

---

## 6 PRE-PROCESSING DAT

Před vlastní analýzou dat je nutné zkontrolovat jejich kvalitu, vyřadit respondenty, u kterých došlo k chybě záznamu, a případně převést data do jiného softwaru. V této kapitole bude popsán převod dat z programů od společností SMI a Gazepoint do open-source nástroje OGAMA. Kromě rozdílné funkcionality aplikací je jednou z hlavních výhod OGAMA to, že ji lze využívat bez nutnosti licence či hardwarového klíče. V další části kapitoly bude popsán princip detekce fixací a sakád a doporučeno optimální nastavení pro kartografické experimenty. Závěr kapitoly bude věnován čištění dat a odhalování chyb v datech.

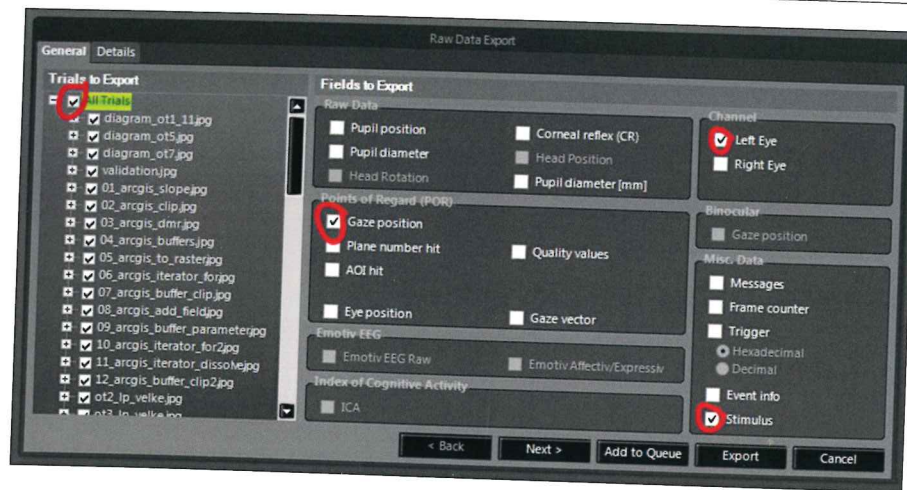
### 6.1 Převod dat ze SMI do OGAMA

Pro převod dat z programu SMI BeGaze do OGAMA byl vytvořen online nástroj SMI2OGAMA.<sup>65</sup> Funkcionalita nástroje je velice jednoduchá. Program čte vyexportovaná raw data z programu SMI a převádí je do formátu vhodného pro import do programu OGAMA. Díky změně způsobu exportu raw dat z programu SMI od verze 3.6 nyní existují dvě možnosti, jak export provést.

Původní způsob se nazýval Export Raw Data to File a v nových verzích je dostupný v nabídce Export/Legacy. V tomto případě si uživatel zvolí, pro které stimuly chce data vyexportovat, a z polí pro export vybere „Gaze position“ a „Stimulus“. OGAMA neumí zpracovat binokulární data. Je na uživateli, které oko si zvolí. V případě, že vybere obě dvě, do OGAMA budou naimportovány pouze záznamy pro levé oko. Nastavení je znázorněno na obrázku 86. V dalším kroku exportu je důležité zatrhnout volbu „Single file output“. Výsledná data budou mít formu textových dokumentů a budou uložena pro každého respondenta zvlášť.

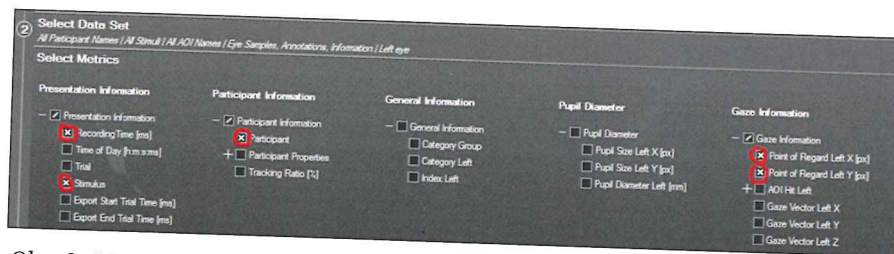
---

<sup>65</sup> [www.eyetracking.upol.cz/smi2ogama](http://www.eyetracking.upol.cz/smi2ogama)



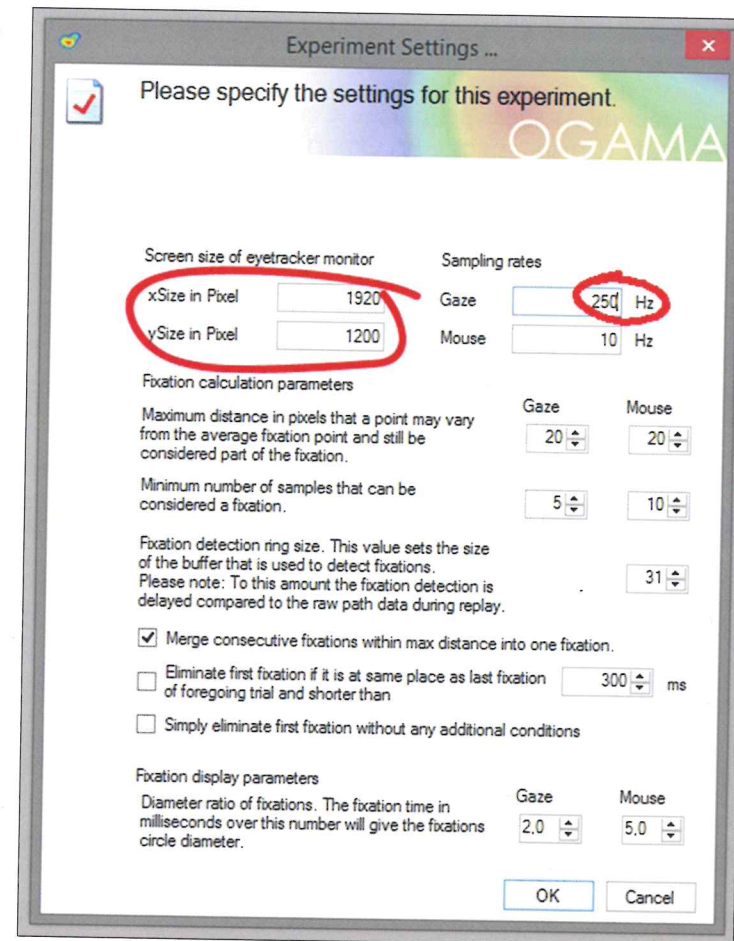
Obr. 86 Nastavení exportu raw dat pomocí Legacy:  
Export raw data to file v programu SMI BeGaze.  
Je důležité, aby byly zaškrtnuty pouze volby označené červenou barvou.

Druhou variantou, dostupnou od SMI BeGaze verze 3.6 a vyšší, je export raw dat pomocí nabídky Metrics Export. Výhodou tohoto postupu je to, že lze exportovat raw data i pro tzv. Custom Trials (viz kapitola 6.5). V tomto případě je nutné zvolit v nabídce Metrics Export šablonu Raw Data, dále vyfiltrovat respondenty a stimuly, pro které má být export proveden. Stejně jako v předchozím případě budou do OGAMA importovány záznamy pouze pro jedno oko. Nejdůležitější je nabídka Select Metrics, kde budou vybrána pole „Recording Time“, „Stimulus“, „Participant“ a „Point of Regard“ pro osu X a Y tak, jak je to znázorněno na obrázku 87. V posledním kroku je opět nutné vybrat, aby byla data uložena pro jednotlivé respondenty zvlášť. To je označeno jako „Individual participant files“.



Obr. 87 Nastavení exportu raw dat pomocí Metrics Export v programu SMI BeGaze.  
Je důležité, aby byly zaškrtnuty pouze volby označené červenou barvou.

