

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A
INFORMATIKY

METÓDY VYHODNOCOVANIA
PSYCHOFYZIKÁLNYCH
EXPERIMENTOV

Bakalárska práca

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A
INFORMATIKY

METÓDY VYHODNOCOVANIA
PSYCHOFYZIKÁLNYCH
EXPERIMENTOV

Bakalárska práca

Študijný program: Aplikovaná informatika
Študijný odbor: 2511 aplikovaná informatika
Školiace pracovisko: Katedra Aplikovanej Informatiky
Školiteľ: RNDr. Elena Šikudová, PhD.



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Paula Budzáková
Študijný program: aplikovaná informatika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: 9.2.9. aplikovaná informatika
Typ záverečnej práce: bakalárska
Jazyk záverečnej práce: slovenský

Názov: Metódy vyhodnocovania psychofyzikálnych experimentov / *Psychophysical experiments' evaluation methods*

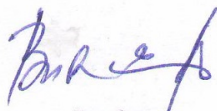
Cieľ: Prehľad štatistických metód na vyhodnocovanie psychofyzikálnych experimentov. Naprogramovanie SW na vyhodnocovanie dát získaných z experimentov.

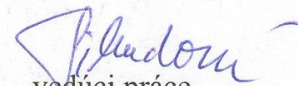
Anotácia: Psychofyzikálne experimenty sa používajú na subjektívne porovnanie výsledkov rôznych metód. Subjekty zadávajú svoje preferencie, ktoré sa potom vyhodnocujú.

Vedúci: RNDr. Elena Šikudová, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. PhDr. Ján Rybár, PhD.
Dátum zadania: 23.09.2013

Dátum schválenia: 23.10.2013

doc. RNDr. Damas Gruska, PhD.
garant študijného programu


študent


vedúci práce

Čestné vyhlásenie

Čestne prehlasujem, že som túto bakalársku prácu vypracovala samostatne s použitím citovaných zdrojov.

V Bratislave

.....

Pod'akovanie

Chcela by som sa pod'akovať mojej školiteľke, RNDr. Elene Šikudovej, PhD, za pomoc a rady pri výbere a spracovávaní štatistických metód a za obetovaný čas pri vylepšovaní celej bakalárskej práce.

Taktiež ďakujem mojej Marienke za jazykovú korektúru a hlavne za rodičovskú podporu. Takisto ďakujem Rozálke za pomoc pri prekladoch článkov a Jurkovi za to, že to so mnou zvládol.

Abstrakt

Pomocou štatistických metód dokážeme vyhodnotiť aj experimenty z oblasti psychofyziky. Je veľmi potrebné pri výbere metód uvedomiť si, ako náš experiment prebieha, čo probanti hodnotia a aké dáta budeme mať k dispozícii. Je veľmi dôležité správne vybrať metódy. Nesprávna metóda by nielen nepresne vyhodnotila výsledky, no mohla by byť aj úplne nefunkčná. V práci je popísaných šesť štatistických metód, ktoré sme použili pri spracovaní obrazu pomocou psychofyzikálnych experimentov. V experimentoch boli použité nepriame metódy škálovania, a to Hodnotenie (rating), Usporiadanie (ranking) a Párové porovnanie (pair comparison).

Kľúčové slová: *psychofyzikálny experiment, Analýza rozptylu, Kruskal-Wallisov test, Kendallov koeficient zhody W*

Abstract

Using statistical methods, we can also evaluate the area of psychophysics experiments. When we are choosing the methods it is very important to realize how our experiment takes place, what participant of experiment evaluated and what data will be available for us. The process of choosing correct method is very significant. The incorrect method would not only assess the results inaccurately, but could be also completely non-functional. Paper described six statistical methods that we used for image processing by using psychophysical experiments. In the experiments were used these indirect methods of scaling: Evaluation (rating), disposing (ranking) and pairwise comparisons (pair comparison).

Keywords: *psychophysical experiments, Analysis of variance, Kruskal-Wallis, Kendall's W*

Obsah

1 Úvod	12
2 Východiská práce	13
2.1 Psychofyzika	13
2.1.1 Gustáv Fechner (1801-1887)	13
2.1.2 Erns Weber (1795 – 1878)	14
2.2 Psychofyzikálne kvantity	14
2.2.1 Psychofyzikálne prahy	14
2.2.2 Prahový problém a prahová koncepcia	14
2.2.3 Psychofyzikálne prahové metódy	15
2.2.3.1 Klasické experimentálne metódy	15
2.2.3.2 Súčasné experimentálne metódy	17
2.2.3.3 Priame metódy psychofyziky	18
2.2.4 Psychofyzikálne škálovanie	20
2.2.5 Psychofyzikálne metódy škálovania	21
2.2.5.1 Priame metódy škálovania	21
2.2.5.2 Nepriame metódy škálovania	22
2.3 Implementácia	24
2.3.1 Matlab	24
3 Metódy vyhodnocovania	25
3.1 Usporiadanie (ranking)	25
3.1.1 Psychofyzikálny experiment	25
3.1.1.1 Testy štatistických hypotéz, P-hodnota	26
3.1.1.2 Jednofaktorová analýza rozptylu (one-way ANOVA)	28
3.1.1.3 Výsledky	30
3.2 Párové porovnanie (pair comparison)	31
3.2.1 Psychofyzikálny experiment 1	31
3.2.1.1 Údaje párového porovnávanía	31

3.2.1.2	Interval spoľahlivosti	32
3.2.1.3	Chi-kvadrát test dobrej zhody	33
3.2.1.4	Výsledky	34
3.2.2	Psychofyzikálny experiment 2	36
3.2.2.1	Kruskalov-Wallisov test	37
3.2.2.2	Výsledky	38
3.3	Hodnotenie (rating)	39
3.3.1	Psychofyzikálny experiment 1	39
3.3.1.1	Kendallov koeficient zhody W	40
3.3.1.2	Výsledky	40
4	Implementácia	42
4.1	Psycho-Stats	42
4.1.1	Usporiadanie (ranking)	42
4.1.1.1	Vstup	43
4.1.1.2	Výstup	43
4.1.2	Párové porovnanie (Pair comparison)	45
4.1.2.1	Vstup	45
4.1.2.2	Výstup	45
4.1.3	Hodnotenie (Rating)	47
4.1.3.1	Vstup	47
4.1.3.2	Výstup	48
4.1.4	Budúca práca	49
5	Záver	50
	Literatúra	52

Zoznam obrázkov

2.1	Hodnotenie	22
2.2	Párové porovnanie	23
2.3	Usporiadanie	23
3.1	Stredné usporiadanie z experimentu. Hdr predstavuje fyzikálne správny HDR obraz, bila, phot, hist reprezentujú použité tone mappingové metódy. Obje a subj reprezentujú najlepšie objektívne a subjektívne obrázky. Vľavo sú rebríčky jednotlivých scén. Vpravo je súhrnný výsledok všetkých scén. Spodný riadok ukazuje vyššie preferencie. Chybové úsečky označuje smerodajná odchylka ± 1 [2].	30
3.2	Podobnosť skupín v experimente. Položky v rovnakom súbore sú štatisticky nerozoznatel'né [2].	30
3.3	Interval škálových hodnôt experimentu, ktoré obsahovali buď zmenu (a) pozadia alebo (b) orámovanie medzi originálom a reprodukciou. Chybové úsečky predstavujú 95 % interval spoľahlivosti okolo priemernej hodnoty [10].	35
3.4	Porovnanie výsledkov hodnôt škály získaných použitím párového porovnania, kde 1 je výborný, 2 - dobrý, 3 - prijateľný, 4 - neprijateľný, 5 - zlý, 6 - hrozný. Zahŕňa údaje všetkých porovnaní v tomto psychofyzikálnom teste [10].	35
3.5	Vzdialenosti medzi načítanými obrázkami [13]	38
3.6	Vyvolané obázky na psychofyzikálnej škále [13]	38
4.1	Úvodná obrazovka softvéru Psycho-Stats.	42
4.2	Vstupné dáta pre experiment Hodnotenie.	43
4.3	Štatistické spracovanie experimentu Usporiadanie.	44
4.4	Vstupné dáta pre experiment Párové porovnanie.	45
4.5	Štatistické spracovanie dát experimentu Párové porovnanie.	46
4.6	Viacnásobný porovnávací test bez odlišných skupín.	46
4.7	Viacnásobný porovnávací test s odlišnými skupinami.	47
4.8	Vstupné dáta pre experiment Hodnotenie.	48

4.9	Štatistické spracovanie dát experimentu Hodnotenie.	48
-----	---	----

1. Úvod

Každý z nás sa určite vo svojom živote stretol so situáciou, v ktorej potreboval niečo ohodnotiť, zmerať, odvážiť. V podstate skoro každá veda disponuje so spoľahlivými metódami hodnotenia. Vo fyzike máme štandardné postupy hodnotenia fyzikálnych javov, ktoré ovláda už dieťa na základnej škole. Chémia nám zase približuje postupy hodnotenia energie rôznych reakcií. Všetky štandardné postupy, v spomínaných vedách, dokážu zhodnotiť namerané výsledky priamo a jasne. Merať ľudské pocity a vnímanie je už o čosi komplikovanejšie. Vizuálne skúsenosti človeka sú subjektívne a ťažko jednoznačne zmerateľné. Jednou z možností je využiť počítačovú grafiku. Pomocou rôznych obrazov, obrázkov, zvukov a animácií vieme odmerať ľudské vnímanie.

Cieľom bakalárskej práce nie je podrobné popísanie samotných psychofyzikálnych experimentov, ale ich štatistické vyhodnotenie. Hlavnou časťou je priblíženie jednotlivých vyhodnocovacích metód a ich následná implementácia v praxi.

V prvej kapitole bakalárskej práce si priblížime psychofyziku ako vedu, psychofyzikálne prahy a prahové metódy. Ďalej si priblížime psychofyzikálne škálovanie, metódy škálovania, od ktorých záviseli naše psychofyzikálne experimenty.

V druhej kapitole sú jednotlivé typy experimentov priblížené prostredníctvom reálne vykonaných experimentov. Pri každom type experimentu sme popísali a vysvetlili metódy, ktorými je možné tento experiment vyhodnotiť.

Tretia kapitola približuje reálnu implementáciu jednotlivých vybraných vyhodnocovacích metód vo vývojovom prostredí Matlab, s využitím štatistického toolboxu. Opisujeme v nej vytvorený štatistický softvér Psycho-Stats a jeho používanie. Taktiež popisujeme jednotlivé štatistické vyhodnotenia pomocou tohto softvéru. V závere tretej kapitoly predstavujeme návrh riešenia do budúcnosti pre vylepšenie jedného z experimentov.

2. Východiská práce

Psychofyzika je exaktná veda, ktorá meria citlivosť zmyslov. Je odborom experimentálnej psychológie. Zaoberá sa rozvojom experimentálnych metód pre objektívne meranie ľudskej subjektívnej vnemovej skúsenosti. A tu nastáva otázka, či je to jednoduché a nie je to problém. Objekty v živote majú určité fyzikálne vlastnosti, ako dĺžka, intenzita. Všetky fyzikálne vlastnosti dokážu byť priamo merané nejakým nástrojom alebo človekom. Každý objekt vytvára pre človeka rozdielne vnemové dojmy. Či je objekt svetlejší ako ten vedľa neho, či je hrubší alebo tenší. Nie je možné napojiť sa priamo na ľudský mozog a odmerať vnemové city pomocou nejakého nástroja. Jedinou možnosťou je odvodiť si výsledky, čo ľudia vnímajú z ich odpovedí a pocitov. Psychofyzika nám poskytuje sadu nástrojov, ktoré umožňujú objektívne merať toto ľudské vnímanie. Psychofyzika jednoducho kvantifikuje vzťah medzi fyzickou simuláciou vonkajšieho sveta a výstupným vnímaním sveta vnútorného. Experimentálne psychofyzikálne metódy využíva v psychofyzikálnych experimentoch.

2.1 Psychofyzika

2.1.1 Gustáv Fechner (1801-1887)

Zakladateľom psychofyziky bol nemecký vedec **Gustáv Fechner** [1], ktorý napísal základné dielo „Elementy psychofyziky“. Fechner chápal psychometriku ako funkčný vzťah medzi vnútorným svetom vnímania človeka a vonkajším fyzickým svetom. Pri riešení tohto problému sa najviac zamerával na experimentálne najdostupnejšie oblasti vzťahov medzi subjektívnym a objektívnym, čiže na meranie pocitov. Zaviedol základné pojmy psychofyziky, sformuloval psychofyzický zákon a vypracoval psychofyzikálne metódy. Fechner veľmi podrobne prepracoval problematiku prahov - upresnil pojem prah, navrhol experimentálne metódy na ich priame meranie.

Pojem prah sa vo vede využíval na označenie hranice, nad ktorou nastáva u človeka pocit a pod ktorou pocit nevzniká. Psychologický prah podľa Fechnera je prah medzi pocitom a vzruchom. Ak je prah akási hranica, hodnota deliaca podnety na vyvolávajúce pocity a na tie, pri ktorých nič nepocitujeme, potom môžeme prah skúmať ako bod vzniku pocitu.

2.1.2 Erns Weber (1795 – 1878)

Nemecký fyzik, anatóm a psychológ **Erns Weber** [1] sa taktiež venoval problematike psychofyziky. Weber skúmal vzťah medzi podnetom a jeho prírastkom, prírastkom intenzity, ktorý zodpovedá najmenšiemu pozorovanému rozdielu. Meral najmenšiu zmenu vo veľkosti podnetu, ktorý môže účastník experimentu (ďalej len probant) detekovať.

2.2 Psychofyzikálne kvantify

2.2.1 Psychofyzikálne prahy

Problém merania prahu je spojený s problémom merania senzorickej citlivosti. Pojem prah ako prvý použil roku 1824 Herbart, ktorý prah definoval ako pomyselnú hranicu medzi časťou poznania a časťou ešte nepoznaného.

