

DE INVLOED VAN INSTANCE LEARNING BIJ ABSOLUTE IDENTIFICATIE VAN TIJDSINTERVALLEN DOOR MENSEN

Bachelorproject

Theije Visser

Samenvatting: Wij hebben een model in ACT-R ontwikkeld die een timingmodule combineert met een geheugengebaseerd schalingsmodel (gebaseerd op het anchor-model van Petrov & Anderson, 2005). Hiervoor is een experiment gedaan waarbij proefpersonen als taak hadden om aangeboden tijdsintervallen te categoriseren in een oplopende schaal. Het model vergelijkt de ontvangen magnitudes van stimuli met *chunks* in het geheugen. Elke unieke combinatie van een tijdsinterval met een categorie wordt opgeslagen in het geheugen. Tijdsintervallen worden door het model gecategoriseerd door lokale vergelijking met voorgenoemde chunks; een voorbeeld van *instance learning*. Uit de resultaten van het experiment bleek dat het model en de experimentele resultaten redelijk overeenkomen. De representatie van tijd bij mensen is redelijk goed uit het model te verklaren door gebruik te maken van het instance learning-mechanisme. Het ontwikkelde temporele schalingsmodel is een opzet voor later onderzoek en er is nog veel ruimte voor verbetering.

1. Inleiding

Het vermogen om korte tijdsintervallen te schatten speelt een belangrijke rol in ons dagelijks leven. Tijdschattingen zijn belangrijk in situaties waarin we handelen en een reactie verwachten, bijvoorbeeld als we moeten beoordelen of we wel of niet moeten remmen voor een oranje verkeerslicht. Dit is een voorbeeld van *prospective time interval estimation*; waar bij het begin van een interval bekend is dat een schatting moet worden gemaakt. Dit soort van tijdschatting is in het echte leven waarschijnlijk impliciet, automatisch en nauw verweven met andere soorten van cognitie, zoals perceptie, leren en beslissingen nemen.

Dit onderzoek is gericht op de wijze waarop tijdsintervallen worden gerepresenteerd in de hersenen. We kijken of zo'n representatie vergelijkbaar is met de representatie van andere sensorische modaliteiten, zoals de perceptie van *ruimtelijke* vormen. Om het mechanisme achter de categorisatie van tijdsintervallen te onderzoeken is cognitieve modellering een veelgebruikte benadering. We hebben een temporeel schalingsmodel gemaakt in de cognitieve architectuur ACT-R (Anderson & Lebiere, 1998). Dit model combineert theorieën van timing en anchoring. Hiervoor zijn twee reeds bestaande en geteste modellen gecombineerd tot een model die het mechanisme van prospectieve categorisatie van

tijdsintervallen door mensen reproduceert. Dit cognitieve model moet zo exact mogelijk de vele fenomenen reproduceren die voorkomen bij menselijke tijdsschatting. Aan de hand van de mechanismen van het model die de fenomenen veroorzaken kunnen we assumpties maken over de oorsprong van de fenomenen. Het uiteindelijke doel is een dieper inzicht in de cognitieve mechanismen die mensen gebruiken bij prospectieve tijdsschatting.

Bij de schatting op grootte van eendimensionale, geordende domeinen speelt vergelijking een grote rol. Alle schattingen worden gedaan relatief aan alle andere objecten in het domein. *Anchoring* (Tversky & Kahneman, 1974) is een cognitieve bias die de menselijke neiging beschrijft om hun oordeel te laten afhangen van recentelijk aangeboden informatie. ANCHOR (Petrov & Anderson, 2000; Petrov, 2001; Petrov, 2003; Petrov & Anderson, 2005) is een implementatie in ACT-R die psychofysische perceptuele theorieën en theorieën van het menselijke geheugen combineert tot een enkel mechanisme dat in staat is om unidimensionale continua op grootte in te schatten. De theorie bestaat grofweg uit drie onderdelen. Het eerste onderdeel is de perceptuele module die input codeert als een interne representatie; de *magnitude*. Het tweede onderdeel is geheugengebaseerd. Hierbij is de aanname dat het geheugen bestaat uit chunks die *anchors*

worden genoemd. Deze chunks bevatten een categorie en de bijbehorende magnitude. De magnitude wordt vergeleken met aanwezige anchors in het geheugen, op basis van deze vergelijking en activatiemechanismen wordt een winnende anchor en de daarbijbehorende categorie gekozen. Het laatste onderdeel is een correctiemechanisme. De grootte van de gekozen categorie wordt nog eens vergeleken met de inputgrootte. Bij een te groot verschil wordt de categorie naar boven of beneden bijgesteld.