Další postup už je velmi podobný pro oba způsoby exportu. Vytvořené textové soubory s daty jednotlivých respondentů je nutné zabalit do ZIP archivu (přímo soubory, ne složku) a tento ZIP nahrát do webové aplikace SMI2OGAMA. Po chvíli dojde ke stažení archivu s daty připravenými pro import do OGAMA. Nevýhodou tohoto importu je fakt, že časy jednotlivých záznamů nejsou unikátní (čas se počítá od zapnutí počítače), proto není možné importovat všechny respondenty najednou (došlo by k promíchání dat mezi respondenty). V programu OGAMA je nutné nejprve založit nový experiment. Zde je důležité specifikovat zejména rozlišení obrazovky, na které byla data nahrávána, a frekvenci záznamu (obr. 88). Další informace týkající se detekce fixací lze změnit i později. Detailněji je toto nastavení popsáno v kapitole 6.3.3.



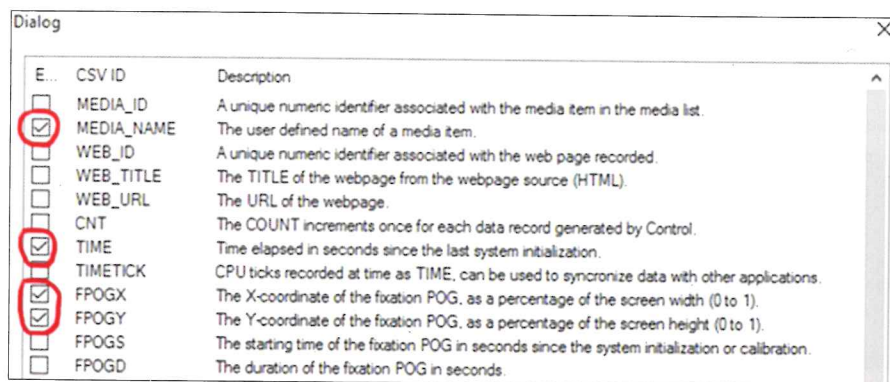
Obr. 88 Tvorba nového experimentu v OGAMA

Po vytvoření experimentu už lze importovat data. Výstupem z programu SMI2OGAMA je sada textových souborů obsahující data za jednotlivé respondenty a soubor ZERO.txt. Ten je nutné nainportovat jako první. Pomocí tohoto souboru se vytvoří ID jednotlivých stimulů, a ty tak budou zobrazené v OGAMA ve správném pořadí. Bez použití tohoto souboru by mohlo docházet k problémům například u experimentů, ve kterých byly stimuly prezentovány v náhodném pořadí. Import v OGAMA probíhá formou průvodce. Výstup z SMI2OGAMA je však vytvořen tak, aby nebylo nutné nic měnit (a lze tedy rovnou stisknout tlačítko Next). Po nainportování všech respondentů už zbývá pouze zkopírovat obrázky stimulů do složky SlideResources daného experimentu v OGAMA. Je nezbytné, aby měly obrázky stejné rozlišení, jako bylo nastavené při tvorbě experimentu. V opačném případě by trajektorie oka neodpovídala zobrazenému stimulu.

## 6.2 Převod dat z Gazepoint do OGAMA

Aplikace Gazepoint Analysis neumožňuje výpočet eye-tracking metrik pro jednotlivé stimuly či respondenty. Při exportu dat dochází pouze k exportu raw dat či fixací za jednotlivé respondenty. Pro statistickou analýzu dat by tedy bylo nutné tato data agregovat například pomocí tabulkového processoru, což by ale bylo velmi časově náročné. Rovněž vizualizace výsledků není příliš povedená a například heat mapy jsou vyloženy nevhodně, neboť není možné jakkoliv změnit barevnou stupnici.

Proto je pro analýzu dat ze zařízení Gazepoint vhodnější využít softwaru OGAMA. Pro převod dat byl vytvořen nástroj s názvem GP2OGAMA.<sup>66</sup> Jeho funkcionality je velmi podobná převodu dat ze SMI popsaném v předchozí kapitole.



Obr. 89 Parametry exportu pro převod dat z Gazepoint Analysis do OGAMA

V prvním kroku je nutné vyexportovat data z programu Gazepoint Analysis. V dialogovém okně exportu je na výběr velké množství parametrů, nicméně důležité jsou pouze čtyři z nich. Jedná se o název stimulu, časovou značku a souřadnice pohledu X a Y (obr. 89). Ty jsou na rozdíl od programů ostatních výrobců uváděny v % monitoru, a nikoliv v pixelech.

Vyexportovány budou dva soubory za každého uživatele. První z nich obsahuje raw data a je zde označen jako „all\_gaze“, druhý z nich pak fixace (označen jako „fixations“). Jelikož v Gazepoint Analysis není možné specifikovat parametry identifikace fixací a import fixací do programu OGAMA nefunguje, je důležitý pouze první typ souborů. Soubory s přídomkem „all\_gaze“ pro všechny respondenty je nutné zabalit do archivu formátu ZIP a nahrát do webového rozhraní aplikace gp2ogama. Zde je rovněž nutné nastavit rozlišení monitoru, na kterém probíhalo testování. Po odeslání dat bude stažen ZIP archiv obsahující soubory připravené pro import do OGAMA. Další postup už je totožný s importem dat ze SMI (viz kapitola 6.1 a obr. 88).

## 6.3 Detekce fixací a sakád

Ve všech eye-tracking studiích je velice důležitá volba a nastavení metody identifikace fixací a sakád, protože toto nastavení může ovlivnit výsledky studie. Důležité je také zvolené nastavení ve studii popsat, protože bez této informace nelze výsledky srovnávat s ostatními pracemi.

Eye-tracker SMI RED 250 ukládá naměřená eye-tracking data do formátu \*.IDT. Pomocí aplikace SMI BeGaze je možné tato data uložit do textového dokumentu ve formě \*.txt. Exportovat lze raw data i identifikované fixace a sakády.

Tab. 7 Vyexportovaná prvotní naměřená (raw) data

Time	Trial	L POR X [px]	L POR Y [px]	Stimulus
2864092592	1	842,89	543,96	010 - 3d.jpg
2864100960	1	845,4	543,45	010 - 3d.jpg
2864109212	1	843,58	544,34	010 - 3d.jpg
2864117593	1	838,73	542,74	010 - 3d.jpg
2864125956	1	839,69	544,45	010 - 3d.jpg
...	...	...	...	...

Exportovaná data obsahují časovou značku a s ní související měřené parametry očí. Jedná se především o polohu pohledu (Point of Regard, POR) ve směru osy X a Y. Tato hodnota je udávána v pixelech použitého monitoru a pro většinu studií, včetně aplikace v kartografii, je tato hodnota nejdůležitější. Mezi další parametry patří například velikost zornice, pozice středu

66 <http://eyetracking.upol.cz/gp2ogama/>

zornice, pozice a úhel natočení hlavy, pozice korneálního odrazu a další. Ukázka vyexportovaných raw dat je v tabulce 7. Tabulka obsahuje pouze časovou značku (Time), souřadnice pozice pohledu levého oka (L POR X, L POR Y) a název obrázku, který se k daným datům vztahuje (Stimulus).

Tato data mohou být v aplikaci SMI BeGaze, případně pomocí dalších nástrojů, např. EMD (Komogortsev a kol., 2010), OGAMA a další, klasifikována na fixace a sakády. Ukázka takto klasifikovaných dat je v tabulce 8. U každé fixace nebo sakády je uvedena časová značka jejího začátku a konce (Start, End), dále délka trvání v mikrosekundách (Duration) a souřadnice v pixelech. U sakád jsou zde uvedeny souřadnice začátku i konce.