Psychofyzikálne prahy [5] môžeme deliť na:

1. **Absolútny prah** - Absolútny prah môžeme definovať ako minimálnu veľkosť podnetu alebo najnižšiu intenzitu podnetu, ktorá sa dá jasne odlíšiť od neprítomnosti podnetu. Napríklad veľmi slabé svetlo, ktoré sa dá jednoducho rozlíšiť od tmy.
2. **Dolný pocitový prah** - Najnižšia intenzita podnetu, ktorá vyvoláva pocit. Nižšie intenzity podnetu už nepocitujeme. Nazývame ich podprahové intenzity.
3. **Horný pocitový prah** - Najvyššia intenzita podnetu, ktoré ešte vyvoláva pocit. Vyššie intenzity podnetu už nepocitujeme príslušným receptorom a môžu vyvolávať bolesť. Nazývame ich nadprahové intenzity.
4. **Rozdielový (diferenciálny) prah** - Diferenciálny prah definoval Fechner ako minimálny rozdiel medzi podnetmi, nad ktorými človek pocituje rozdiely a pod ktorými sa tieto podnety zdajú rovnaké.

2.2.2 Prahový problém a prahová koncepcia

S prahovým problémom je priamo spojený problém merania senzorickej citlivosti. Čo je vlastne citlivosť? Citlivosť nie je nič iné ako obrátená hodnota prahu. Najbežnejším spôsobom odhadu citlivosti nejakého zmyslu je určenie minimálnej veľkosti podnetu. Práve táto minimálna hodnota je označovaná ako **absolútny prah**.

2.2.3 Psychofyzikálne prahové metódy

Fechner predstavil tri klasické experimentálne metódy [5] pre meranie minimálneho podnetu:

- *metóda priemernej chyby* (method of adjustment)
- *metóda minimálnych zmien* (method of limit)
- *metóda konštantných podnetov* (method of constant stimuli)

Bardin zase predstavil súčasné metódy [5] pre meranie minimálneho podnetu:

- *metóda „áno a nie“*
- *hodnotiaca metóda*
- *metóda nútenej voľby*

2.2.3.1 Klasické experimentálne metódy

Metóda priemernej chyby

Táto metóda sa môže využívať aj na meranie diferenciálneho prahu. Metóda priemernej chyby je zvláštna práve tým, že zmena veľkosti premenného podnetu nie je navodená experimentátorom, ale samotným probantom. Postup metódy je nasledovný: Probantovi sa podáva podnet, ktorý slúži ako štandard, zadáva sa úloha a zobrazujú sa testovacie sady premenných podnetov. Probantovou úlohou je pri zmene premenného podnetu nastaviť takú hodnotu, pri ktorej budú oba podnety rovnaké - hodnota premenného podnetu bude rovnaká ako hodnota štandardu. Pri tejto metóde probant neinformuje experimentátora o svojich pocitoch a dojmoch, ale zmenšovaním alebo zväčšovaním mení hodnotu premenného podnetu a snaží sa nastaviť ju na hodnotu štandardu. Hodnotu, pri ktorej probant zastavuje svoj výber a určil ju za rovnakú alebo veľmi podobnú štandardu, možno považovať za bod lokalizácie prahu.

Metóda minimálnych zmien

Určenie absolútnej citlivosti

Metóda minimálnych zmien určuje prahové hodnoty. Prah, ktorý vyvoláva pocit, sa dá nájsť postupne po krokoch zmenšovaním hodnoty, ktorá prah prevyšuje. Po každom kroku je tento podnet slabší a slabší, až sa dostaneme do štádia, kedy podnet prestáva vyvolávať pocit. Taktiež sa dá prah nájsť aj opačným spôsobom. Za výhodiskovú hodnotu môžeme vziať nulovú hodnotu podnetu a postupne po krokoch ju budeme zvyšovať. Kým hodnota podnetu bude podprahová, nebude vyvolávať pocity. Pri nejakom kroku prechádzame hodnotu prahu, a tu u probanta vzniká pocit. Prah bude teda ležať medzi poslednou hodnotou podnetu, ktorá ešte nevyvolávala pocit (resp. vyvolávala), a prvou hodnotou podnetu, ktorá pocit vyvolala (resp. nevyvolala). Teoreticky by hodnoty prvej a druhej metódy mali fluktuovať okolo spoločného priemeru. Experimenty však ukazujú, že zistené prahy v oboch metódach sú natoľko odlišné, že hodnoty prahu, získané v podaniach, kde veľkosť podnetu klesala, sa rozkladajú okolo jedného priemeru, a hodnoty prahu, získané v podaniach, kde sa veľkosť prahu zvyšovala, sú rozložené okolo druhého priemeru.

Určenie diferenciálnej citlivosti

Pri určení diferenciálnej citlivosti sa dávajú probantovi dva podnety. Podobne ako pri metóde priemernej chyby sa probantovi dáva podnet s konštantnou hodnotou, čiže štandard, ktorý slúži na porovnanie, a premenný podnet, ktorého veľkosť sa od jedného kroku k druhému postupne mení. Postupným zmenšovaním alebo zväčšovaním premenného podnetu experimentátor určuje bod, pod ktorým probant prestáva registrovať rozdiely medzi porovnávacími podnetmi a bod, z ktorého probant začína tieto rozdiely registrovať. Meraním, o koľko musí byť premenný podnet väčší oproti štandardu, aby ich probant vnímal ako rozdielne, dostaneme *horný prah*. Pri meraní, o koľko musí byť premenný podnet menší oproti štandardu, aby probant vnímal ich rozdiel, dostávame *spodný prah*. Rozdiel medzi spodným a horným prahom je pomenovaný ako *prahový interval* alebo ako *interval neurčitosti*.

Metóda konštantných podnetov

Tretou metódou klasických experimentálnych metód je metóda konštantných podnetov. Pokým v metóde priemernej chyby a aj v metóde minimálnych zmien dochádzalo k zmene hodnoty premenného podnetu až po dosiahnutie senzorickeho efektu, v tejto metóde experiment pozostáva z podania podnetov probantovi, ktoré zostávajú nemenné počas celého pokusu. Práve vďaka tejto skutočnosti je odvodený názov metódy - metódy konštantných podnetov. Ak sa táto metóda používa pri zisťovaní absolútnych prahov, probantovi sa podáva jeden podnet a ak chceme zistiť diferenciálny prah, probantovi

sa pri skúške podávajú dva podnety. V tomto prípade jeden podnet slúži ako štandard a veľkosť druhého sa v jednotlivých skúškach mení. V každej jednotlivej skúške musí proband odpovedať podľa svojich dojmov už v dohodnutých kategóriách, ako napr. áno - nie, počujem - nepočujem, vidím - nevidím. Tento systém sa nazýva dvojkategoriálny. Pri určovaní diferenciálnej citlivosti sa používa trojkategoriálny systém ako „väčší“ „menší“, alebo neutrálny. Výsledok, ktorý experimentátor dostáva pri meraní absolútneho alebo diferenciálneho prahu, je počet odpovedí rôznych kategórií, ktoré pripadajú na každý z podaných stimulov. Ďalším podstatným rozdielom v tejto metóde oproti ostatným je, že merania v tejto metóde nám nedajú bezprostredne prahové hodnoty. Výsledky sa vyjadrujú vo forme frekvencií odpovedí, na základe ktorých treba nájsť veľkosť prahu pomocou výpočtu. V tomto prípade sa nám prah odhaluje ako výsledok štatistického spracovania údajov. Z hľadiska štatistiky sa prah určuje ako medián získaného rozloženia odpovedí.

Metóda konštantných podnetov disponuje množstvom matematických aparátov. Základné postupy spracovávania údajov sú:

- *Spracovanie metódou lineárnej interpolácie*
- *Spracovanie metódou Spearmana*
- *Spracovanie metódou sumácie*
- *Fechnerov postup*
- *Prah ako modus*

2.2.3.2 Súčasné experimentálne metódy

Metóda „áno a nie“

Princíp metódy spočíva v tom, že buď sa probantovi podáva alebo nepodáva podnet. Jeho úlohou je odpovedať áno, ak podnet spozoroval a nie, ak podnet nespozoroval. Pri podávaní podnetu sa táto situácia nazýva signálna skúška, a ak sa podnet nepodával, situácia sa nazýva prázna alebo chytáková skúška. Na základe týchto signálových skúšok a probantových odpovedí v nich sa počíta pravdepodobnosť spozorovania stimulu, $P(Y|s)$, podľa probantových odpovedí v prázdnych skúškach sa zase počíta pravdepodobnosť falošného posudku.

Hodnotiaca metóda

Táto metóda je založená na princípe striedania signálu a prázdnej skúšky. Probandovi sa počas skúšky podávajú signály alebo prázdne skúšky. Probandovou úlohou je zhodnotiť, opierajúc sa na svoje dojmy, aká je pravdepodobnosť podania signálu v skúške.

K dispozícií má niekoľko hodnotiacich kategórií. Prvá kategória zodpovedá pravdepodobnosti blízkej alebo rovnej 1,0. Posledná kategória zodpovedá pravdepodobnosti blízkej alebo rovnej 0,0. Ostatné hodnotiace kategórie sú rozložené medzi nimi podľa stupňa poklesu pravdepodobnosti podľa toho, že pri pozorovaní sa vyskytol signál. Proband má k dispozícií určitý súbor odpovedí.

Metóda nútenej voľby

Skúška pozostáva z dvoch alebo viacerých skúšok porovnávania. Proband má vymenovať tú skúšku, v ktorej sa podľa jeho názoru signál podal. Každé meranie pozostáva z niekoľkých prázdnych skúšok pozorovania, v ktorých sa podáva len šum a z jedného, v ktorom sa podáva signál. Pohyb skúšok sa pohybuje medzi dvoma až ôsmimi.

2.2.3.3 Priame metódy psychofyziky

- **Frakcionálna metóda:** probant má za úlohu vybrať z predložených podnetov práve ten, o ktorom si myslí, že je dvakrát menší ako zadaný vzor.
- **Pokusy:** probant má za úlohu určiť, v akom pomere je veľkosť podnetu k veľkosti zadaného vzoru.

Schopnosť ľudí pociťovať podnety s čo najmenším pocitovým prahom a rozlišovať podnety s čo najmenším rozdielovým prahom sa nazýva *zmyslová citlivosť*. Ak niekto pociťuje podnety s veľmi nízkym pocitovým prahom, hovoríme o *rozvinutej absolútnej zmyslovej citlivosti*. V prípade pociťovania podnetov s veľmi malým rozdielovým prahom, hovoríme o rozvinutej rozlišovacej zmyslovej citlivosti.

Zmyslová citlivosť sa rozvíja v praktickej činnosti. Citlivosť sa znižuje pri poruchách alebo aj pri zníženej kvalite činnosti analyzátor, ale aj vekom, spánkovým deficitom alebo únavou. Rozdiely v zmyslovej citlivosti možno pozorovať nielen medzi jednotlivými pozorovacími objektami, ale aj u toho istého človeka, napr. pri zmene podnetov, alebo pri dlhšom pôsobení toho istého podnetu. V tomto prípade hovoríme o *zmyslovej adaptácii*.

Procesy rozhodovania v detekcii

Chápanie absolútneho prahu ako pevnej hranice znamená, že subjekt podáva v pokuse správu, či podnet prekročil hranicu alebo nie. Zmyslové orgány však nemusia fungovať dokonale, môže kolísať pozornosť probanta a tak výsledky pokusov nemusia byť vždy totožné. Zmyslové orgány nemusia vždy fungovať dokonale, taktiež pozornosť probanta môže kolísať.

To, či podnet prekročil hranicu, probant neoznamuje jednoduchým spôsobom. Namiesto toho uskutočňuje zložité rozhodovanie, či je zmyslový vnem spôsobovaný podnetom alebo len akousi chybou v senzorickom systéme probanta. Ak je podnet v experimente vysielaný vždy, môže nastať situácia, že probant si môže byť v niektorých prípadoch neistý, či má odpovedať „áno“ alebo „nie“. Ak chce probant zapôsobiť na experimentátora, vo väčšine prípadov odpovedá áno. Na obmedzenie tohto problému vedci vymysleli falošné pokusy, pri ktorých nie je podnet podaný, aby sa zistilo, ako bude probant reagovať. Odpoveď „áno“ vo falošných pokusoch sa označuje ako „falošný posudok“.

Teória detekcie signálu

Na to, aby sa vysvetlili niektoré problémy s prahmi, slúži teória detekcie signálu. Táto teória predpokladá, že v senzorickom systéme je vždy nejaká náhodná aktivita alebo šum, a že teda neexistuje nulový podnet. Probant sa v experimente nachádza v situácií, keď sa má rozhodnúť, či je pravdepodobnejšie, že zmyslový vnem je spôsobený podaným podnetom alebo šumom v senzorickom systéme. V tomto prípade majú na rozhodovanie vplyv dva faktory:

1. *citlivosť subjektu k podnetu* - ako dobre vidí slabé svetlo alebo počuje slabý tón.
2. *miera subjektu* - nakoľko je schopný odpovedať „áno“.

Detekcia signálu môže závisieť nielen od kritérií, ktoré si určil probant, ale aj od vlastností šumu a intenzity podávaného signálu. Závislosť týchto jednotlivých faktorov možno znázorniť pomocou *kriviek recepčných operačných charakteristík* (ROC). Hodnotenie vzťahu medzi podaným podnetom a hodnotením probanta možno charakterizovať pomocou štyroch možností:

1. *A - A* - podnet bol podaný, probant podnet detekoval.
2. *A - N* - podnet bol podaný, probant podnet nezachytil.
3. *N - A* - podnet podaný nebol, probant oznámil, že podnet zachytil.
4. *N - N* - podnet podaný nebol, probant oznámil, že podnet nezachytil.