Onderzoek naar tijdsschatting bij mensen is niets nieuws. De rol van de context van het geschatte tijdsinterval speelt hierbij een belangrijke rol. Al anderhalve eeuw geleden publiceerde Karl von Vierordt zijn boek *Der Zeitsinn nach Versuchen* (1868), waarin werd geobserveerd dat mensen bij het identificeren van tijdsintervallen een structurele bias naar het midden van het domein van aangeboden tijdsintervallen vertonen. Naast het onderzoek naar de (context-) effecten die volgen uit menselijke tijdsschattingen is ook veel onderzoek gedaan naar het biologische mechanisme achter ons gevoel van tijd. Dit ligt niet binnen het bereik van dit onderzoek. Wat we wel kunnen doen is een biologisch plausibel mechanisme implementeren in een cognitief model. Hier zijn al reeds meerdere implementaties voor beschikbaar. We gebruiken een mechanisme ontwikkeld door Taatgen, Van Rijn & Anderson (2005), die hun model hebben gebaseerd op een pacemaker-gebaseerde interne klok die tijd modelleert als *ticks*.

Tussen het ANCHOR-model en ons *temporele schalingsmodel* zijn enkele belangrijke verschillen te onderscheiden. Het ANCHOR-model maakt gebruik van een lokaal correctiemechanisme, dat voor ons model niet cognitief plausibel is, dit mechanisme is daarom ook niet geïmplementeerd. Verder gebruikt ons model een dynamisch aantal anchors, in tegenstelling tot ANCHOR, waarbij het aantal anchors vaststaat.

We hebben zelf een experiment opgezet naar voorbeeld van Petrov & Anderson (2005). Uit het model kunnen we bestaande, bekende effecten bevestigen. Maar belangrijker is dat we het model kunnen testen door de experimentele

resultaten te vergelijken met de gemodelleerde resultaten.

1.1. Overzicht

Het belangrijkste doel van dit onderzoek is om een geheugengebaseerd schalingsmodel voor tijdsintervallen te implementeren en te testen. We streven naar een integratie van psychofysische schaling met een geheugenfunctie, door het combineren van een temporele module met een geheugengebaseerde schalingsmodule. Hierdoor verkrijgen we een temporeel schalingsmodel. Het model wordt getest aan de hand van menselijke data en de resultaten van het originele model. Het originele model, genaamd ANCHOR (Petrov & Anderson, 2005), is gericht op schaling van geordende, unidimensionele continua, waaronder ook tijdsintervallen vallen.

Het temporele schalingsmodel is gebaseerd op het ANCHOR-model. Zoals gezegd zijn er enkele veranderingen doorgevoerd ten opzichte van het ANCHOR-model. Het temporele schalingsmodel kan worden gezien als een cognitief plausibelere versie van het ANCHOR-model aangepast op het temporele domein. Beide modellen bevatten gelijkende mechanismen, daarom verwachten we ook dat beide resultaten nagenoeg gelijkblijven. Hieruit kunnen we hypothetiseren dat de resultaten van het temporele schalingsmodel net zo goed overeen moeten komen met de experimentele data als het geval was bij het ANCHOR-model.

Naast de implementatie en het testen van het model willen we uit de experimentele resultaten de bekende (context- en transfer-) effecten die voorkomen bij tijdsschatting bevestigen. Door naar de implementatie van het model te kijken, kunnen we een indicatie verkrijgen van de mechanismen die sommige effecten veroorzaken.

Om het temporele schalingsmodel goed te kunnen evalueren en te kunnen vergelijken met het originele ANCHOR-model, hebben we een experiment gereproduceerd uit Petrov & Anderson (2005), die gebruikt is om het originele model te testen.