Tab. 8 Eye-tracking data klasifikovaná na fixace a sakády

Event	Start	End	Duration	StartX	StartY	EndX	EndY
Fixation	2864092592	2864850875	758283	842.69	519.71		
Saccade	2864850875	2864875886	25011	844.16	489.87	614.32	473.56
Fixation	2864875886	2865475793	599907	577.39	437.20		
Saccade	2865475793	2865509248	33455	586.56	455.07	944.01	679.79
Fixation	2865509248	2865717480	208232	965.64	687.75		
Saccade	2865717480	2865742492	25012	963.91	682.63	966.17	689.33
Fixation	2865742492	2865867448	124956	968.57	687.98		
Saccade	2865867448	2865900712	33334	970.97	674.98	106.58	455.91
Fixation	2865900782	2866117401	216619	1059.43	451.11		
...	...	...	...	...	...	...	...

### 6.3.1 Algoritmy pro detekci fixací a sakád

Fixace je možné identifikovat manuálně (Harris a kol., 1988), ale existuje také velké množství klasifikačních algoritmů. Komogortsev (2010) zmiňuje několik studií, které se těmito algoritmy zabývaly, např. McConkie, 1981; Salvucci a Goldberg, 2000; Sauter a kol., 1991. Mezi nejčastěji používané patří algoritmy I-VT (Velocity Threshold Identification) a I-DT (Dispersion Threshold Identification), jejichž srovnání je znázorněno v tabulce 9.

Tab. 9 Porovnání algoritmů I-VT a I-DT (upraveno podle Salvucci a Goldberg, 2000)

Metoda	Přesnost	Rychlost	Robustnost	Jednoduchost	Počet Parametrů
I-VT	✓	✓✓	✗	✓✓	1
I-DT	✓✓	✓	✓✓	✓	2

Legenda: ✓✓ = velmi dobré, ✓ = dobré, ✗ = špatné

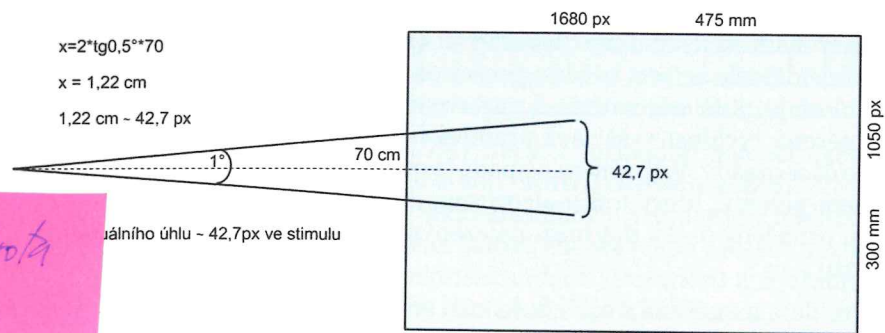
Algoritmus **I-VT** (Velocity Threshold Identification) je využíván pro data naměřená s frekvencí 250 Hz a vyšší (Holmqvist a kol., 2011). V naměřených datech se pro každou pozici oka vypočítá rychlost pohybu. Tato rychlost je poté porovnána s nastavenou prahovou hodnotou. Pokud je naměřená rychlost vyšší než prahová hodnota, je tento záznam označen jako část sakády. V opačném případě je označen jako část fixace. Primárním typem pohybu, který tento algoritmus detekuje, jsou sakády. Jako fixace jsou označeny úseky dat mezi nalezenými sakádami (Salvucci a Goldberg, 2000).

Pro data naměřená s nižší frekvencí se nejčastěji využívá algoritmu **I-DT** (Dispersion Threshold Identification). Ten vyžaduje dva parametry – prahovou hodnotu pro rozptyl (Dispersion) a délku trvání fixace (Duration). Hlavní nevýhodou tohoto algoritmu je vzájemná vysoká závislost těchto dvou parametrů (Salvucci a Goldberg, 2000). Malá prahová hodnota pro rozptyl a velká hodnota pro délku mohou vést k tomu, že nebudou identifikovány žádné fixace. Z tohoto důvodu je u I-DT algoritmu velice důležité správné nastavení parametrů. Při vhodně zvolených prahových hodnotách však I-DT poskytuje velice dobré výsledky (Salvucci a Goldberg, 2000). Algoritmus I-DT je zaměřen na detekci fixací na základě prostorové a časové blízkosti naměřených pozic oka. V algoritmu je definováno časové okno, které se pohybuje naměřenými daty. Prostorová blízkost bodů obsažených v časovém okně je porovnávána s definovanou prahovou hodnotou. V případě, že je rozptyl nižší než prahová hodnota, body obsažené v časovém okně jsou označeny jako část fixace. V opačném případě se okno o jeden záznam posune a první bod předchozího okna je klasifikován jako část sakády (Komogortsev a kol., 2010).

### 6.3.2 Srovnání nastavení algoritmu I-DT

Hledáním optimální prahové hodnoty rozptylu se zabýval Blignaut (2009). Ve své detailní studii došel k závěru, že optimální prahová hodnota pro rozptyl by měla být 1°. Podobnou hodnotu uvádějí i Salvucci a Goldberg (2000), kteří doporučují nastavení prahové hodnoty mezi 0,5° a 1° vizuálního úhlu. Holmqvist (2011) doporučuje stanovení prahové hodnoty v intervalu mezi 0,5° a 2°.

V některých aplikacích se prahové hodnoty uvádějí ve stupních, v jiných v pixelech. V případě, že je oko respondenta vzdálené od roviny obrazovky 70 cm, pak hodnota 1° vizuálního úhlu odpovídá 42,7 px. Výpočet podle goniometrických funkcí ostrého úhlu je znázorněn na obrázku 90.



Obr. 90 Převod mezi prahovou hodnotou udávanou ve stupních a pixelech

Ve většině studií zaměřených na hodnocení kartografických děl údaj o nastavení identifikačních algoritmů chybí. Hermans a Laarni (2003), kteří se zabývali analýzou hledání konkrétního objektu v interaktivní mapě, použili prahovou hodnotu rozptylu  $2^\circ$ . Coltekin (2009) zkoumala efektivitu interaktivních map a prahovou hodnotu rozptylu nastavila na 50 px. Sama autorka však uvádí, že se toto nastavení neopírá o žádná předem stanovená pravidla. Stejné nastavení využila i ve své další studii (Coltekin a kol., 2010) a uvádí příklady dalších autorů, kteří použili stejné nastavení (např. Goldberg a Kotval, 1999).

Irwin (1992) se zabýval stanovením optimální prahové hodnoty délky fixace a zjistil, že 100 ms je vhodný a praktický kompromis mezi teoretickým minimem a maximem. Ke stejnému zjištění došli také Manor a Gordon (2003). Podle Holmqvista (2011) by se měla prahová hodnota délky fixace pohybovat v rozmezí 50–250 ms. Hodnota 100 ms je často využívána v kartografických studiích (Coltekin a kol., 2010; Coltekin a kol., 2009; Fabrikant a kol., 2010). Oproti tomu Hermans a Laarni (2003) nastavili prahovou hodnotu délky fixace pouze na 80 ms.

Stanovení prahových hodnot algoritmu I-DT se může lišit podle oblasti studia a konkrétní studie. Doporučuje se také, aby byly vytvořené fixace porovnány s původními naměřenými daty (Holmqvist a kol., 2011). Pro určení nejvhodnějšího nastavení algoritmu I-DT, který je při analýze eye-tracking dat na KGI UP nejčastěji využíván, byla srovnána čtyři nastavení algoritmu I-DT označená jako Fix 1–Fix 4.