2.2.4 Psychofyzikálne škálovanie

V psychofyzike máme mnoho typov psychofyzikálneho škálovania [5], ktoré je možné použiť na meranie vlastností objektov. Každé z rozdielných typov meraní škálovania má svoje matematické vlastnosti. Poznáme tieto typy meraní škálovania:

- **Nominálne meranie (nominal)** - je to najjednoduchšia forma spracovania údajov. Používa sa bežne v každodennom živote. Podstata nominálneho merania je v klasifikácii objektov. Objekty zaradíme do určitých tried podľa zvolených kritérií a taktiež podľa ich spozorovaných znakov. Prvky tried môžeme počítat' a aj porovnávat'. Zo štatistických mier môžeme z tohto merania získať *modus*.
- **Poradové meranie (ordinal)** - taktiež sa nazýva aj ordinálne meranie. Toto meranie vychádza z predpokladu, že všetky objekty určitej triedy môžu byť zoradené podľa definovanej vlastnosti. V triede, ktorá bola vytvorená rovnakým kritériom ako na vytvorenie skupiny, zist'ujeme rozdiely v určitej vlastnosti. V tomto meraní zo štatistických mier získavame hodnotu *medián*.
- **Intervalové meranie (interval)** - Škály v intervalovom meraní majú všetky charakteristiky škál v nominálnom aj poradovom meraní, hlavne charakteristiku poradia. Jednotlivé intervaly môžu byť sčítané a odčítané. Hodnoty, ktoré vzniknú sčítaním a odčítaním intervalov sú odčítané a sčítané intervaly, vzdialenosti, ale nie množstvá. Preto nemajú absolútnu hodnotu na stupnici nulová ale len dohodnutú nulová hodnotu. Čo sa týka štatistického spracovania, máme oveľa väčšie možnosti ako v predchádzajúcich meraniach. Môžeme vypočítat' *aritmetický priemer a smerodajnú odchylku*.
- **Pomerové meranie (ratio)** - Pri tomto meraní majú škály absolútnu nulová hodnotu. Tým, že obsahujú absolútnu nulová hodnotu je možné vykonať porovnanie pomerov. Možnosti štatistického spracovania sú podobné ako pri intervalovom meraní.

Tieto štyri spomenuté typy meraní škálovania sa v podstate rozdeľujú do dvoch základných druhov. Nominálne a poradové typy merania sa zarad'ujú do nižších stupní merania, ktoré predstavujú zist'ovanie početnosti a poradia - **homográdny prípad**. Intervalové a pomerové meranie predstavujú metrické úrovne škálovania - **heterográdny prípad**.

2.2.5 Psychofyzikálne metódy škálovania

Táto široká škála metód bola vytvorená na základe odvodzovania psychofyzikálneho škálovania. Rozdeľuje sa na dve základne kategórie:

1. **Priame** - Proband, podľa svojho vnímania, priamo priradí číselné hodnoty podnetom alebo ich upraví pomocou nejakých matematických vzťahov. Číselné stupnice, ktoré probant vytvorí, sú akýmsi vedľajším produktom experimentu, na ktoré je potrebná len minimálna štatistická analýza.
2. **Nepriame** - Proband je požiadaný, aby jednoducho posúdil vlastnosti podnetov. Výsledkom štatistickej analýzy týchto posudkov sú hodnoty číselnej stupnice.

2.2.5.1 Priame metódy škálovania

Metóda delenia intervalu

Je to akýsi spôsob postupu nastavenia. Test obsahuje sériu pokusov. V každom pokuse je testovací podnet ohraničený na oboch stranách od noriem. Probandovou úlohou je nastaviť jas testovacieho podnetu, kým spadá do polovice medzi normami. Pri druhom pokuse dostávajú objekty nižší štandard a takto to pokračuje ďalej.

Výroba veľkosti

Na začiatku sa probantovi ukáže štandardný jas testovacieho podnetu. Počas pokusu probant nastavuje jas testovacieho objektu, kým nie je z polovice tak jasný, kým nie je dvakrát jasnejší, atď... Nastavenia poskytujú priame odhady a mierky.

Odhad veľkosti

Odhad veľkosti je najznámejšia metóda z priamych metód. Probandovi sa ukáže testovací podnet v štandarde a je mu povedané, že má jas „10“. Potom sa mu zobrazí séria testovacích podnetov a od probanta sa žiada, aby im priradil čísla podľa jasov. Voľba rozsahu čísel je na probantovi.

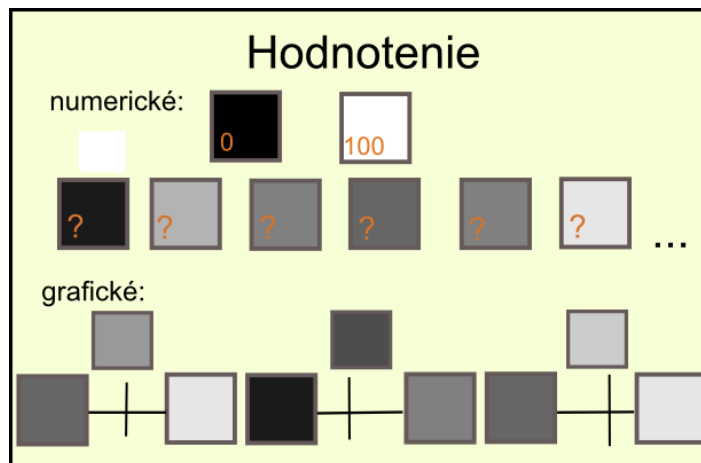
2.2.5.2 Nepriame metódy škálovania

Hodnotenie (rating)

Metóda rating má dve formy, a to *numerickú a grafickú*.

Podstata numerického hodnotenia spočíva v tom, že probant na začiatku dostane dva hraničné koncové body, ktoré sú numericky označené. Prvý je najtmavší, čiže čierny s hodnotou 0 a druhý je najsvetlejší, čiže biely s hodnotou 100. Probandovi sa postupne v sérii testov ukazujú testovacie podnety, ktorým má určiť numerickú hodnotu 0-100 podľa zadanej testovacej otázky. Keďže v našom experimente sme sa zameriavali na hodnotenie jasu, probant zadával jednotlivé numerické hodnoty na základe jasu obrazu.

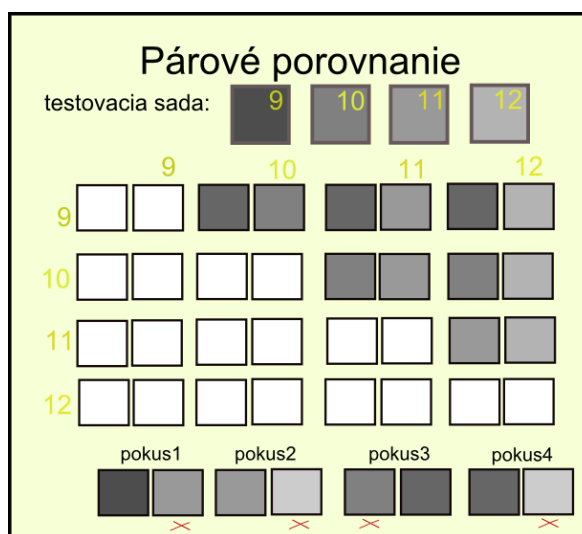
Pri grafickom hodnotení probant dostáva dva koncové body spojené úsečkou, ktorá predstavuje akúsi pomyselnú stupnicu jasu. Proband postupne dostáva sériu testovacích podnetov a jeho úlohou je znázorniť na úsečke miesto, kde by mal testovací podnet padnúť medzi dva koncové body. Numerické hodnoty získavame meraním vzdialenosti od jedného koncového bodu. Grafické hodnotenie využíva skutočnosť, že na krátku vzdialenosť sú skutočná vzdialenosť a vnímaná vzdialenosť priamo spojené.



Obr. 2.1: Hodnotenie

Párové porovnanie (pair comparison)

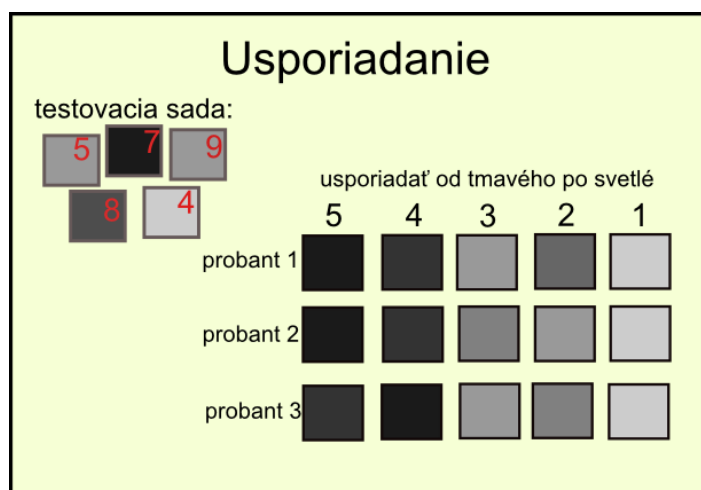
Už ako nám vyplýva z názvu, podnety z testovacích sád sú usporiadané do dvojíc. V každej sérii je prezentovaný jeden pár a probantovou úlohou je povedať, napríklad, ktorý testovací podnet z tejto dvojice je svetlejší. Množstvo testovacích párov na stupnici je $(n)(n-1)/2$. Hodnoty testovacích podnetov musia byť blízko seba, aby bol probant zmätený a robil chyby. Probandovi sa táto metóda môže vidieť ako nudná a veľmi vyčerpávajúca. Štatistické merania pre odvodenie hodnôt stupnice závisia na párových rozdieloch prahu.



Obr. 2.2: Párové porovnanie

Usporiadanie (ranking)

V tejto metóde je probantovi zobrazený celá sada testovacích podnetov v kódovanom poradí. Probant je požiadaný, aby každý testovací podnet opravil podľa jeho jasnosti. Probantovi táto metóda môže pripomínať puzzle a môže sa pri tom dokonca aj zabávať. Spokojný a šťastný probant poskytuje aj dobré údaje.



Obr. 2.3: Usporiadanie

Kategorické škálovanie (category scaling)

Pri tejto metóde je probantovi zobrazený celá sada testovacích podnetov a kategórie jasnosti týchto podnetov. Jeho úlohou je postupne všetky testovacie podnety rozčleniť do daných kategórií.

2.3 Implementácia

2.3.1 Matlab

Systém Matlab je interaktívne programovacie prostredie pre numerické výpočty, vizualizácie a programovanie. Názov Matlab vznikol skrátením slov MATrix LABoratory, čím poukazuje, že kľúčovou dátovou štruktúrou pri výpočtoch v Matlabe sú matice. Použitím Matlabu je možné analyzovať dáta, vyvíjať algoritmy, vytvárať modely a aplikácie. So systémom Matlab je možné pracovať viacerými rôznymi spôsobmi. Pokiaľ potrebujeme spočítať len niečo narychlo, jednoducho použijeme riadkový dialogový spôsob. Ak je našim cieľom práca, ktorá je rozsiahlejšia, alebo po zavretí programu nechceme, aby sa zapísané výrazy stratili, prejdeme na spôsob podobný iným vývojovým prostrediam. Všetky výrazy a príkazy si zapíšeme do editora a potom uložíme na disk. Takýto zdrojový kód je možné spúšťať, ladiť, optimalizovať a prepájať s inými zdrojovými kódmi v iných súboroch.

Pomocou Matlabu dokážeme naprogramovať aplikácie s podporou grafiky. Súčasťou aplikácie môžu byť okrem grafu taktiež grafické prvky ako tlačítka, posuvníky, pole pre editáciu či zápis textu... Vytvárať grafické prvky je možné buď pomocou špeciálnych príkazov s mnohými možnými parametrami v editore zdrojového kódu, alebo môžeme použiť automatický generátor grafických prvkov. Pre optimálny zdrojový kód je lepšie vytváranie grafických prvkov manuálne.

Dôležitou súčasťou Matlabu sú knižnice funkcií, tzv. toolboxy. Toolboxy obsahujú problémovo orientované funkcie pre jednotlivé oblasti použitia, či už v numerickej matematiky, analytickej matematike, v spracovaní obrazu a v iných odvetviach, v ktorých nachádza Matlab uplatnenie.

Matlab má veľmi dobre spracovaný štatistický toolbox, ktorý nám pomáha pri implementácii vyhodnocovacích metód psychofyzikálnych experimentov. Vďaka grafickým prvkom implementácia obsahuje jednoduché GUI, ktoré slúži na lepšie ovládanie softvéru.

3. Metódy vyhodnocovania

Existuje široké spektrum štatistických metód na spracovanie experimentálnych dát. Každá štatistická metóda je svojím spôsobom odlišná. Je dobré si uvedomiť, čo od experimentu a výsledkov očakávame a podľa toho si zvolit' správnu metódu na analýzu dát. V našej práci sme sa zamerali na tri metódy škálovania, ktoré boli použité v psychofyzikálnych experimentoch a na analýzu ich dát.

3.1 Usporiadanie (ranking)

Metóda usporiadania dodáva probantovi pocit šťastia a spokojnosti. Od týchto pocitov sa odvíjajú aj výsledné dáta, ktoré môžu byť oveľa presnejšie. Podstata tejto metódy je v usporiadaní testovacieho setu, ktorý sa probantovi prezentuje neusporiadaný, zle ohodnotený alebo zakódovaný. Úlohou probanta je testovací set preusporiadať do správneho poradia podľa jeho uváženia. Metóda usporiadania sa môže pre probanta javiť ako skladanie puzzle.

3.1.1 Psychofyzikálny experiment

Psychofyzikálny experiment [2] je zameraný na porovnanie dvoch zobrazení obrázkov. HDR (High dynamic range) obrázky a LDR (low dynamic range) obrázky, ktoré sa od seba líšia hodnotou rozsahu jasu. HDR obrázok je vo všeobecnosti považovaný za lepší obrázok ako LDR. Experiment to buď vyvráti alebo potvrdí.