2. Experimentele opzet

We hebben een experiment uitgevoerd waarbij proefpersonen tijdsintervallen absoluut moesten identificeren. Absolute identificatie houdt in dat de aangeboden stimuli elk een label krijgen: de categorie. Er zijn dus evenveel categorieën als stimuli (zie tabel 2.1). Daarnaast wordt na elke trial (aangeboden stimulus en respons) feedback gegeven in de vorm van de juiste categorie. Dit experiment is een reproductie van een experiment uitgevoerd door Petrov (Petrov & Anderson, 2005, p. 386), waarbij het om absolute identificatie van de lengte van lijnstukken ging (zie ook de methode-sectie).

Tabel 2.1: Alle stimuli met bijbehorende categorie. Responses worden gegeven in categorieën.

Categorie	Interval (ms)
1	500
2	750
3	1000
4	1250
5	1500
6	1750
7	2000
8	2250
9	2500

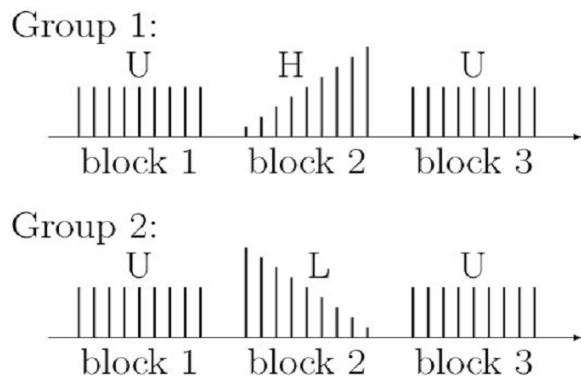
Er zijn enkele aanpassingen doorgevoerd ten opzichte van het experiment van Petrov. Ten eerste zijn de stimuli veranderd van ruimtelijk naar temporeel. De opzet blijft verder hetzelfde; proefpersonen moeten nog steeds de grootte van stimuli schatten, de stimuli blijven geordend en evenredig. De tweede aanpassing is kwantitatief. In plaats van 5 blokken van 90 stimuli aan te bieden, zoals Petrov, hebben we gekozen voor 3 blokken van 90 stimuli. Deze keuze is gebaseerd op resultaten van een pilot-experiment. Proefpersonen gaven aan dat de concentratie zeer moeilijk langer dan een half uur vast te houden is. Omdat de aanname was dat concentratieverlies een te grote invloed heeft op de prestatie van proefpersonen, is gekozen voor

3 blokken. Het experiment werd hierdoor ingekort tot ongeveer dertig minuten. We hebben verder gekozen voor het visueel aanbieden van stimuli. We reproduceren hiermee het experiment van Petrov het beste, aangezien hier ook visuele stimuli werden gebruikt.

Het doel van dit experiment is om een dataset te verkrijgen die te vergelijken is met het experiment van Petrov. Daarnaast willen we bekende context- en transfereffecten voor tijd en geordende continua afleiden. Als deze effecten zijn geverifieerd, zijn de resultaten te gebruiken als test voor het model. Om deze context-effecten te veroorzaken in de resultaten, is het nodig dat we niet-uniforme stimulusdistributies gebruiken (zie figuur 2.1). We gebruiken drie soorten distributies: uniform (U), scheef en oplopend (H, hoog) en scheef en aflopend (L, laag). Context-effecten kunnen ons meer vertellen over de verdeling van de interne representaties van tijdsintervallen.

2.1. Effecten

Context-effecten worden veroorzaakt door de relatie van stimuli met gehele spectrum van stimuli waarin ze voorkomen. De beoordeling van een stimulus wordt dus niet alleen veroorzaakt door zijn absolute eigenschappen, maar hangt ook af van het bereik, de frequentie en de grootte van andere stimuli aangeboden in de experimentele opzet. We hebben het hier dus over effecten *binnen* een blok van aangeboden stimuli. Er bestaan twee soorten contexteffecten: *assimilatieve* contexteffecten en *compenserende* contexteffecten. Bij assimilatieve contexteffecten worden de responses richting het dichtstbevolkte einde van het spectrum van stimuli getrokken, bij een scheve verdeling van stimuli wordt de verdeling van responschaal nog schever. Als de verdeling van de responschaal minder scheef is dan de verdeling van de stimuli, spreken we over een compenserend contexteffect (Petrov & Anderson, 2005).