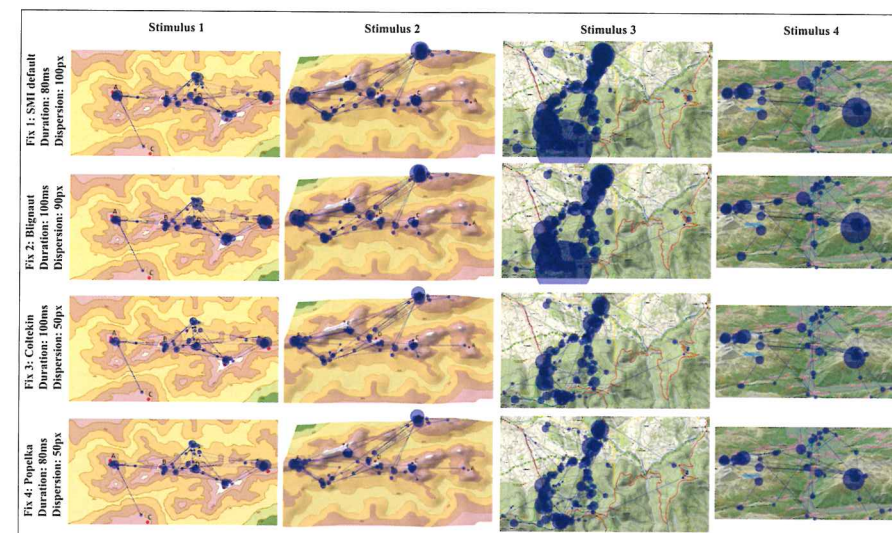
Fix 1 je základní nastavení používané programem SMI BeGaze. Nastavení Fix 2 použité ve studii Blignauta (2009) je velice podobné. Obě vytvářejí menší počet delších fixací. Takové nastavení je vhodné například pro úkoly, kdy uživatel nad stimulem neřeší žádný specifický úkol (free-viewing). Pro účely kartografie je však vhodnější identifikovat větší počet fixací s menší průměrnou délkou. Proto je nutné zvolit menší prahovou hodnotu rozptylu. Nastavení používané například ve studii Coltekin (2009), ale i dalšími

autory je označeno jako Fix 3. Nastavení Fix 4 je definováno autorem na základě porovnání identifikovaných fixací s původními naměřenými daty. Prahové hodnoty jednotlivých nastavení jsou uvedeny v tabulce 10.

Tab. 10 Srovnávaná nastavení identifikačního algoritmu v programu SMI BeGaze

	Fix 1	Fix 2	Fix 3	Fix 4
<b>Autor nastavení</b>	SMI Default	Blignaut (2009)	Coltekin (2009)	Popelka (2014b)
<b>Délka (ms)</b>	80	100	100	80
<b>Rozptyl (px)</b>	100	90	50	50

Rozdíly mezi Fix 1–Fix 2 a Fix 3–Fix 4 jsou jasně patrné z ukázky na obrázku 91. Kružnice v prvních dvou řádcích (Fix 1, Fix 2) jsou jasně větší než v řádcích následujících (Fix 3, Fix 4), protože délka fixací je větší.

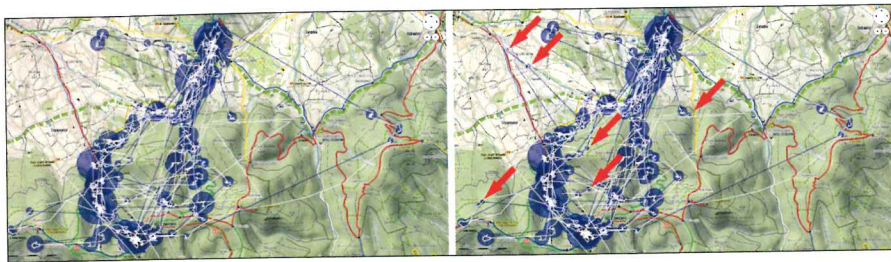


Obr. 91 Porovnání čtyř nastavení identifikace fixací pomocí I-DT algoritmu

Autorem vytvořené nastavení Fix 4 je modifikací nastavení využívaného v kartografických studiích (např. Coltekin a kol., 2009). Pro ověření korektní identifikace fixací doporučuje Holmqvist (2011) vizuální porovnání fixací s prvotními (raw) daty.

Protože nastavení Fix 3 a Fix 4 jsou si velmi podobná, identifikované fixace jsou téměř stejné. Při detailním pohledu je ale možné nalézt několik důležitých rozdílů. Výřez scanpath na obrázku 92 zobrazuje data jednoho uživatele

(PO4) na podkladu stimulu obsahujícího turistickou mapu. Úkolem respondentů v tomto experimentu bylo nalezení optimální trasy mezi dvěma body. Rozdílně identifikované fixace jsou na obrázku 92 označeny pomocí šipek. Při použití nastavení Fix 4 bylo identifikováno více fixací než při Fix 3. Při srovnání s původními naměřenými daty (bílá linie) je patrné, že se v těchto místech oko opravdu zastavilo (fixovalo).



Obr. 92 Porovnání rozdílů výsledku identifikačního algoritmu. Při použití nastavení Fix 4 (vpravo) bylo nalezeno více fixací než v případě nastavení Fix 3 (vlevo). Rozdíly jsou označeny pomocí šipek. Bílá linie reprezentuje původní naměřená data.

### 6.3.3 Nastavení detekce fixací v SMI BeGaze a OGAMA

Pro nastavení prahových hodnot identifikačního algoritmu v SMI BeGaze slouží nabídka File – Adjust Event Detection. Pokud zařízení umožňuje záznam s frekvencí alespoň 250 Hz, jsou přístupné nabídky jak pro Low Speed, tak pro High Speed (obr. 93).

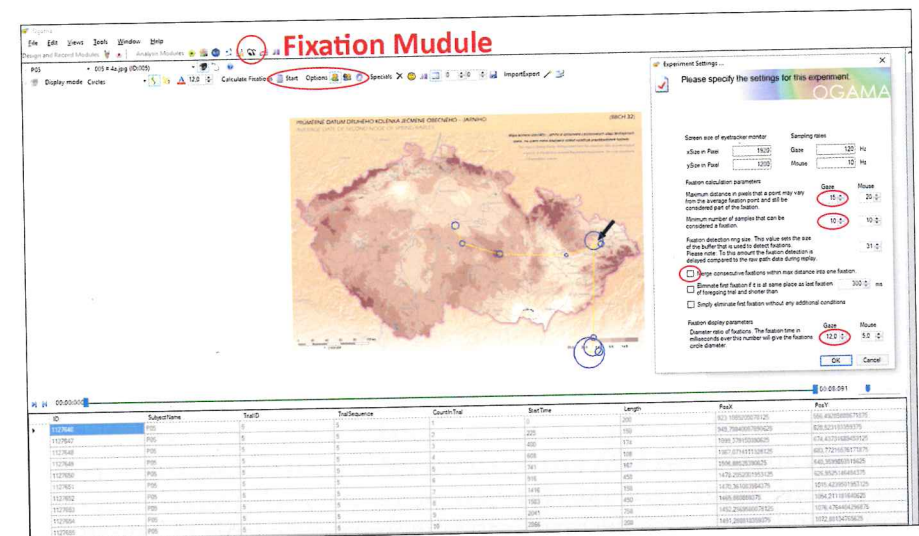


Obr. 93 Nastavení detekčního algoritmu v prostředí SMI BeGaze

Kromě prostého zadání prahové hodnoty délky (duration) a rozptylu (dispersion) umožňuje dialogové okno pouze vyřazení první fixace. V experimentech prováděných na KGI UP tato volba není používána, protože díky fixačnímu kříži všichni respondenti začínají prohlížet stimulus ze stejného místa na monitoru.

Jak bylo uvedeno v kapitole 6.1, do OGAMA lze importovat data naměřená například pomocí eye-trackeru SMI. OGAMA by rovněž měla umožňovat import fixací, avšak bylo zjištěno, že tato funkce nefunguje. Z toho důvodu byla importována pouze raw data a následně byly nastaveny parametry pro identifikaci fixací přímo v aplikaci OGAMA. OGAMA provádí detekci fixací pomocí algoritmu I-DT, což je jeden z důvodů, proč jej volíme i v SMI BeGaze.