Experiment chcel zistiť subjektívne hodnotenie nasledovných troch zobrazovacích metód:

1. Vytvorenie obrázka pomocou HDR zariadenia, zobrazenie na HDR zariadení- 1 obrázok
2. Vytvorenie obrázka pomocou HDR zariadenia, zobrazenie na LDR zariadení- 3 obrázky metód tone mappingu
3. Vytvorenie obrázka pomocou LDR zariadenia, zobrazenie na LDR zariadení- 2 obrázky najlepšej subjektívne a objektívnej expozície

Všetky tri procesy zobrazenia boli rovnaké s rozdielom, že v každom zobrazení boli použité im určené druhy obrázkov. Na experiment bolo vytvorených 10 obrázkov rôznych scén ako deň, noc, interiér, krajina a zátišie. Každý proces začal prezentáciou po sebe idúcich

šiestich obrázkov v náhodnom poradí. Každý obrázok bol zobrazený po dobu dvoch sekúnd. Následne sa probantovi zobrazila prezentácia obrázkov v tvare tabuľky 2x3, ktorá bola na obrazovke až do doby, kým nebol zostavený rebríček. Rebríček bol udávaný stlačením príslušného tlačidla probantom. Proband vyberal usporiadanie obrázkov nasledovne: na prvé miesto si vybral obrázok, ktorý v danom zobrazení uprednostňoval najviac, a potom pokračoval podľa preferencií. Proband mal možnosť preklikávať si medzi dvoma obrázkami, pre ich lepšie porovnanie. V prípade chyby mohli probanti obnoviť ich rozhodnutie pre aktuálny proces. Časový limit nebol určený, no experiment by nemal presiahnuť dve minúty.

3.1.1.1 Testy štatistických hypotéz, P-hodnota

Štatistická hypotéza [12] je tvrdenie o základnom súbore. Výskumu sa podrobuje iba časť základného súboru - vzorka, aby bolo možné rozhodnúť o správnosti alebo nesprávnosti hypotézy. Testovanie štatistických hypotéz je proces, ktorý overuje správnosti alebo nesprávnosti hypotézy pomocou vzorky výsledkov vybraných náhodným výberom.

Postup testovania hypotéz:

1. *Formulácia nulovej hypotézy (H_0)*

Formulujeme nulovú hypotézu. Nulová hypotéza vyjadruje nezávislosť medzi premennými v základnom súbore.

2. *Formulácia alternatívnej hypotézy (H_α)*

Väčšinou sa dokazuje pravdivosť alternatívnej hypotézy, ktorá najčastejšie dokazuje štatistickú závislosť premenných. Pravdivosť alternatívnej hypotézy sa dokazuje vždy nepriamo a to tak, že ukážeme, že nulová hypotéza je nepravdivá a alternatívna je tým pádom pravdepodobná.

3. *Stanovenie hladiny významnosti (α)*

Hladina významnosti je pravdepodobnosť chyby I. druhu, ktorú urobíme, keď zamietneme nulovú hypotézu, ktorá v skutočnosti platí. α sa tradične stanovuje na 5 % (=0,05) alebo na 1 % (=0,01) .

4. *Výpočet testovacej štatistiky a pravdepodobnosti*

Testovacia štatistika sa vypočíta zo vzorky, ktorá má za predpokladu pravdivosti nulovej hypotézy príslušné pravdepodobnostné rozdelenie - príslušnú štatistickú metódu. *P-hodnota* predstavuje pravdepodobnosť, že testovacia štatistika za predpokladu pravdivosti nulovej hypotézy dosiahne pri najmenšom tak extrémnu hodnotu, ako je hodnota vypočítaná zo vzorky. P-hodnota je teda pravdepodobnosť, že vzťah zistený z údajov je iba dôsledkom zlej vybranej vzorky. Je teda možné, že aj keby sme vybrali inú náhodnú vzorku, nemuseli by sme nič nájsť. P-hodnota je

najnižšia hodnota hladiny významnosti, ktorá vedie k zamietnutiu nulovej hypotézy. *P-hodnota je odhadovaná pravdepodobnosť zamietnutia pravdivej nulovej hypotézy.* Čím menšia je P-hodnota, tým sme si viac istí, že nulová hypotéza nie je pravdivá a mala by byť zamietnutá.

5. Rozhodnutie

Ak je $P < \alpha$, nulová hypotéza sa voči alternatívnej hypotéze zamietne. Medzi premennými existuje vzťah. Sú štatisticky významné. Ak je $P \geq \alpha$ to znamená, že nulovú hypotézu nie je možné zamietnuť. Rozdiel nameraný vo vzorke môže byť iba náhodný, teda nemáme dostatočný počet dôkazov na to, aby sme nulovú hypotézu zamietli. Tým pádom vzťah medzi premennými neexistuje a nie sú štatisticky významné.

Hranice oblasti prijatia nulovej hypotézy

$$I_\alpha = (t_1, t_2). \quad (3.1)$$

Zamietnutie

$$(-\infty, t_1) \cup (t_2, \infty). \quad (3.2)$$

Výpočet t_1, t_2

$$F(t_1) = \frac{\alpha}{2} \quad (3.3)$$

$$F(t_2) = 1 - \frac{\alpha}{2}. \quad (3.4)$$

Body oddeľujúce oblasť prijatia nulovej hypotézy od oblasti jej zamietnutia sa nazývajú *kritické hodnoty*. V tabuľke 3.1 sú znázornené simulácie, ktoré môžu nastať pri testovaní hypotéz:

		Skutočná situácia	
		H_0 je pravda	H_0 nie je pravda
Naše rozhodnutie	H_0 je pravda	správne rozhodnutie pravdepodobnosť $1 - \alpha$	chyba typu II pravdepodobnosť β
	H_0 nie je pravda	chyba typu I pravdepodobnosť α hladina významnosti	správne rozhodnutie pravdepodobnosť $1 - \beta$ sila testu

Tabuľka 3.1: Simulácie pri testovaní hypotéz [7]

3.1.1.2 Jednofaktorová analýza rozptylu (one-way ANOVA)

Analýza rozptylu (ANOVA) [8] patrí medzi najčastejšie používané metódy vo výskumnej praxi. ANOVA je štatistická metóda na porovnanie stredných hodnôt niekoľkých základných súborov. Od jednej alebo od viacerých kvalitatívnych premenných, tzv. faktorov, môže závisieť hodnota kvantitatívnej premennej. Podľa toho, či sledujeme vplyv jedného alebo viacerých faktorov, hovoríme o *jednofaktorovej*, *dvojfaktorovej* alebo *viacfaktorovej analýze rozptylu*. Jednotlivé varianty faktora nazývame *úrovne faktora*.

Jednofaktorová analýza rozptylu je najjednoduchšou formou ANOVA. Pri jednofaktorovej analýze rozptylu sa predpokladá, že úroveň priemerov niekoľkých základných súborov závisí len od jedného faktora. Jednofaktorová ANOVA skúma vzťah medzi intervalovou a nominálnou premennou (faktorom).

Sleduje sa faktor na K rôznych úrovniach. Na každej úrovni sme previedli n_i meraní $y_{ij}, j = 1, \dots, n_i$. Jednofaktorová ANOVA je jednoduchý špeciálny prípad lineárneho modelu. Lineárny model jednofaktorovej analýzy rozptylu je možné zapísať

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad (3.5)$$

kde

- μ - celkový priemer,
- y_{ij} - matica pozorovaní, v ktorej každý stĺpec prezentuje inú skupinu,
- α_i - príspevok i -tej úrovne faktora A,
- e_{ij} - nezávislé náhodné veličiny.

Celkový počet meraní je

$$N = \sum_{i=1}^K n_i. \quad (3.6)$$

Celkový priemer μ je súčet všetkých hodnôt y_{ij} delení celkovým počtom dát N .

Testovanie

Súčet štvorcov odchýliek od celkového priemeru μ , ktorý je definovaný vzťahom

$$S_e = \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \mu)^2 \quad (3.7)$$

sa dá rozložiť na dve zložky

$$S_e = \sum_{i=1}^K \sum_{j=1}^{n_i} [(y_{ij} - \mu_i) + (\mu_i - \mu)]^2 = S_A + S_R. \quad (3.8)$$

S_A predstavuje rozptyl medzi jednotlivými úrovňami daného faktora

$$S_A = \sum_{i=1}^K n_i (\mu_i - \mu)^2. \quad (3.9)$$

S_R je reziduálny rozptyl, čiže rozptyl vo vnútri jednotlivých úrovní

$$S_R = \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2. \quad (3.10)$$

Nevychýleným odhadom rozptylu chýb σ_e^2 , je priemerný reziduálny štvorec MS_R

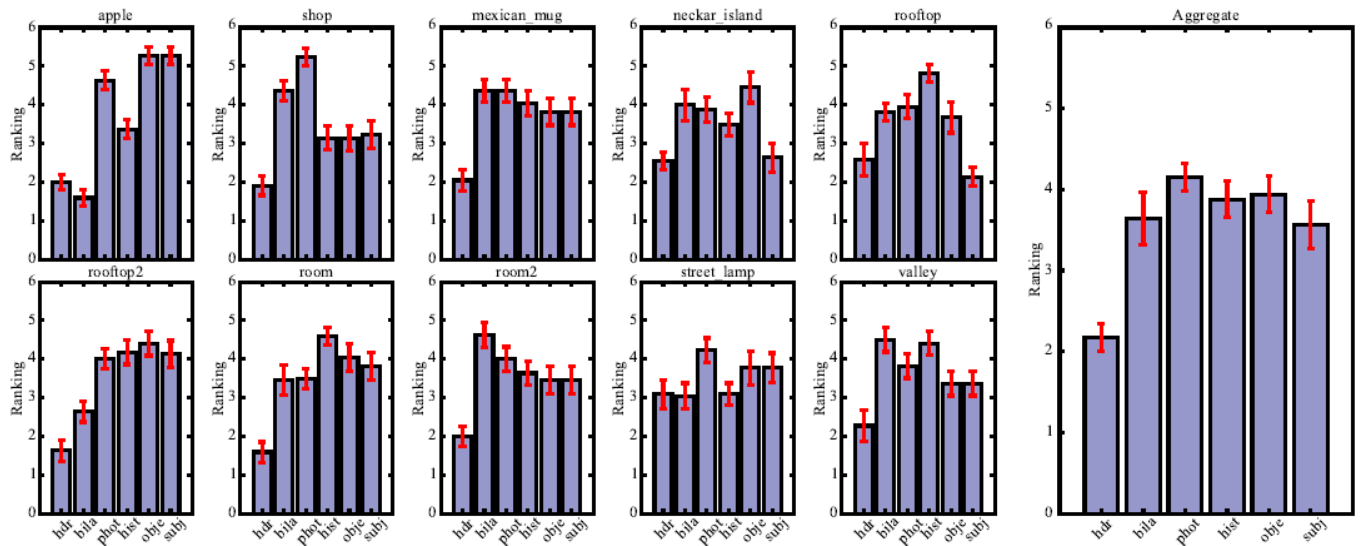
$$MS_R = \frac{S_R}{N - K} = \sigma_e^2. \quad (3.11)$$

Cieľom jednofaktorovej analýzy rozptylu je odhaliť, či vo vzorke zistené rozdiely priemerov jednotlivých skupín (podľa úrovné faktora) sú štatisticky významné (medzi premennými je vzťah) alebo môžu byť iba náhodné (medzi premennými nie je vzťah). Testuje sa nulová hypotéza $H_0 : \alpha_i = 0, i = 1, \dots, K$ oproti alternatívnej hypotéze $H_A : \alpha_i \neq 0, i = 1, \dots, K$. Pri testovaní sa využíva fakt, že veličina S_A/σ_e^2 má χ^2 -rozdelenia s $(K - 1)$ stupňami voľnosti a veličina S_R/σ_e^2 má nezávislé χ^2 -rozdelenia s $(N - K)$ stupňami voľnosti. Ich podiel má F-rozdelenie s $(K - 1)$ a $(N - K)$. Testovacia štatistika F_e má tvar

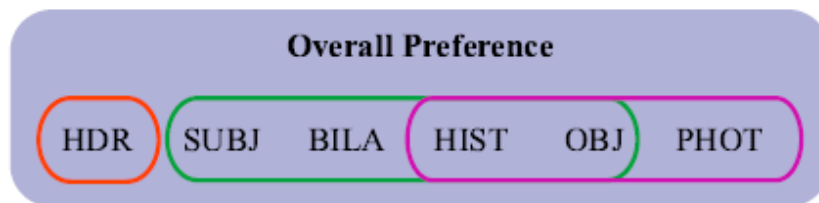
$$F_e = \frac{MS_A}{MS_R} = \frac{S_A(N - K)}{S_R(K - 1)}. \quad (3.12)$$

Pri platnosti nulovej hypotézy H_0 má F_e štatistika rozdelenie s $(K - 1)$ a $(N - K)$ stupňami voľnosti. Pri F_e väčšom než kvantil $F_{1-e}(K - 1, N - K)$, je nutné nulovú hypotézu H_0 na hladine významnosti α zamietnuť a efekty považovať za štatisticky významné.

3.1.1.3 Výsledky



Obř. 3.1: Stredné usporiadanie z experimentu. Hdr predstavuje fyzikálne správny HDR obraz, bila, phot, hist reprezentujú použité tone mappingové metódy. Obje a subj reprezentujú najlepšie objektívne a subjektívne obrázky. Vľavo sú rebríčky jednotlivých scén. Vpravo je súhrnný výsledok všetkých scén. Spodný riadok ukazuje vyššie preferencie. Chybové úsečky označuje smerodajná odchylka ± 1 [2].



Obř. 3.2: Podobnosť skupín v experimente. Položky v rovnakom súbore sú štatisticky nerozoznateľné [2].

