Fougny, D., & Alvarez, G. A. (2017). Looking inwards and back: realtime monitoring of visual working memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(4), 660-668.
- Suchow, J. W., Fougny, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception, & Psychophysics*, 76(7), 2071-2079.
- Šikl, R. (2012). *Zrakové vnímání*. Praha: Grada.
- Tan, D. S., & Czerwinski, M. (2003). Effects of visual separation and physical discontinuities when distributing information across multiple displays. In *OZCHI 2003 Conference for the Computer-Human Interaction Special Interest Group of the Ergonomics Society of Australia* (pp. 1–8).
- Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, 61(6), 1140-1153.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):4, 1-17.

- Tatler, B. W. (2008). On the perception of natural scenes: An introduction to the special issue. *Journal of Eye Movement Research*, 2(2):1, 1-4.
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45, 643-659.
- Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, 46(12), 1857-1862.
- Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files. *The Quarterly Journal of Experimental Psychology*, 58A(5), 931-960.
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5):5, 1-23.
- Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society*, 366, 596-610.
- Tatler, B. W., & Melcher, D. (2007). Pictures in mind: Initial encoding of object properties varies with the realism of the scene stimulus. *Perception*, 36, 1715-1729.
- Torralba, A. (2003a). Contextual priming for object detection. *International Journal of Computer Vision*, 53(2), 169-191.
- Torralba, A. (2003b). Modeling global scenes factors in attention. *Journal of the Optical Society of America*, 20(7), 1407-1418.
- Torralba, A. (2005). Contextual influences on saliency. *Neurobiology of Attention*, 586-593.
- Torralba, A. (2009). How many pixels make an image? *Visual Neuroscience*, 26(1), 123-131.
- Torralba, A., Murphy, K. P., & Freeman, W. T. (2010). Using the forest to see the trees: exploiting context for visual object detection and localization. *Communications of the ACM*, 53, 107-114.
- Torralba, A., & Oliva, A. (1999). Semantic organization of scenes using discriminant structural templates. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 1253-1258.
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, 14(3), 391-412.
- Torralba, A., Oliva, A., Castelhana, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113(4), 766-786.
- Torralba, A., & Sinha, P. (2001a). Recognizing indoor scenes. *Technical Report 2001-015, AI Memo*.
- Torralba, A., & Sinha, P. (2001b). Statistical context priming for object detection. In *IEEE Proceedings of International Conference in Computer Vision (ICCV)*, 1-8.
- Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, 34(8), 1704-1719.
- Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and subjective expansion of time. *Perception & Psychophysics*, 66(7), 1171-1189.

- Van Den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780-8785.
- VanRullen, R., Reddy, L., & Fei-Fei, L. (2005). Binding is a local problem for natural objects and scenes. *Vision Research*, *45*(25–26), 3133–3144.
- Varakin, D. A., Klemes, K. J., & Porter, K. A. (2013). The effect of scene structure on time perception. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1639-1652.
- Varakin, D. A., & Levin, D. T. (2008). Scene structure enhances change detection. *The Quarterly Journal of Experimental Psychology*, *61*(4), 543-551.
- Velisavljević, L., & Elder, J. H. (2002). What do we see in a glance? [Abstract]. *Journal of Vision*, *2*(7): 493, 493a.
- Velisavljević, L., & Elder, J. H. (2008a). Visual short-term memory for natural scenes: Effects of eccentricity. *Journal of Vision*, *8*(4):28, 1-17.
- Velisavljević, L., & Elder, J. H. (2008b). Visual short-term memory of local information in briefly viewed natural scenes: Configural and non-configural factors. *Journal of Vision*, *8*(16):8, 1-17.
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748-751.
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500-503.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92-114.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436-1451.
- Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology*, *54*, 298-303.
- Wang, M., Lai, Y.-K., Liang, Y., Martin, R. R., & Hu, S.-M. (2014). BiggerPicture: Data-driven image extrapolation using graph matching. *ACM Transactions on Graphics*, *33*(6), 173-186.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48-64.
- Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.
- Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, *10*(1), 80-87.
- Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review*, *15*(1), 223-229.
- Xiao, J., Ehinger, K. A., Hays, J., Torralba, A., & Oliva, A. (2016). SUN database: Exploring a large collection of scene categories. *International Journal of Computer Vision (IJCV)*, *119*(1), 3-22.

- Xiao, J., Hays, J., Ehinger, K., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Xiao, J., Hays, J., Russell, B. C., Patterson, G., Ehinger, K. A., Torralba, A., & Oliva, A. (2013). Basic level scene understanding: Categories, attributes and structures. *Frontiers in Psychology, 4*, 1-10.
- Xu, Y. (2002a). Encoding color and shape from different parts of an object in visual short-term memory. *Perception & Psychophysics, 64*(8), 1260-1280.
- Xu, Y. (2002b). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 458-468.
- Xu, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Perception & Psychophysics, 68*(5), 815-828.
- Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440*, 91-95.
- Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Science, 13*(4), 167-174.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235.
- Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science, 20*(4), 423-428.
- Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science, 22*(11), 1434-1441.
- Zhang, Y., Xiao, J., Hays, J., & Tan, P. (2013). FrameBreak: Dramatic image extrapolation by guided shift-maps. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1171-1178*.
- Zhou, B., & Zhang, L. (2010). Scene gist: A holistic generative model of natural image. In H. Zha, R. Taniguchi, S. Maybank (Eds.), *Computer Vision – ACCV 2009. Lecture notes in Computer Science, 5995* (pp. 395-404). Berlin, Heidelberg: Springer.

Internetové zdroje

PsychoPy [Online]. Retrieved April 20, 2019, from <http://www.psychopy.org>

SUN database [Online]. Retrieved April 10, 2019, from <https://groups.csail.mit.edu/vision/SUN/>

SUN Database: Scene Categorization Benchmark [Online]. Retrieved April 10, 2019, from <https://vision.princeton.edu/projects/2010/SUN/>