V OGAMA se detekční algoritmus fixací nastavuje už při zakládání nového projektu, ale samozřejmě lze toto nastavení měnit i později. K tomu slouží Fixation Module (obr. 94). V okně modulu je nutné vybrat jednoho respondenta a konkrétní stimulus. Poté se zobrazí jeho fixace nad daným stimulem, a to jak graficky, tak jejich výčet v tabulce v dolní části modulu. Dialogové okno s nastavením detekčního algoritmu zobrazíme po stisku ikony ozubeného kola. Změna prahových hodnot se v datech projeví až po stisku tlačítka „Start“ v okně modulu. Pomocí ikon napravo od tohoto tlačítka můžeme zvolit, zda bude změna provedena pouze pro vybraného respondenta nebo pro všechny respondenty v experimentu. Tato funkce může být užitečná v případě, že hledáme ideální nastavení identifikačního algoritmu a nechceme čekat, až se fixace přepočítají pro všechny respondenty.



Obr. 94 Nastavení fixací v Fixation Module programu OGAMA

Nastavení parametrů algoritmu I-DT v programu OGAMA je od SMI BeGaze poměrně odlišné. Je nutné nastavit následující parametry:

- **Maximum distance** – maximální vzdálenost v pixelech, kterou od sebe mohou být body vzdálené, aby byly stále považovány jako součást fixace;
- **Minimum number of samples** – minimální počet bodů, které tvoří fixaci;
- **Merge** – sloučení po sobě jdoucích fixací, které jsou blíže, než je maximální vzdálenost;
- **Diameter ratio of fixation** – poloměr vykreslené kružnice bude zvětšen o rozdíl mezi délkou konkrétní fixace a touto hodnotou.

Klíčové jsou především první dva parametry. Rovněž je nutné dávat pozor, zda je správně nastavena hodnota frekvence záznamu dat. K problémům může docházet zejména v situaci, kdy jsou nejdříve nastaveny parametry pro identifikaci fixací a až poté je stanovena hodnota frekvence. V této situaci totiž dojde ke změně parametru Minimum number of samples.

Při hledání optimálního nastavení parametrů pro detekci fixací v OGAMA byl využit snímek scanpath z SMI BeGaze, vytvořený pomocí výše zmíněného nastavení I-DT algoritmu (80 ms, 50 px). Přes tento obrázek byly vykreslené fixace detekované programem OGAMA. Parametry byly modifikovány, dokud nebylo nalezeno takové nastavení, které nejlépe odpovídá fixacím identifikovaným pomocí SMI BeGaze (tab. 11).

Nejdůležitější parametry jsou „Maximum distance“ a „Minimum number of samples“, které odpovídají parametrům zadávaným do SMI BeGaze – rozptylu a délce fixace. Experimentálně bylo zjištěno, že optimální nastavení pro frekvenci 120 Hz je 15 px a 10 bodů. S použitím tohoto nastavení je výsledná trajektorie oka (scanpath) téměř shodná s trajektorií vytvořenou v aplikaci SMI BeGaze. Důležité je zmínit, že v tomto případě měla data frekvenci 120 Hz. V případě záznamu s frekvencí 250 Hz by měl parametr Maximum number of samples mít hodnotu 20. Pro 60 Hz by to potom byla hodnota 5.

Hodnota parametru „Diameter ratio of fixation“ ovlivní pouze grafické zobrazení fixací prostřednictvím scanpath. Se zobrazením scanpath z aplikace SMI BeGaze nejlépe korespondovala hodnota parametru 12, pouze u velmi dlouhých fixací docházelo k rozdílům, jak je patrné z obrázku 95. Zde je porovnán scanpath vzniklý v aplikaci SMI BeGaze (fialové kruhy) s fixacemi identifikovanými v aplikaci OGAMA (modré kružnice) pomocí výše uvedeného nastavení algoritmu. K jediné odchylce došlo v levé části obrázku vlevo, kdy aplikace OGAMA identifikovala jednu krátkou fixaci navíc oproti SMI BeGaze. Rozdíl v grafické vizualizaci velmi dlouhých fixací je znázorněn na obrázku vpravo. Obě nesrovnalosti je možné vyřešit změnou parametrů identifikačního algoritmu, ale tato změna by ovlivnila všechny ostatní fixace.



Obr. 95 Srovnání fixací identifikovaných na dvou různých stimulech programu SMI BeGaze (fialové kruhy) a OGAMA (modré kružnice). Rozdílně identifikované fixace jsou označeny červenou šipkou.

Závěrem kapitoly lze shrnout optimální nastavení detekce fixací pro SMI BeGaze i pro program OGAMA. Optimální hodnoty pro kartografický výzkum jsou uvedeny v tabulce 11.

Tab. 11 Optimální nastavení detekce fixací pro kartografický výzkum v programech SMI BeGaze a OGAMA

	Parametr algoritmu I-DT	Optimální nastavení
SMI BeGaze	Min. duration	80 ms
	Max. dispersion	50 px
OGAMA	Maximum distance	15
	Minimum number of samples	5 (60 Hz) 10 (120 Hz) 20 (250 Hz)
	Merge	NO
	Diameter ratio of fixation	12

## 6.4 Čištění dat

Před vlastní analýzou naměřených dat je nutné zkontrolovat, zda jsou naměřená data dostatečně kvalitní. Častým problémem může být (prostorový) posun dat způsobený nedostatečnou kalibrací nebo chybějící data (DataLoss).

### 6.4.1 DataLoss

DataLoss neboli ztráta dat může být způsobena ztrátou signálu (eye-tracker nevidí oko), ale také třeba prostým mrknutím. Z toho plyne, že téměř nikdy není dosaženo zaznamenání 100 % dat. Mnohdy se ale stane,

že je ztraceno větší množství dat, a analytik by měl takového respondenta z testování vyřadit, aby neovlivňoval výsledek celého experimentu.

V prostředí SMI BeGaze lze vyexportovat tzv. Participant Overview (sekce Specialized Statistics). Zde je pro jednotlivé respondenty vidět odchylka kalibrace ve směru os X a Y ve stupních (obr. 96). Součástí Participant Overview je rovněž tzv. Tracking Ratio, což je hodnota udávající, kolik procent z celkové doby trvání experimentu byl zaznamenáván pohyb očí. To znamená, pro kolik procent záznamů jsou v datech uloženy jiné hodnoty než 0;0. V SMI BeGaze bohužel nelze tuto statistiku zobrazit pro jednotlivé stimuly, ale pouze pro celý experiment.

Participant	Right Eye Deviation X [°]	Right Eye Deviation Y [°]	Tracking Ratio [%]
P01	0.4	0.2	97.6
P03	0.4	0.6	97.0
P04	0.3	0.4	96.7
P05	2.0	2.0	97.2
P06	0.7	0.4	98.8
P08	0.6	0.9	98.3
P09	0.3	0.4	92.5
P10	0.3	0.3	99.0

Obr. 96 Ukázka Participant Overview. Tabulka ukazuje odchylku kalibrace a Tracking Ratio pro jednotlivé respondenty.