Obrázok 3.1 zobrazuje priemerný rebríček všetkých probantov. Spodný riadok usporiadania označuje vyššie preferencie.

Významné celkové rozdiely boli medzi spracovaním obrazu číslo 6, ktoré boli odhalené opakovou dvojfaktorovou analýzou rozptylu 1 : $F(5, 105) = 20.132, p < 0.001$. Výsledne podobnosti skupín na 95 % úrovni významnosti sú uvedené na obrázku 3.2 .

Tieto testy ukazujú, že probanti vo všeobecnosti uprednostňovali HDR zobrazenie, aj keď existujú výnimky. V experimente sa zistilo, že bilaterálny filter mal vo všeobecnosti lepšie výsledky ako fotografický tone mappingový operátor, ale nebola preukázaná jednoznačná výhoda obrázkov spracovaných tone mappingom nad obrazmi v tej najlepšej expozícii.

3.2 Párové porovnanie (pair comparison)

Metóda párového porovnania je veľmi obľúbená pri experimentoch zameraných na spracovanie obrazu. Podstata tejto metódy spočíva v porovnávaní párov, ktoré sa skladajú z testovacích podnetov. Hodnoty testovacích podnetov musia byť blízko seba, aby bol probant zmätený a robil chyby. Na odvodenie hodnôt stupnice potrebujeme párové rozdiely prahu, od ktorých nám závisí štatistické meranie.

3.2.1 Psychofyzikálny experiment 1

V experimente [10] bolo použitých päť nasledovných farebných obrazov.

- päť mužov a päť žien rôznych rás sediacich na sedačke
- košík plný ovocia a zeleniny na tmavohnedom pozadí
- dvaja golfisti v pestrofarebnom oblečení na golfovom ihrisku s jasnou modrou oblohou
- žltý motýľ sediaci na niekoľkých ružových kvetoch obklopených zeleňou
- červená stodola a biely dom stojaci pred diaľnicou s modrou oblohou

Týchto päť scén bolo starostlivo vybraných pre obsahovosť rôznych odtieňov.

V experimente bolo použitých 5 transformácií na prezentáciu obrazov na CRT monitore - von Kries, CIELAB, RLAB, Hunt a Nayatani. Predkladané obrázky boli zobrazované na 100dpi CRT monitore.

V psychofyzikálnom experimente bolo použité párové porovnávanie na základe porovnania originálnej scény s upravenou scénou pre CRT monitor. Originál obrázku a reprodukcia CRT boli umiestnené 90° od seba ohľadom na pozorovateľa. Keď probanti cítili, že si dostatočne zapamätali farby obrazu, obrátili svoju pozornosť na monitor. Na monitore probanti porovnávali dvojicu reprodukcí. Reprodukcie sa postupne zobrazovali v strede CRT monitora a probant rozhodoval, ktorá z dvoch reprodukcí sa približuje viac k originálu. Každá reprodukcia bola vyhotovená pomocou piatich transformácií, čo vyústilo do desiatich párových porovnaní. Poradie zobrazovania reprodukcí bolo náhodné. Tento proces sa opakuje pre päť rôznych pôvodných scén.

3.2.1.1 Údaje párového porovnania

Použitím Thurstonhovo zákona porovnávacích rozhodnutí boli vybrané CRT reprodukcie prevedené na škále intervalov kvality reprodukcie farieb pre rôzne modely. Následne bola skonštruovaná matica p_{ij} . Matica p_{ij} obsahuje stĺpec matice, ktorý prezentuje

pomer pozorovateľov posudzujúcich CRT reprodukciu a riadok matice, ktorý reprezentuje bližšiu vizuálnu zhodu s originálom než s reprodukciou. Tieto pomery boli prevedené na *z-skóre* (jednotka odchýlky) s odvolaním na tabuľku oblastí v rámci jednotky normálnej krivky. Každý spriemerovaný stĺpec matice *z-skóre* dal hodnotu intervalu škály pre daný model, ktorý bol použitý pri vizuálnej zhode CRT reprodukcie a tlačeného originálu. Niekoľko pomerov matíc obsahovalo nulové vstupy, z ktorých vyplývalo, že žiadny z pozorovateľov si nevybral danú reprodukciu pred ďalšou reprodukciou.

Ak mala matica celý stĺpec núl, tak tento postup nebolo možné použiť. Údaje v týchto prípadoch boli vylúčené z analýzy.

3.2.1.2 Interval spoľahlivosti

Cieľom je na základe dát zostrojiť interval [14], o ktorom môžeme s primeraným stupňom dôvery prehlásiť, že obsahuje skutočnú hodnotu parametra θ . Úlohou je nájsť na základe náhodného výberu taký interval, ktorý bude obsahovať neznámu hodnotu parametra θ s pravdepodobnosťou $1 - \alpha$. Takéto intervaly sa nazývajú $100(1 - \alpha)\%$ *intervaly spoľahlivosti s koeficientom spoľahlivosti* $1 - \alpha$, $\alpha = 0.05$, $\alpha = 0.01$, α je volené štatistikom. Intervaly spoľahlivosti (IS) rozlišujeme:

- Dvojstranné (DIS)

pre DIS sú koncové body $DH(X_1, \dots, X_n)$ a $HH(X_1, \dots, X_n)$ vytvorené tak, aby platilo

$$P(DH(X_1, \dots, X_n) \leq \theta \leq HH(X_1, \dots, X_n)) = 1 - \alpha. \quad (3.13)$$

- Jednostranné (JIS)

Pre JIS volíme konkrétnu hranicu tak, aby mohlo dôjsť len k podceneniu neznámeho parametra (pravostranný) alebo preceneniu neznámeho parametra (ľavostranný), aby platilo

$$P(DH_*(X_1, \dots, X_n) \leq \theta) = 1 - \alpha, \text{ pre } \forall \theta \in \Theta(\text{dolný}) \quad (3.14)$$

$$P(\theta \leq HH^*(X_1, \dots, X_n)) = 1 - \alpha, \text{ pre } \forall \theta \in \Theta(\text{horný}). \quad (3.15)$$

V experimente bol IS škálových hodnôt vypočítaný na základe škálovacích jednotiek. Jedna škálovacia jednotka sa rovnala $\sqrt{2}\sigma$. Preto má smerodajná odchylka σ hodnotu $1/\sqrt{2}$ alebo 0.707 jednotiek. 95 % intervalu spoľahlivosti sa vypočíta rovnicou:

$$R \pm (1.96\sigma/N^{1/2}) = R \pm (1.96 * 0.707/N^{1/2}) = R \pm (1.39/N^{1/2}) \quad (3.16)$$

kde N je počet pozorovaní každej dvojice. Keď výsledky obrázkov boli spriemerované, interval spoľahlivosti bol vypočítaný uvedenou rovnicou a predelený druhou mocninou počtu obrázkov - $\sqrt{5}$.

3.2.1.3 Chi-kvadrát test dobrej zhody

χ^2 test [14] vo všeobecnosti slúži na určenie dobrej zhody medzi teoretickými a experimentálnymi údajmi. Porovnávajú sa pozorované hodnoty s očakávanými hodnotami. Pozorované hodnoty sú tie, ktoré sú získané pomocou priameho pozorovania a očakávané sú vyvinuté na základe hypotézy. Štatistický výsledok je definovaný rovnicou:

$$\chi^2 = \sum_{i=1}^k \frac{(O - E)^2}{E} \quad (3.17)$$

kde

- O = pozorovaná hodnota,
- E = očakávaná hodnota,
- k = počet kategórií alebo skupín.

Štatistický test ω je daný rovnicou

$$\omega = \sum \frac{(\theta'_{ij} - \theta_{ij})^2}{(821/N)}, \quad (3.18)$$

kde N je číslo pozorovaní, $\theta_{ij} = \arcsin[(p_{ij})^{1/2}]$ a $\theta'_{ij} = \arcsin[(p'_{ij})^{1/2}]$.

Nulová hypotéza bola odvodená a pozorované podiely neboli štatisticky odlišné. Ak štatistický test ω bol väčší ako kritická hodnota $\chi^2_{\sigma, \nu'}$, potom nulová hypotéza bola odmietnutá, čo naznačuje, že jeden alebo viac predpokladov modelu boli nesprávne. ν je počet voľných stupňov, ktoré sa rovnajú $(n - 1)(n - 2)/2$, kde n je počet podnetov, čiže v našom prípade 5.

3.2.1.4 Výsledky

χ^2 test bol použitý na test za predpokladu, že psychofyzikálne údaje vznikli za normálneho behu experimentu. Test meria predpokladaný pomer dobrej zhody hodnôt škály pozorovaných pomerov.

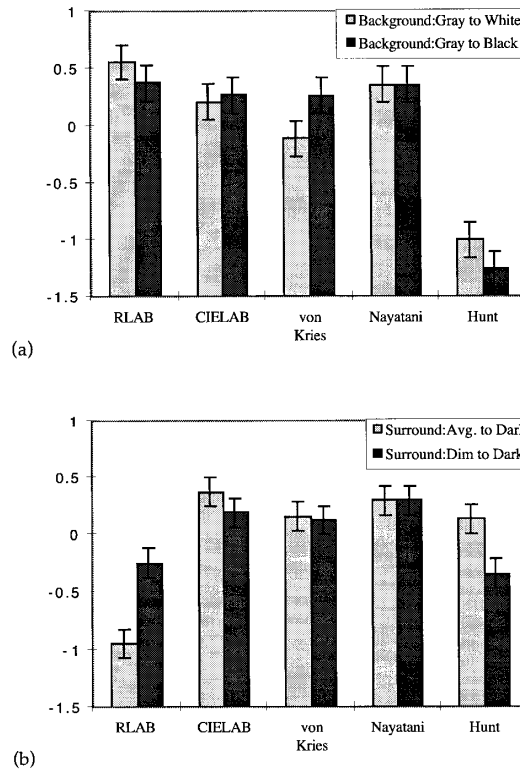
Keď pozorovaný podiel matíc obsahoval hodnoty núl a jednotiek, nebolo možné odvodiť pomer vo výške jedna alebo nula z odhadovanej z-skóre matice. Tieto odhadované pomery v matici spôsobili, že štatistický χ^2 test bol neúmerne veľký.

Jeden test zo všetkých testov obsahujúci len obrázky, ktoré obsahovali jednoznačnú zhodu, neprešiel χ^2 testom. Zvýšenie počtu pozorovateľov alebo odstránenie modelov, ktoré nikdy neprodukujú akceptovateľnú zhodu, by znížilo výskyt núl a jednotiek v týchto podieloch matice.

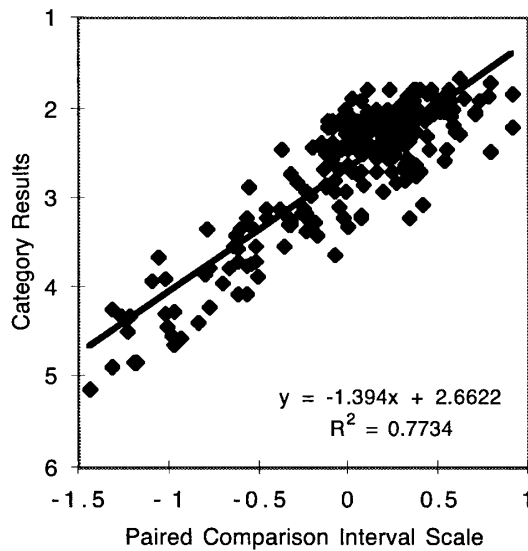
V experimente 5, ktorý obsahoval najviac rozdielov medzi pozorovanými podmienkami, štyri z piatich obrázkov týmto testom neprešli. Tri z podielov matíc nemali problém z jednoznačným určením. Ako povedal Mosteller: „Alternatívna hypotéza je celkom všeobecná: výlučne nulová hypotéza nie je správna.“ To môže nastať v jednodimenziálnosti alebo v zlyhaní aditívnosti údajov.

Nulová hypotéza predpokladá, že ak x je vzdialenosť medzi modelmi a a b na obrázku 3.3, a ak je y vzdialenosť medzi modelmi b a c na obrázku 3.3, potom $x + y$ musí byť vzdialenosť medzi modelmi a a c na obrázku 3.3. Pokiaľ to tak nie je, tak χ^2 test sa pravdepodobne nepodarí. Druhým dôvodom zamietnutia nulovej hypotézy môže byť, že štandardné odchýlky podnetov nie sú rovnaké. Predpokladá sa, že vo výpočte hodnôt škály sú rovnaké. Ak nie sú, reprodukcia pomerov z hodnôt škály nie je možná.

Obrázok 3.4 znázorňuje výsledky všetkých experimentov. Hodnotenie „dobre“ znázorňuje na škále hodnotu 0,475. Lineárnou regresiou sa zistilo, že hodnotenie „dobre“ u probantov sa pohybovalo na škále medzi 0,25 a 1.



Obr. 3.3: Interval škálových hodnôt experimentu, ktoré obsahovali buď zmenu (a) pozadia alebo (b) orámovanie medzi originálom a reprodukciou. Chybové úsečky predstavujú 95 % interval spoľahlivosti okolo priemernej hodnoty [10].



Obr. 3.4: Porovnanie výsledkov hodnôt škály získaných použitím párového porovnania, kde 1 je výborný, 2 - dobrý, 3 - prijateľný, 4 - neprijateľný, 5 - zlý, 6 - hrozný. Zahŕňa údaje všetkých porovnaní v tomto psychofyzikálnom teste [10].

3.2.2 Psychofyzikálny experiment 2

V druhom psychofyzikálnom experimente [13] sú skúmané emócie človeka na základe emocionálnych algoritmov a interaktívneho genetického algoritmu. Pomocou psychofyzikálneho škálovania, meraním pocitov pozorovateľov, je možné posúdiť, či emocionálne algoritmy sú dobré alebo nie. Metóda párového porovnania vytvára interval hodnôt škály. Tým sú získané kvantitatívne vyhodnotenia rôznych emočných algoritmov.