Oproti tomu v aplikaci OGAMA je možné vyexportovat procento chybějících dat pro jednotlivé stimuly. Ve Statistics module je nutné zvolit Trial information – Additional parameters a zde je nabídka DataLoss. Vyexportovat jdou jak absolutní, tak i relativní data. Kromě počtu chybějících záznamů (tj. záznamů se souřadnicemi 0;0) lze zobrazit i počet záznamů naměřených mimo monitor (např. –200 px). Data jsou vždy vztažena k jednotlivému respondentovi a jednotlivému stimulu. Pro hlubší analýzu těchto dat je vhodné využít tabulkový procesor a pomocí kontingenční tabulky si zobrazit, u kterých stimulů a respondentů byla zaznamenána chybějící data. To lze udělat například tak, že v řádcích kontingenční tabulky budou zobrazeny jednotlivé stimuly, ve sloupcích respondenti a jako hodnota pole bude vybrán součet DataLoss (relativní hodnota). Protože pro každou kombinaci respondenta a stimulu je pouze jedna hodnota, bude tento „součet“ odpovídat původním vyexportovaným datům. Výslednou kontingenční tabulku lze upravit tak, jak je znázorněno na obrázku 97. Pomocí funkce podmíněné formátování byly označeny ty buňky, kde byla ztráta dat vyšší než 10 %. Z tabulky je vidět, že u respondentů a618, a624 a a626 byly vysoké hodnoty DataLoss u velkého množství stimulů. Tato konkrétní data byla zaznamenávána pomocí zařízení EyeTribe, takže úplně stoprocentní data se nedala očekávat. Přesto by bylo asi vhodnější tyto respondenty z následné analýzy

vyřadit, a to i přesto, že celková průměrná hodnota DataLoss za všechny stimuly je nižší než 100 %. Velmi důležitá jsou pole označena oranžovou barvou, kde OGAMA nevyexportovala žádné hodnoty. To je způsobeno tím, že byl proces nahrávání nečekaně přerušeno a nebyla zaznamenána vůbec žádná data. Z toho plyne, že respondenta a606 je nutné z analýzy rovněž vypustit.

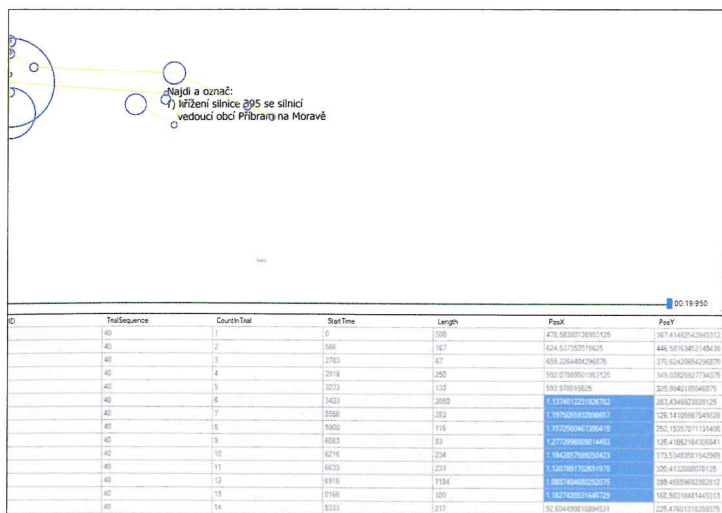
Data Loss (%)	a600	a602	a604	a606	a608	a616	a618	a620	a622	a624	a626	Průměr
uloha1_1.jpg	3,92	0,85	0,00	2,23	0,34	6,44	29,51	3,24	1,89	7,86	4,51	5,52
uloha1_2.jpg	3,44	0,77	1,53	2,50	1,07	2,52	17,06	4,29	1,46	13,44	5,57	4,88
uloha1_3.jpg	3,93	1,90	0,50	1,88	1,58	0,00	13,20	5,15	1,97	24,48	16,21	6,44
uloha1_4.jpg	5,81	6,65	1,16	5,09	1,04	1,08	10,48	1,64	2,23	10,19	8,73	4,92
uloha1_bod1.jpg	3,07	1,08	0,34	0,00	0,00	2,85	0,00	4,66	3,23	6,25	14,35	3,26
uloha1_bod2.jpg	8,51	8,29	1,01	2,19	0,34	1,85	0,20	3,45	1,55	11,92	16,38	5,06
uloha1_linie2.jpg	2,76	0,76	0,00	1,77	0,14	2,98	0,34	2,66	0,81	8,75	8,25	2,66
uloha2_1.jpg	4,80	0,00	1,19		0,39	0,66	18,94	1,98	3,21	8,75	8,37	4,83
uloha2_2.jpg	1,62	7,43	2,10		1,38	0,00	8,63	3,01	0,00	1,83	12,35	3,84
uloha2_3.jpg	5,96	2,63	0,00		0,00	1,46	15,71	0,00	2,12	3,83	4,37	3,61
uloha2_4.jpg	7,21	6,98	1,22		0,73	0,63	22,86	4,71	1,80	3,60	17,14	6,69
uloha2_bod1.jpg	6,20	0,69	0,00		0,76	35,76	0,40	4,24	2,02	15,22	12,18	7,75
uloha2_bod2.jpg	4,01	0,36	0,00		0,00	0,00	0,33	3,33	1,73	2,63	8,63	2,10
uloha2_linie2.jpg	2,43	3,52	0,38		0,00	0,97	0,79	2,39	0,40	1,78	0,78	1,34
uloha3_1.jpg	5,36	5,85	0,00		0,00	0,35	8,06	4,71	1,62	19,98	9,67	5,56
uloha3_2.jpg	1,23	0,63	1,52		0,00	0,00	5,73	4,86	2,78	3,66	5,98	2,64
uloha3_3.jpg	2,90	4,87	0,00		1,69	1,87	2,20	0,00	2,04	8,13	13,12	3,68
uloha3_4.jpg	2,44	28,92	1,65		1,37	1,25	14,43	5,50	1,72	23,95	8,48	8,97
uloha3_bod1.jpg	0,00	13,78	1,31		0,60	1,71	0,00	3,20	2,90	11,15	8,57	4,32
uloha3_bod2.jpg	17,80	3,53	0,00		0,26	1,10	0,00	3,39	1,46	4,52	20,75	5,28
uloha3_linie2.jpg	2,28	8,17	0,00		0,12	1,81	0,09	2,78	1,44	8,94	2,19	2,78
Průměr	4,56	5,13	0,66	2,24	0,56	3,11	8,04	3,29	1,83	9,56	9,84	4,44

Obr. 97 Upravená kontingenční tabulka ukazující DataLoss pro 11 respondentů

Je na analytikovi, jak si nastaví hraniční hodnotu a která data do analýzy použije a která nikoliv. Toto rozhodnutí záleží především na použitém zařízení a také na typu úkolu. Pokud měl například respondent během experimentu za úkol něco hledat v zadání na papíře, dá se předpokládat že DataLoss bude u těchto experimentů vyšší (protože musel hledat v zadání, a tím pádem se nedíval na monitor a souřadnice jeho pohledu tedy byly 0;0).

Ne všechny chyby v datech je lze takto jednoduše odhalit. Příkladem může být situace na obrázku 98. DataLoss v tomto případě nepřesahovalo ani 7 %, přesto však tato data nelze použít. Polovina ze zaznamenaných fixací se nachází na levém okraji stimulu. Tento typ chyby je typický pro záznam pomocí EyeTribe trackeru. Pokud je toto zařízení použito, je nutné věnovat hlubší pozornost kontrole dat a stimuly či respondenty, u kterých se tyto chyby vyskytují, z výsledného hodnocení vyřadit.

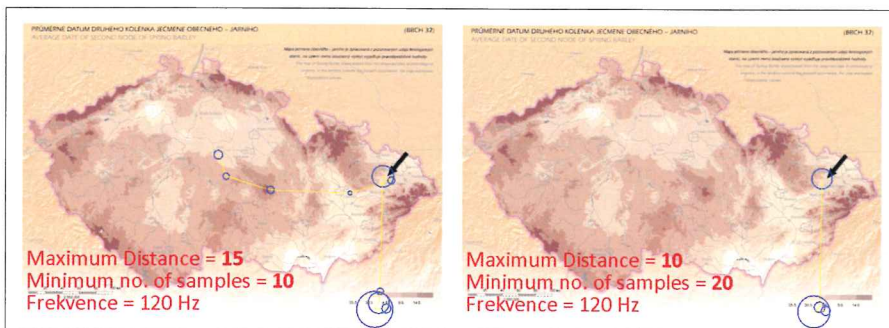




Obr. 98 Chyba v datech, kterou je nutné odhalit vizuální inspekci.

### 6.4.2 Chyby způsobené nevhodným nastavením detekce fixací

Dalším častým zdrojem chyb při analýze eye-tracking dat je nevhodné nastavení prahových hodnot detekčního algoritmu. Optimální nastavení identifikace fixací jak v SMI BeGaze, tak v OGAMA bylo popsáno v kapitole 6.3.3. Korektní nastavení identifikace fixací je velmi důležité, protože valná většina metod analýzy pracuje právě s fixacemi. Na obrázku 99 je znázorněno, jak změna dvou nejdůležitějších parametrů identifikace fixací v OGAMA ovlivní výsledek. V levé části obrázku je korektní nastavení algoritmu tak, jak bylo popsáno v kapitole 6.3.3. V pravé části je upravené nastavení. Je jasně patrné, že tato změna způsobila, že z původních deseti fixací zůstaly pouze čtyři.

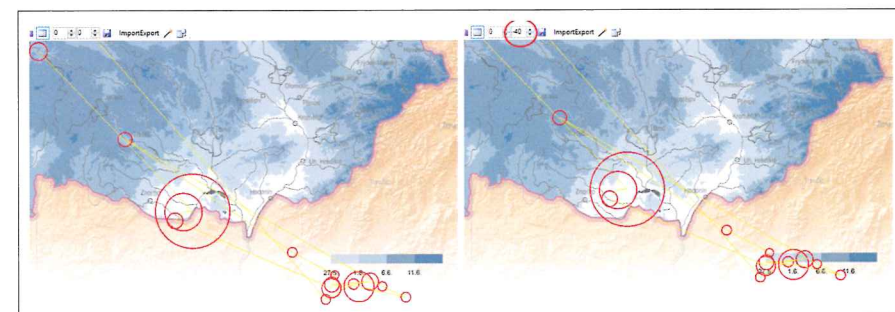


Obr. 99 Porovnání korektního (vlevo) a nesprávného (vpravo) nastavení identifikace fixací v programu OGAMA

K chybám velmi často dochází v případě, že uživatel nesprávně nastaví frekvenci záznamu dat a poté ponechá defaultní prahové hodnoty identifikačního algoritmu. To se stává zejména při použití EyeTribe trackeru, kdy si uživatel myslí, že zaznamenával data s frekvencí 60 Hz, ale zapomněl změnit nastavení a ve skutečnosti byla data zaznamenávána pouze s frekvencí 30 Hz. Pokud v takovémto případě ponechá defaultní nastavení detekce fixací, může se stát, že dosáhne průměrného počtu fixací na stimulus menší než 1. Z takových dat poté samozřejmě nelze vyvozovat žádné závěry.

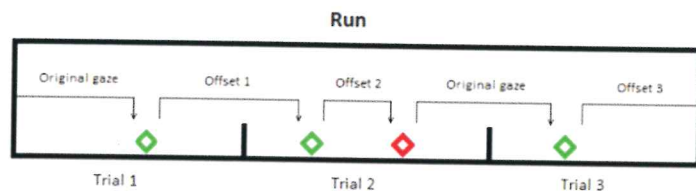
### 6.4.3 Odchylka dat

V kapitole 6.4 bylo uvedeno, že v prostředí SMI BeGaze lze vyexportovat odchylku kalibrace. Ta by měla být co nejmenší. Většinou se uvádí méně než 0,5° či 1°, ale vždy záleží na charakteru stimulů a účelu testování. Pokud budou na stimulu zaznačovány malé oblasti zájmu, je nutné mít odchylku kalibrace co nejmenší. Ve všech případech je ale vhodné vizuálně zkontrolovat, zda scanpath odpovídá stimulu. Například zda je při čtení textu zaznamenána trajektorie pohybu očí na dané větě nebo jestli není posunuta. Tato kontrola by měla probíhat i během vlastního eye-tracking testování, aby bylo možné například upravit pozici respondenta před monitorem. Pokud jsou naměřená data posunuta, je možné tento nedostatek opravit i zpětně posunutím celého scanpath. Tato možnost je dostupná v programu OGAMA, nicméně je nutné opravit data pro každého respondenta a každý obrázek, tudíž je tato operace velmi časově náročná a navíc je do dat vnášena určitá nejistota, protože záleží na analytikovi, o kolik px data posune. Ukázka takového posunu je vidět na obrázku 100. V levé části jsou zobrazena data jednoho respondenta, která jsou od reality posunuta směrem dolů (respondent se díval na legendu, ale data jsou níž). V pravé části obrázku došlo k posunu dat o 40 px směrem nahoru. Důležité je každou změnu v databázi OGAMA uložit. Tato operace ale většinou trvá dlouho a OGAMA během ní bohužel někdy nečekaně spadne.



Obr. 100 Data zaznamenaná s posunem směrem dolů (vlevo) a stejná data po korekci v programu OGAMA (vpravo)

Obdobně lze fixace posunout i v SMI BeGaze. Výhodou je, že lze posunout fixace za určitý časový úsek, tedy ne za celou dobu sledování stimulu. Nevýhodou je to, že je celý proces o něco složitější. Pokud výzkumník zjistí, že od určitého okamžiku došlo k posunu zaznamenaných eye-tracking dat, což se může stát například díky změně pozice respondenta, může data opravit pomocí nástroje Offset Correction. Ten je dostupný v nástrojích Custom Trial Selector, Gaze Replay, Scan Path a Bee Swarm. Výzkumník neprve umístí posuvník časové osy do času, kdy došlo k posunu. Následně v okně stimulu stiskne pravé tlačítko myši a zvolí možnost Offset Correction. Poté drží levé tlačítko a pomocí tahu myši posune kurzor na místo, kde má být další fixace. Po uvolnění dojde k posunu všech dalších fixací. Po tomto posunu se v menu zobrazeném po stisku pravého tlačítka objeví možnost Reset Offset Correction, která vrátí fixace na původně naměřené pozice, ovšem až od času, ve kterém byla tato možnost zvolena. V SMI BeGaze je možné tyto nástroje různě kombinovat a řetězit tak, jak je znázorněno na obrázku 101.



Obr. 101 Řetězení nástrojů Offset Correction a Reset Offset Correction v prostředí SMI BeGaze (SMI, 2008)

Velice důležité je pamatovat na to, že posun fixací v SMI BeGaze není omezen pouze na jeden stimulus, ale dojde k posunu všech dat od zadaného času až do konce experimentu.