V experimente je používaná Scheffenova metóda párového porovnávania, pretože probanti nemajú za úlohu len vybrať lepší obraz, ale aj priradiť skóre rozdielu podľa zadaného kritéria rozhodnutia. Predpokladajme, že máme n prvkov, potom počet porovnaní je C_n^2 . Pár sa skladá z položiek i a z položiek j . Proband k priradí uje hodnotenie podľa kritéria, aké vyvolá obraz emócie podľa zadanej emocionálnej požiadavky. Toto hodnotenie je označované ako v_{ijk} . Všetky hodnotenia probantov dostaneme podľa rovnice

$$v_{ij} = \sum_{i=1}^k v_{ijk}. \quad (3.19)$$

Sématická štruktúra obrazov zahŕňa objekty, priestorové vzťahy, udalosti, správanie a emócie. Emocionálne sématická sú veľmi abstraktné a zvyčajne sú popísané prídavnými menami ako št'astný, romantický. Hodnotia sa obrázky získavané interaktívnym genetickým algoritmom v porovnaní s náhodným emočným algoritmom, ktorý používa rovnaké obrázky a rovnaké prídavné mená.

Boli vytvorené a porovnané dva druhy emocionálnych obrázkov - obrazy oblečenia (thick clothes images) a obrazy s pohľadom na krajinu (wide view landscape images). Pre každé prídavné meno bolo vytvorených šesť snímok. Tri z nich boli vytvorené interaktívnym genetickým algoritmom a boli označené A, C a E. Ďalšie tri obrazy boli vybrané náhodne a boli označené B, D a F. Počet obrázkov použitých na subjektívne porovnávanie bolo C_6^2 .

Zobrazený bol vždy pár obrázkov a probanti mali ohodnotiť obrázky podľa svojej spokojnosti. Napríklad, pre obrazy oblečenia sa zobrazia súčasne obrazy A a B. Ak pozorovateľ považoval za oveľa silnejšie oblečenie v obraze A ako tie v obraze B, dal hodnotenie 2. Ak považuje oblečenie v obraze A len o niečo silnejšie ako v obraze B, dal hodnotenie 1. Ak považuje oblečenie obrazov za rovnako silné, dal hodnotenie 0. V opačnom prípade dával hodnotenie -2, -1. Na záver sa všetky hodnotenia probantov zhromaždili a štatisticky vyhodnotili.

3.2.2.1 Kruskalov-Wallisov test

Kruskalov-Wallisov test [11] je neparametrickou obdobou jednofaktorovej analýzy rozptylu. Umožňuje testovať hypotézu H_0 tak, že k ($k \geq 3$) nezávislých súborov pochádza z toho istého rozdelenia.

Rozsahy jednotlivých testovacích súborov sme si označili n_1, n_2, \dots, n_k . Zoradili sme všetkých n prvkov do neklesajúcej postupnosti a každému prvku sme priradili jeho poradie. Súčet poradí i -teho testovacieho súboru ($i = 1, 2, \dots, k$) sme si označili T_i . Pretože musí platiť $T_1 + T_2 + T_3 + \dots + T_k = \frac{n(n+1)}{2}$, môžeme tento vzťah použiť na kontrolu výpočtu hodnôt charakteristík T_i ($i = 1, 2, \dots, k$). Testovacia štatistika má tvar

$$K = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{T_i^2}{n_i} - 3(n+1). \quad (3.20)$$

Ak sa v postupnosti získaných údajov objavia zhodné hodnoty, ktorým sa priraduje priemerné poradie, je potrebné hustotu testovacieho kritéria K deliť tzv. *korekčným faktorom*. Jeho hodnotu vypočítame nasledujúcim vzťahom

$$f = 1 - \frac{\sum_{i=1}^p (t_i^3 - t_i)}{n^3 - n}, \quad (3.21)$$

kde:

- p - počet tried s rovnakým poradím
- t_i - počet poradí v i -tej triede.

Testovacia štatistika [3] má potom tvar

$$K_{opr.} = \frac{K}{f}. \quad (3.22)$$

Ak zamietneme testovanú hypotézu H_0 v prospech alternatívnej hypotézy H_1 , ktorá znamená, že výbery nepochádzajú z toho istého rozdelenia, ostáva nezodpovedaná otázka, ktoré výbery sa od seba štatisticky významne líšia. Pri Kruskalovom-Wallisovom teste sa postupuje nasledovne. Pre každú dvojicu porovnávaných súborov sme vypočítali priemerné poradie

$$\bar{T}_i = \frac{T_i}{n_i} \quad (3.23)$$

$$\bar{T}_j = \frac{T_j}{n_j} \quad (3.24)$$

Testovacím kritériom nulovej hypotézy H_0 , ak rozdelenia i -teho súboru a j -teho súboru sú rovnaké, absolútna hodnota rozdielu ich priemerných poradí je

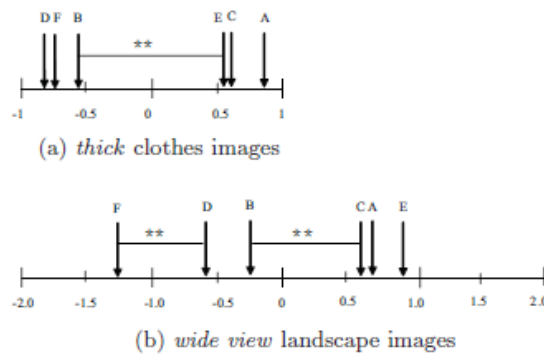
$$D = |\bar{T}_i - \bar{T}_j|. \quad (3.25)$$

3.2.2.2 Výsledky

<i>thick clothes images</i>				
distance	$\alpha = 0.01$		$\alpha = 0.05$	
AC	0.267 < 0.439		0.267 < 0.369	
CE	0.063 < 0.439		0.063 < 0.369	
EB	1.008 > 0.439	**	1.008 > 0.369	**
BF	0.131 < 0.439		0.131 < 0.369	
FD	0.045 < 0.439		0.045 < 0.369	

<i>wide view Landscape images</i>				
distance	$\alpha = 0.01$		$\alpha = 0.05$	
EA	0.225 < 0.454		0.225 < 0.382	
AC	0.088 < 0.454		0.088 < 0.382	
CB	0.845 > 0.454	**	0.845 > 0.382	**
BD	0.305 < 0.454		0.305 < 0.382	
DF	0.705 > 0.454	**	0.705 > 0.382	**

Obr. 3.5: Vzďalenessi medzi načítanými obrázkami [13]



Obr. 3.6: Vyvolané obázky na psychofyzikálnej škále [13]

Obrázok 3.5 a 3.6 zobrazuje vzdialenessi medzi šiestimi obrazmi. Vodorovná os na obrázku 3.6 predstavuje postavenie na psychofyzikálnej škále s lepším hodnotením na pravo v stupnici. Ak sú vzdialenessi väčšie ako prahová hodnota, znamená to, že rozdiel medzi týmito dvoma obrazmi je štatisticky významný.

Pre obrazy oblečenia vzdialenessť medzi E a B na intervale hodnôt rozsahu je väčšia ako prahová hodnota. To znamená, že obrazy A, C a E, ktoré boli vyvolané interaktívnym genetickým algoritmom, sú štatisticky významnejšie a lepšie ako tie, ktoré boli vybrané náhodne a to na úrovniach $\alpha = 0.05$ a $\alpha = 0.01$.

Podľa toho sme prišli k záveru, že interaktívny genetický algoritmus je lepší ako náhodný algoritmus. Obrazky so širokým pohľadom na krajinu C, A a E sú tiež štatisticky výrazne lepšie ako B, D a F. Experiment dokazuje, že interaktívny genetický algoritmus je lepší ako náhodný algoritmus.

3.3 Hodnotenie (rating)

Metóda hodnotenia má dve podoby: numerickú, v ktorej probant udáva hodnoty a grafickú, v ktorej hodnotu zadáva na graficky zadanej škále. Psychofyzikálne experimenty sú zamerané na numerické hodnotenie.

3.3.1 Psychofyzikálny experiment 1

V experimente [15] boli použité obrázky z verejne prístupnej databázy, ktorá sa skladala zo sady 80 poškodených obrázkov. Sada obrázkov bola vytvorená zo sady 7 referenčných obrázkov v stupňoch šedej. Obrázky obsahujú vzory poškodenia, ako napríklad blokovanie, intenzita, rozostrenie, maskovanie a vysokofrekvenčný šum. Tieto poškodenia sa objavujú v rôznych kombináciách v rámci jedného obrázka. Pokrývajú širokú škálu veličín od takmer neviditeľných až po veľmi spozorovateľné. Okrem toho, že obrázok obsahuje rovnomerne rozložené poškodenie, obsahuje aj klastre obrazu, ktoré poškodené nie sú.

Experiment bol vykonaný v miestnosti za zhoršených svetelných podmienok. Pozorovacia vzdialenosť bola štvornásobkom výšky pozorovaného obrazu. Experimentu sa zúčastnilo 15 probantov. Väčšina probantov boli neodborníci až na troch, ktorí uviedli, že pracovali so spracovaním obrazu. Experiment bol rozdelený do dvoch čiastkových experimentov s prestávkou, pre lepšiu koncentráciu probantov.

V každom z čiastkových experimentov bola použitá sada obrázkov, ktorá pozostávala z polovice poškodených obrázkov a zo 7 referenčných obrázkov, čiže zo 47 obrázkov. Testovacie obrázky v každej sade boli vybrané náhodným spôsobom. Probandi ohodnocovali obrázky podľa kvality a spoľahlivosti v bodovej škále 1-5. Taktiež sa meral reakčný čas probanta pri každom ohodnotení.

Na štatistické spracovanie boli vytvorené štyri hypotézy.

H1: Jednoduchšie je ohodnotiť obraz vysokej alebo nízkej kvality ako obraz strednej kvality. Očaká sa nižšia hodnota intervalu spoľahlivosti pri obraze strednej kvality v porovnaní s obrazom vysokej alebo nízkej kvality.

H2: Spoľahlivosť probanta, kedy hodnotenie kvality obrazu súvisí s reakčnou dobou hodnotiaceho procesu. Očakáva sa dlhší reakčný čas na obrázky, ktoré boli ťažko hodnotiteľné.

H3: Probandov nesúhlas sa odráža v intervale spoľahlivosti. Je ovplyvnený mnohými faktormi a spoľahlivosť probanta je jedným z nich.

H4: Spôľahlivosť probanta možno predpovedať s rozumnou presnosťou na základe jeho daných hodnotení kvality a reakčného času.

Pre overenie týchto hypotéz bola vykonaná podrobná štatistická analýza.

3.3.1.1 Kendallov koeficient zhody W

Štatistickú závislosť môžeme merať nielen medzi číselnými hodnotami dvoch premenných ale aj medzi poriadami týchto hodnôt v celom súbore. Kendallov koeficient zhody W [4] je neparametrický štatistický test, ktorý predstavuje mieru zhody medzi probantmi, ktorí posudzujú rovnaké objekty.

Nech máme n pozorovaní a zisťovali sme hodnoty k probantov (x_1, x_2, \dots, x_k) . Pozorovania vyjadríme indexom i ($i = 1, 2, \dots, n$) a probantov indexom j ($j = 1, 2, \dots, k$). Napozorované hodnoty x_{ji} transformujeme na ich poradia $R_{x_{ji}}$. Pri každom pozorovaní zistíme celkový súčet poradí R_i vzť ahom

$$R_i = \sum_{j=1}^k R_{x_{ji}}. \quad (3.26)$$

Stredná hodnota týchto radov je definovaná vzť ahom

$$\bar{R} = \frac{1}{2}k(n+1). \quad (3.27)$$

Súčet štvorcov odchýlok, S , je definovaný

$$S = \sum_{i=1}^n (R_i - \bar{R})^2. \quad (3.28)$$

Kendallov koeficient zhody W potom vypočítame podľa vzť ahu

$$W = \frac{12S}{k^2(n^3 - n)}. \quad (3.29)$$

Ak Kendallov koeficient zhody W má hodnotu 1, znamená to dokonalú zhodu hodnôt medzi jednotlivými hodnoteniami probantov. Ak má hodnotu 0, znamená to, že neexistuje žiadna zhoda medzi hodnoteniami probantov, čiže ich odpovede možno považovať za úplne náhodne. Prechodné hodnoty W označujú väčšiu alebo menšiu mieru jednomysel'nosti medzi jednotlivými probantmi.

3.3.1.2 Výsledky

Experiment sa zameriaval na pravdivosť ľudského vnímania pri posudzovaní kvality obrazu. Pomocou experimentu bol určený vzájomný vzť ah medzi kvalitou vnímania, spoľ ahlivosťou a reakčného času. Podrobnou štatistickou analýzou sa potvrdila hypotéza

H1, ktorá hovorí, že je jednoduchšie ohodnotiť obraz vysokej alebo nízkej kvality ako obraz strednej kvality.

Hypotézu H2 , ktorá hovorila o tom, že reakčný čas obrázkov, ktoré boli ťako hodnotiteľné, je dlhší, potvrdzuje inverzná korelácia medzi reakčným časom a hodnotením spoľahlivosti. Bola nájdená podobná korelácia aj medzi hodnotením kvality a reakčným časom.

Hypotézu H3 potvrdzuje korelácia medzi hodnotením spoľahlivosti a štandardnou chybou. Nízka korelácia medzi hodnotením spoľahlivosti a štandardnou chybou znamená, že obsahuje aj iné ovplyvňujúce faktory, ktoré je treba ešte identifikovať.