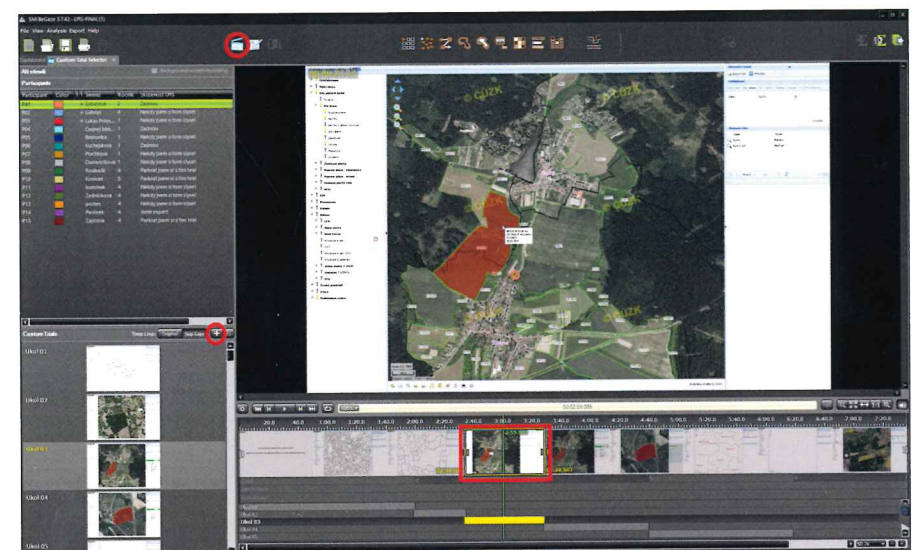
## 6.5 Segmentace záznamu obrazovky

Posledním typem předzpracování eye-tracking dat je takzvaná segmentace záznamu obrazovky. V aplikaci SMI BeGaze se tato funkce jmenuje Custom Trial Selector a jejím účelem je rozdělení stimulu do více trialů či úseků. Typicky se tato funkce používá v případě, že je stimulem záznam obrazovky. V modelovém případě byla analyzována mapová aplikace, nad kterou respondent řešil pět různých úkolů. Při zobrazení naměřených dat v SMI BeGaze je v seznamu stimulů vidět pro každého respondenta videozáznam pro každý úkol zvlášť (obr. 102). Není tedy možné porovnat chování několika respondentů během jednoho úkolu.

Stimuli		
P01-scrrec.mkv	1920 x 1200	<input checked="" type="checkbox"/>
P01-scrrec1.mkv	1920 x 1200	<input type="checkbox"/>
P01-scrrec2.mkv	1920 x 1200	<input type="checkbox"/>
P01-scrrec3.mkv	1920 x 1200	<input type="checkbox"/>
P01-scrrec4.mkv	1920 x 1200	<input type="checkbox"/>
P01-scrrec5.mkv	1920 x 1200	<input type="checkbox"/>
P02-scrrec.mkv	1920 x 1200	<input type="checkbox"/>
P02-scrrec1.mkv	1920 x 1200	<input type="checkbox"/>
P02-scrrec2.mkv	1920 x 1200	<input type="checkbox"/>

Obr. 102 Zobrazení stimulů při záznamu obrazovky v prostředí SMI BeGaze

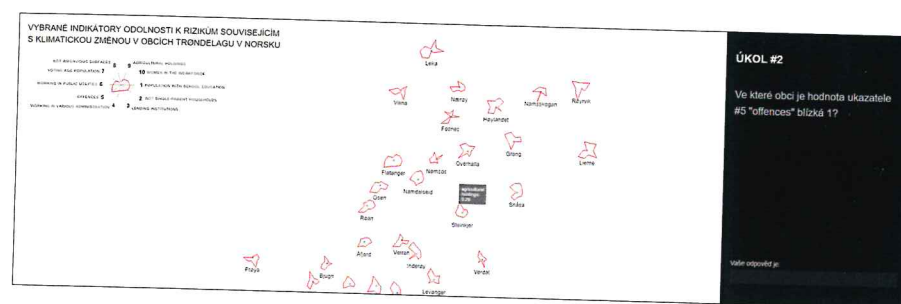
Pomocí Custom Trial Selector lze videozáznam rozdělit na libovolný počet úseků, například na počet odpovídající jednotlivým úkolům. Funkce Custom Trial Selector je dostupná prostřednictvím ikony v horní oblasti prostředí SMI BeGaze (obr. 103 nahoře). Nejprve je nutné vytvořit screenshot, nad kterým budou separovaná data zobrazována. V časové ose v dolní části obrazovky si uživatel zobrazí požadovaný moment a poté pomocí tlačítka „+“ v levé části obrazovky vytvoří screenshot neboli Custom Trial. Ten si může libovolně pojmenovat (např. úkol 1). Následně na tento Custom Trial klikne a v časové ose označí odpovídající úsek videa. Pomocí žlutého pruhu v dolní části obrazovky se zvýrazní, že daná část záznamu odpovídá tomuto konkrétnímu trialu.



Obr. 103 Použití funkce Custom Trial Selector v prostředí SMI BeGaze

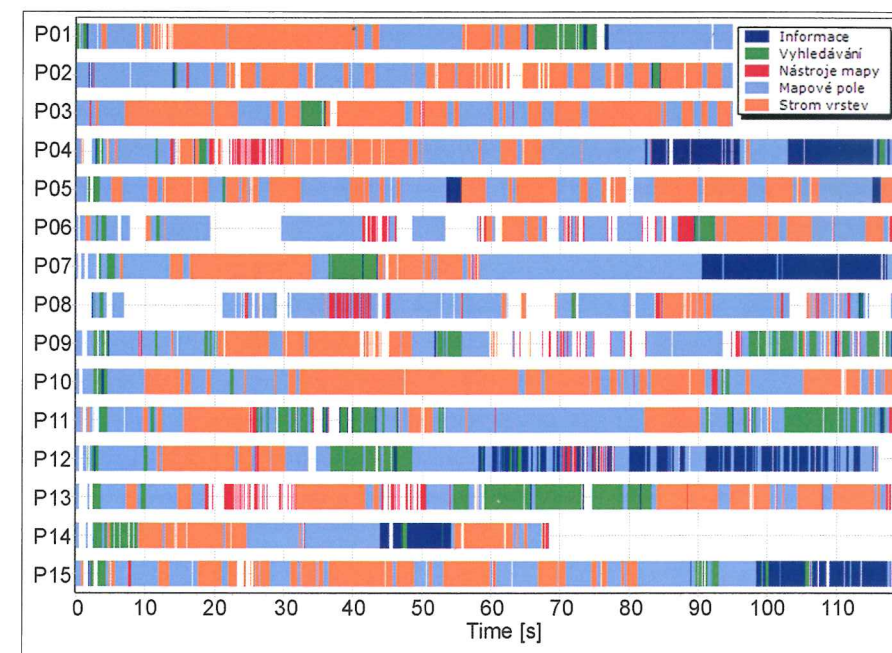
Tento postup je nutné zopakovat pro všechny respondenty, takže tato operace může být časově velmi náročná. Nicméně výsledkem je přiřazení záznamu všech respondentů k jednotlivým screenshotům, které se následně jeví tak, jako by se respondenti dívali na stejný obrázek. Při následném zpracování je nutné pamatovat na to, že vytvořený screenshot nemusí odpovídat tomu, co na obrazovce respondent skutečně viděl.

Využití Custom Trial Selector je tedy vhodné především v těch případech, kdy je obraz víceméně statický, ale z nějakého důvodu bylo využito záznamu obrazovky. K tomu může dojít například v případě, že stimulus obsahuje nějaké interaktivní prvky. Příkladem může být studie Opacha a kol. (2017), která obsahovala množství úkolů nad stejnými mapami obsahujícími glyphy. Při najetí myši nad jednotlivé glyphy byly interaktivně zobrazeny určité informace (obr. 104) a navíc stránka obsahovala pole pro zadávání odpovědi. Z toho důvodu nemohly být jako stimuly použity obrázky.



Obr. 104 Ukázka prostředí obsahujícího interaktivní prvky

Custom Trial Selector lze využít i v případě, kdy se obsah mění velmi výrazně. Příkladem této situace může být analýza mapové aplikace popsána v modelovém příkladě na začátku této kapitoly. Přestože respondenti mohou s mapou pohybovat atd., základní rozdělení obrazovky na jednotlivé části, jako je mapové pole či strom vrstev, je vždy stejné. Po použití Custom Trial Selectoru je tedy možné analyzovat čas či počet fixací zaznamenaný v oblastech zájmu odpovídajících jednotlivým segmentům mapové aplikace. Výsledkem tedy může být například Sequence Chart, který ukazuje, do kterých částí aplikace se respondenti během úvodního seznámení s prostředím dívali (obr. 105).



Obr. 105 Sequence Chart vytvořený z dat záznamu obrazovky segmentovaných pomocí Custom Trial Selectoru