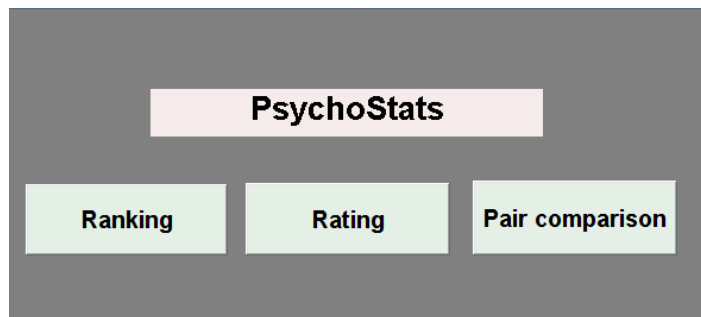
Nakoniec sa ukázalo, že spoľahlivosť probanta možno predvídať s dostatočnou presnosťou s použitím jednoduchých kombinačných modelov založených na hodnotení kvality a reakčného času, čo potvrdzuje hypotézu H4. Zahrnutie štandardnej chyby priemeru do predpovedaných modelov spôsobilo negatívny vplyv na predpovedné modely.

4. Implementácia

4.1 Psycho-Stats

Vybrané štatistické metódy, ktoré sme popísali v predchádzajúcej kapitole, sme aplikovali vo vyhodnocovacom štatistickom softvéri (ďalej len Psycho-Stats). Vybrané metódy jednotlivých experimentov sme implementovali vo vývojovom prostredí **Matlab r2009a** s využitím štatistického toolboxu. Na správne fungovanie Psycho-Stats je potrebné mať nainštalované vývojové prostredie Matlab r2009a a vyšší.

Vstupné dáta všetkých experimentov boli uložené v textovom súbore s koncovkou „*.psy“. Obsahovali meno experimentátora, názov experimentu, dátum vytvorenia experimentu, označenie obrázkov použitých v experimente a hodnotenia probantov. Po otvorení Psycho-Stats sa zobrazilo úvodné okno, ktoré ponúka experimentátorovi vybrať si štatistické spracovanie experimentu podľa jeho druhu, znázornené na obrázku 4.1. Kliknutím na jemu daný zvolený druh experimentu sa experimentátor dostáva do grafického rozhrania štatistického spracovania experimentu.



Obr. 4.1: Úvodná obrazovka softvéru Psycho-Stats.

4.1.1 Usporiadanie (ranking)

Na vyhodnotenie experimentu Usporiadanie sme si vybrali **jednofaktorovú analýzu rozptylu (ANOVA1)**. Na základe vstupných dát, ktoré boli spracovávané, bola táto metóda vybraná ako najvhodnejšia.

4.1.1.1 Vstup

Vstupné dáta experimentu Usporiadanie sú znázornené na obrázku 4.2. Hodnotenie jedného probanta bolo zaznamenané v jednom riadku, čo znamená, že každý nový riadok sa rovná hodnoteniu nového probanta. Probant mal v tomto experimente za úlohu zoradiť testovaciu sadu do jemu zvoleného poradia podľa otázky, ktorá bola zadaná experimentátorom.

```
Paula Budzaková Porovnávanie hrán 5.4.2014
I1 I2 I3 I4 I5 I6 I7 I8 I9
2 3 4 1 5 7 9 6 8
1 6 3 4 2 8 5 7 9
2 1 4 5 3 7 9 8 6
1 4 2 7 3 9 6 5 8
1 6 3 5 4 7 2 8 9
2 3 1 4 5 7 6 9 8
1 3 2 4 5 8 6 7 9
2 1 4 5 3 7 9 8 6
1 4 2 3 5 9 7 6 8
3 6 9 5 4 7 8 2 1
2 3 4 1 5 7 9 6 8
4 6 3 1 2 8 9 7 5
2 1 4 5 3 7 9 8 6
1 4 2 3 5 9 7 8 6
3 6 9 5 4 1 7 2 8
```

Obr. 4.2: Vstupné dáta pre experiment Hodnotenie.

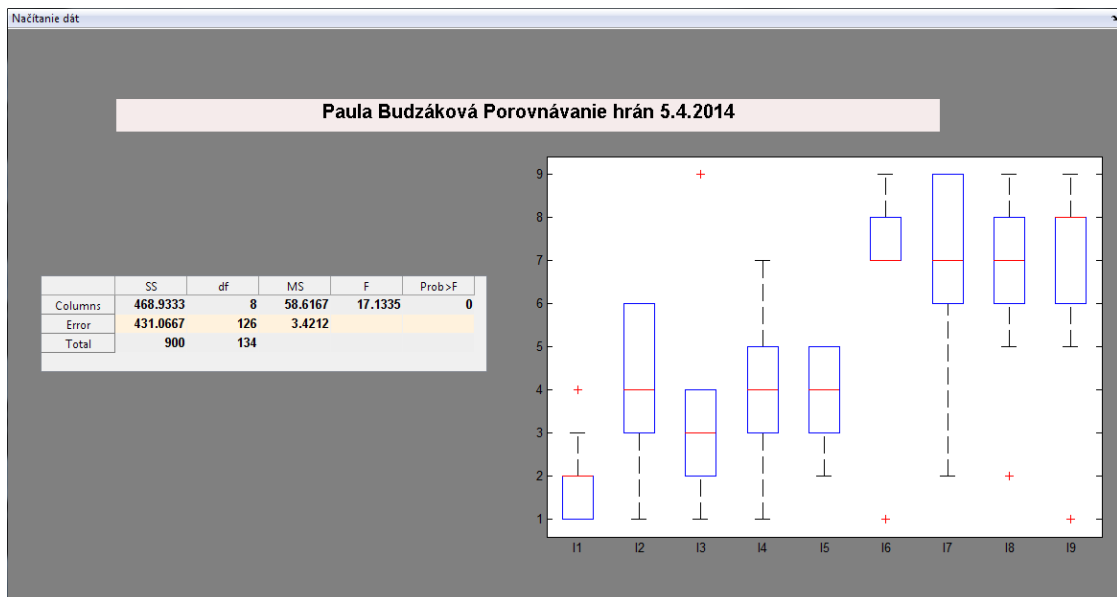
Každý textový súbor obsahoval informácie jedného celého experimentu, ktorý bol hodnotený niekoľkými probantmi. Každý nový experiment, zadaný experimentátorom, mal na výstupe vždy len jeden textový súbor, ktorý bol štatisticky spracovávaný.

4.1.1.2 Výstup

Po načítaní textového súboru dát sa nám dáta automaticky štatisticky spracovali pomocou funkcie

»`[p,table,stats]=anova1(X,group)`.

kde X je matica dát a $group$ je skupina. P vráti danú P-hodnotu, $table$ vráti štandardnú ANOVA tabuľku štatistických údajov a $stats$ štruktúru štatistiky na vykonanie viacnásobného porovnávacieho testu, ktorý sme ale v tomto experimente nevykonali.



Obr. 4.3: Štatistické spracovanie experimentu Usporiadanie.

Obrázok 4.3 zobrazuje štatisticky spracované dáta. Meno experimentátora, názov experimentu a dátum vykonania sa zobrazí v hornom riadku pre lepšiu orientáciu experimentátora.

Tabuľka ANOVA rozdeľuje rozptyl dát na dve časti

- rozptyl medzi jednotlivými úrovňami faktora, čiže rozptyl medzi stĺpcami,
- rozptyl vo vnútri jednotlivých úrovní, čiže rozptyl medzi údajmi v stĺpci.

Tabuľka ANOVA obsahuje šesť stĺpcov

1. zdroj rozptylu,
2. súčet štvorcov (SS) v dôsledku každého zdroja,
3. stupne voľnosti (df) spojené s každým zdrojom,
4. priemerný štvorec (MS) pre každý zdroj, čo je pomer (SS)/(df),
5. testovacia F-štatistika, čo je pomer priemerných štvorcov,
6. P-hodnota.

Graf jednotlivých stĺpcov dát naznačuje veľkosť F-štatistiky a P-hodnoty. Veľké rozdiely stredovej osy jednotlivých boxov zodpovedajú veľkej hodnote F-štatistiky a malej hodnote P-hodnoty. Na základe testovania hypotéz určíme vzťah medzi premennými a ich štatistickú významnosť.

4.1.2 Párové porovnanie (Pair comparison)

V Párovom porovnaní sme vyhodnocovali pomocou metódy **Kruskall-Wallisov test**. Túto metódu sme si vybrali z dôvodu, že umožňuje testovať hypotézu H_0 na $k(k \geq 3)$ nezávislých súborov toho istého rozdelenia. Keďže v tomto experimente netestujeme len jeden súbor vstupných dát, ale celú množinu, bola táto metóda na štatistické spracovanie najvhodnejšia.

4.1.2.1 Vstup

Textový súbor vstupných dát obsahoval hodnotenie len jedného probanta. Probant mal v tomto experimente za úlohu vybrať z dvojice obrázkov ten obrázok, ktorý sa mu zdal vhodnejším podľa zadanej otázky. Dáta v textovom súbore predstavovali maticu porovnaní jednotlivých obrázkov, ktorá je znázornená na obrázku 4.4.

```
Paula Budzaková Porovnanie šumu 20.4.2014
I1 I2 I3 I4 I5
0 2 1 1 2
0 0 0 1 2
1 1 0 1 1
1 0 0 0 0
0 0 0 1 0
```

Obr. 4.4: Vstupné dáta pre experiment Párové porovnanie.

Keďže každý textový súbor obsahoval dáta len jedného probanta, v Párovom porovnaní sa spracovávala celá zložka vstupných dát. Počet textových súborov sa rovnal počtu probantov, ktorí tento experiment vykonali.

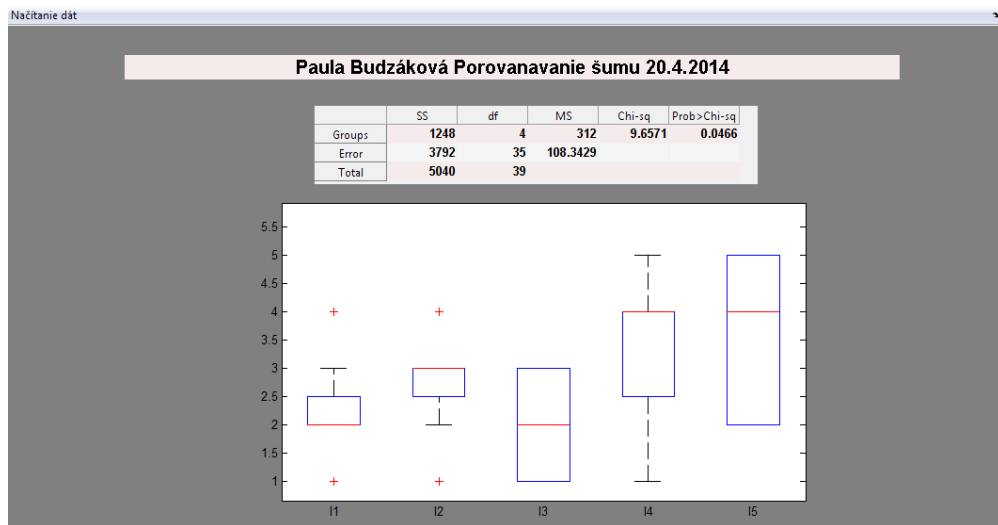
4.1.2.2 Výstup

Po načítaní zložky dát, sme si jednotlivé matice dát spojili do matice, ktorá obsahovala podmatice probantov. Každý stĺpec matice predstavoval počet preferovaní jedného obrázka pred ostatnými. Hodnoty v jednotlivých stĺpcoch boli sčítané, čím sme dostali maticu veľkosti $2 \times n$. Prvý riadok matice predstavoval označenia obrázkov. Druhý riadok obsahoval hodnoty celkových preferovaní každého obrázka u probantov. Takto upravené dáta sme štatisticky spracovali pomocou funkcie

»`[p,tbl,stats] = kruskalwallis(X,group)`,

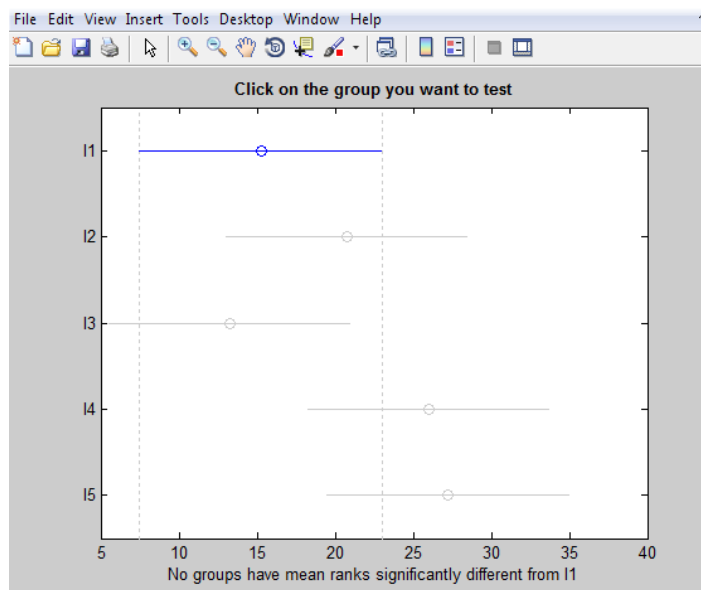
kde X je vektor dát a $group$ je skupina. Keďže Kruskal-Wallisov test je neparametrickou obdoba jednofaktorovej analýzy rozptylu, výstupné argumenty budú podobné ako v

predchádzajúcom experimente. P vráti P - hodnotu pre hypotézu H_0 , tbl vráti štandardnú ANOVA tabuľku a $stats$ štruktúru štatistiky na vykonávanie viacnásobného porovnávacieho testu, ktorý sme v tom experimente vykonali.



Obr. 4.5: Štatistické spracovanie dát experimentu Párové porovnanie.

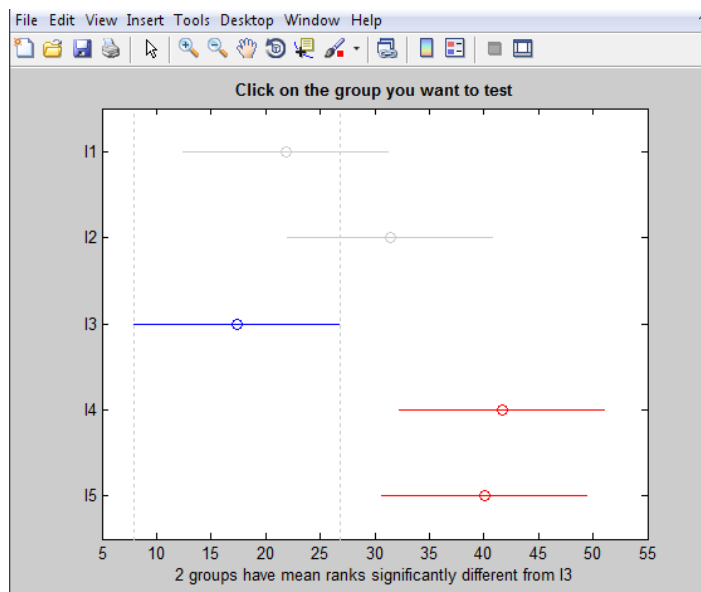
Meno experimentátora, názov experimentu a dátum vykonania sa zobrazili, už ako v predchádzajúcom experimente, v hornom riadku. ANOVA tabuľka vrátila *súčet štvorcov, stupne voľnosti, priemerný štvorec pre každý zdroj, chi-kvadrát rozdelenie a P-hodnotu*. Graf naznačuje veľkosť chi-kvadrátu rozdelenia a P -hodnoty. Stredová os jednotlivých boxov znázorňuje priemerný počet preferovaní jednotlivého obrázka. Symbol + znázorňuje extrémne hodnoty, ako na obrázku 4.5.



Obr. 4.6: Viacnásobný porovnávací test bez odlišných skupín.

V Kurskall-Wallisovom teste je možné porovnávať dáta niekoľkých skupín testovaním hypotézy H_0 , že sú všetky hodnotenia rovnaké oproti alternatívnej hypotéze H_α ,

že všetky hodnotenia nie sú rovnaké, čiže sú odlišné. Tieto informácie nám vráti viacnásobný porovnávací test. Viacnásobný porovnávací test používa informácie zo *stats*. Graf na obrázku 4.6 zobrazuje každú skupinu porovnavania na intervale spoľahlivosti vypočítaného z ich rozdielu. Ak interval spoľahlivosti neobsahoval nulu, rozdiel bol štatisticky významný. Pomocou kliknutia je možné vybrať si skupinu a graf bude upozorňovať na všetky ďalšie skupiny, ktoré sú od tejto odlišné, ako je zobrazené na obrázku 4.7.



Obr. 4.7: Viacnásobný porovnávací test s odlišnými skupinami.

4.1.3 Hodnotenie (Rating)

Experiment Hodnotenie sme štatisticky spracovali pomocou **Kendallovho koeficientu zhody W**. Kendallov koeficient zhody W (ďalej len W) je neparametrický test, ktorý predstavuje mieru zhody medzi probantmi, ktorí posudzujú rovnaké objekty.

4.1.3.1 Vstup

Textový súbor obsahoval dáta zobrazené na obrázku 4.8. Každý riadok dát predstavuje hodnotenie nového probanta. V experimente Hodnotenie má probant za úlohu ohodnotiť obrázok, podľa vybraného štandardu, na škále určenej experimentátorom.

Každý textový súbor obsahuje hodnotenia probantov len podľa jedného štandardu. Ak by sme chceli testovať hodnotenia viacerých štandardov, museli by sme mať zabezpečený rovnaký počet probantov v každom experimente, čo zatiaľ nemáme. Práve z tohto dôvodu bol W zvolený za najvhodnejší. Ak sa W rovnal 1 znamenalo to dokonalú zhodu medzi hodnoteniami probantov. Ak mal W hodnotu 0, hodnotenia sa mohli považovať za úplne

```

Paula Budzáková Hodnotenie jasu 20.4.2014
I1 I2 I3 I4 I5 I6 I7 I8 I9
2 3 6 4 8 9 4 5 1
3 1 5 7 6 3 4 9 1
2 4 6 1 2 8 6 7 4
2 5 4 8 7 5 9 1 3
1 3 2 5 4 6 5 8 5
2 4 3 5 4 6 5 1 3
7 6 8 4 3 4 1 8 5
3 4 2 6 5 7 4 9 6
2 3 1 1 5 4 7 1 1
1 4 3 6 2 8 7 9 5
1 3 4 5 2 9 8 5 6
2 8 7 6 9 4 5 2 1
1 5 4 7 6 2 4 9 7
1 4 3 5 6 2 3 8 6
1 4 3 6 2 8 7 9 5

```

Obr. 4.8: Vstupné dáta pre experiment Hodnotenie.

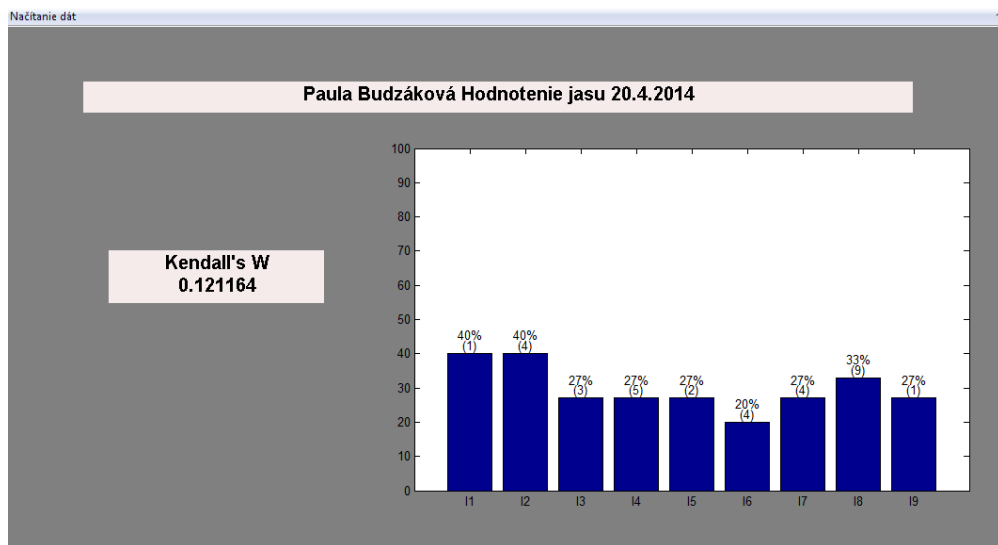
náhodné. Prechodné hodnoty naznačovali väčšiu alebo menšiu mieru jednomysel'nosti medzi hodnoteniami.

4.1.3.2 Výstup

Po načítaní dát sa dáta neupravovali, a hneď sa štatisticky spracovali pomocou funkcie

» **W = KendallCoef(X)**,

kde X predstavuje dáta na spracovanie. Keďže funkcia na výpočet W nebola súčasťou štatistického toolboxu, ktorý sme využívali, použili sme funkciu **KendallCoef(X)** vytvorenú *Lijie Huang 6.5.2010* pod licenciou, ktorá nám dovolila funkciu použiť. Táto funkcia nám vráti hodnotu W , na základe vstupných dát, ako na obrázku 4.9. Graf znázorňuje najviac preferovanú hodnotu škály pre každý obrázok hodnotený v experimente a percentuálny počet ľudí, ktorý touto hodnotou škály obrázkov ohodnotili.



Obr. 4.9: Štatistické spracovanie dát experimentu Hodnotenie.

4.1.4 Budúca práca

Do budúca by sme chceli implementovať lepšie štatistické metódy pre experiment Hodnotenie. Každý textový súbor dát by predstavoval jeden experiment hodnotení na základe jedného štandardu. Štatisticky by sa spracovávalo viacero textových súborov, ktoré by boli hodnotením podľa rôznych štandardov. Na takéto dáta potrebujeme upraviť aj softvér na spracovanie experimentov, aby sme zabezpečili, že každý textový súbor dát bude obsahovať rovnaký počet probantov. Na štatistické spracovanie by sa potom mohla použiť celá zložka takýchto súborov.

Takéto dáta by sme mohli spracovať **Kappa štatistikou** [6] prezentovanú **vnútrotriednym korelačným koeficientom** [6] (ďalej len VKK). VKK predstavuje veľmi kontroverznú mieru asociácie medzi dvomi nominálnymi meraniami rovnakých objektov. VKK by pre nás predstavoval mieru zhody dvoch probantov pri hodnotení obrázkov podľa rôznych štandardov. Ak by VKK vyšiel štatisticky významný, znamenalo by to, že hodnotenia probantov boli príliš zhodné na to, aby mohli byť len náhodné.

VKK sa môže použiť len pri väčšej vzorke dát. Vzorové dáta by mali obsahovať aj približne rovnaký počet hodnotiacich obrázkov na základe každého štandardu. Ak by bol počet obrázkov príliš odlišný, VKK by mohol indikovať rozdiel v zhode probantov, ktorý však nie je zapríčinený samotnými probantmi.

5. Záver

V našej práci sme sa venovali psychofyzikálnym experimentom z ich štatistického hľadiska. Najprv sme opísali psychofyziku ako vedu, základné psychofyzikálne prahy, škálovanie a ich metódy. Potom sme priblížili jednotlivé psychofyzikálne experimenty z reálnej praxe. V každom experimente sme popísali metódy, ktoré sme si vybrali na ich spracovanie.

V experimente *Usporiadanie* sme uviedli *jednofaktorovú analýzu rozptylu*. Táto metóda nám spracovávala dáta získané na základe hodnotenia len jedného faktora. Uviedli sme si, ako táto metóda postupne dáta spracováva a kedy sú pre nás výsledky štatisticky významné. Taktiež sme si priblížili *testovanie štatistických hypotéz* a výpočet *P-hodnoty*. Reálny priebeh experimentu Hodnotenia sme znázornili v psychofyzikálnom experimente. V *Párovom porovnaní* sme popísali dva reálne experimenty, ktoré na svoje hodnotenie využívali *Interval spoľahlivosti*, *chi-kvadrát dobrej zhody* a *Kurskall-Wallissov test*, ktorý je obdobou jednofaktorovej analýzy rozptylu. V *Hodnotení* sme popísali jeden reálny experiment a *Kendallov koeficient zhody W*.

Implementácia metód v softvéri Psycho-stats dopadla podľa predstáv. Softvér je plne funkčný a použiteľný v praxi. Metódy, ktorým sme sa venovali, sa nám podarilo aplikovať na potrebných dátach. Pomocou softvéru vieme povedať, či hodnotenia probantov boli pre nás štatisticky významné alebo nie. V budúcnosti by sme chceli softvér rozšíriť o metódy, ktoré dokážu spracovávať aj väčšie a rôznorodejšie dáta.

Bakalárska práca splnila svoj cieľ. Dozvedeli sme sa nové poznatky z odvetvia psychofyziky a štatistiky. Psychofyzikálne experimenty nám ukázali svoju veľmi dobrú použiteľnosť v odvetví spracovania obrazu. Na základe hodnotenia faktorov probantmi vieme spracovanie obrazu otestovať v praxi a využiť na možné vylepšenia alebo zmeny.

Literatúra

- [1] Boroš Július a kol. *Experimentálna psychológia*. Bratislava: SNP, 1986.
- [2] Ahmet Oguz Akyuz, Roland Fleming, Erik Reinhar. Do HDR displays support LDR content? a psychophysical evaluation. *ACM SIGGRAPH conference*, pages 1–7, June 2007.
- [3] Anna Tirpáková, Dagmar Markechová, Beáta Stehlíková. *Štatistické metódy a ich aplikácie*. Univerzita Konštantína Filozofa v Nitre, 2011.
- [4] Jozef Chajdiak. *Štatistika jednoducho*. STATIS, 2010.
- [5] James A. Ferwerda. Psychophysics 101: how to run perception experiments in computer graphics. *GRAPH International Conference on Computer Graphics and Interactive Techniques*, pages 1–60, 2008.
- [6] Fleiss Joseph L., Bruce Levin, Myunghee Cho Paik. The measurement of interrater agreement. *Statistical methods for rates and proportions 2*, pages 212 – 236, 1981.
- [7] Elena Šikudová. *Prednášky z predmetu rozpoznávanie obrazcov*, 2013.
- [8] Jíří Militký, Milan Maloun. *Kompendium statistického zpracování dat*. Univerzita Karlova v Prahe - Nakladatelství Karolinum, 2012.
- [9] Bardin K.V. *Problémy prahov citlivosti a psychofyzické metody*. Bratislava : SNP, 1981.
- [10] Mark D. Fairchild, Karen M. Braun. Testing five color-appearance models for changes in viewing conditions. *COLOR research and application*, pages 165–173, June 1997.
- [11] Petr Klímeck, František Pavelka. *Aplikovaná štatistika*. Vysoké učení technické v Brne, 2000.
- [12] Marián Rimarčík. *Štatistický navigátor*, 2000.
- [13] Shangfei Wang, Xufa Wang. Evaluating emotional algorithms using psychological scales. *SIGGRAPH '08 ACM SIGGRAPH 2008 classes*, (11):1–60, 2008.

- [14] Katina Stanislav. *Vybrané kapitoly z počítačovej štatistiky 1*. Univerzita Komenského v Bratislave, 2006.
- [15] Ulrich Engelke, Anthony Maeder, Hans-Jurgen Zepernick. Human observer confidence in image quality assessment. *Signal Processing: Image Communication*, 27(9):935–947, 2007.

Prílohy

CD obsahujúce:

- Elektronická verzia Bc. práce
- Zdrojové kódy softvéru