Figuur 2.1: *Presentatieschema's van uniforme (U), lage (L) en hoge (H) blokken. De hoogte van de lijnen geeft de frequenties van aangeboden stimuli binnen een blok aan. Uit: Petrov & Anderson (2005).*

Vierordt's law is het schoolvoorbeeld van een assimilatief contexteffect die alleen bij de identificatie van tijdsintervallen voorkomt (Bobko et. al, 1977). Naast effecten die exemplarisch zijn voor de identificatie van tijdsintervallen, zijn we geïnteresseerd in effecten die ook voorkomen bij identificatie van ruimtelijke objecten, zoals bij Petrov's lijnenidentificatie. Sequentiële effecten zijn contexteffecten die vooral de anchoring-hypothese kunnen bevestigen; hier wordt gekeken of de vorige stimulus invloed heeft op het verschil tussen de huidige respons en stimulus, dat wil zeggen de precisie van de schatting.

Verder wordt gekeken of de besproken contexteffecten worden overgedragen *tussen* blokken. Als bijvoorbeeld de gemiddelde respons door een scheef blok omhooggaat (een assimilatief context effect), blijft het gemiddelde dan hoger in het volgende uniforme blok?

2.2. Methode

Vierentwintig studenten van de Rijksuniversiteit Groningen in de leeftijd van 18-25 jaar werkten mee aan het experiment. Twaalf werden willekeurig aan groep 1 toegewezen, de andere 12 werden aan groep 2 toegewezen.

De stimuli waren tijdsintervallen weergegeven door een grijze bol tegen een witte achtergrond

op een 19-inch CRT scherm. Op het moment dat de bol op het scherm verscheen ving het interval aan, op het moment van verdwijnen eindigde het interval. De tijd tussen het verschijnen en verdwijnen van de bol werd afgewisseld. De set van stimuli bestond uit negen tijdsintervallen van de volgende lengtes: 500, 750, 1000, 1250, 1500, 1750, 2000, 2250, 2500 milliseconden.

De proefpersonen werden uitgelegd dat er negen stimuli en evenveel categorieën zijn om als antwoord te geven en dat het hun taak was om de stimuli in te delen in een categorie met een cijfer van 1 tot en met 9. Daarnaast werd hen verteld dat er 10 punten te verdienen waren voor een goed antwoord, 5 punten als ze het tijdsinterval één te groot of de klein schatten en 1 punt als ze er twee naast schatten. Ook werd expliciet vermeld dat de proefpersonen geen strategieën (zoals meetellen) mochten gebruiken voor het inschatten van de lengtes van de tijdsintervallen. De stimuli werden achter elkaar weergegeven op basis van de hieronder weergegeven schema's. Bij elke trial werd voor het verschijnen van de bol een ster (*) gedurende 500 milliseconden weergegeven op de plaats waar de bol ging verschijnen. Na het verdwijnen van de bol werd steeds een vraagteken (?) weergegeven op de plaats waar de bol zich eerst bevond. Op dat moment kon de proefpersoon zijn antwoord intypen met de cijfertoetsen op het alfanumerieke gedeelte van een toetsenbord. Hier stond geen tijdslimiet op. De rest van de toetsen waren geblokkeerd. Nadat de proefpersoon zijn antwoord intoetste werd er feedback gegeven door gedurende 1300 milliseconden de juiste categorie van het voorgaande tijdsinterval en de huidige score op het scherm te tonen. Hierna begon direct de volgende trial door weer een ster op het scherm te tonen.

Er waren 17 demonstratie-trials en 270 experimentele trials. Tijdens de demonstratie-trials werden alle stimuli eerst in oplopende volgorde en daarna in aflopende volgorde weergegeven: 1, 2, 3, ..., 8, 9, 8, ..., 1. Hierbij waren de omstandigheden gelijk aan de experimentele trials, de proefpersonen kregen feedback en moesten een antwoord intypen. Na

deze demonstratie-trials werd aangegeven dat het daadwerkelijke experiment ging aanvangen.

De experimentele trials waren opgedeeld in 7 perioden met pauzes na 56, 112, 157, 202 en 247 trials (van de in totaal 287 trials). Tijdens elke trial werden de goede categorie, het antwoord van de proefpersoon en de reactietijd van de proefpersoon geadmistreerd door de computer. Het gehele experiment duurde ongeveer 30 minuten. Na het experiment werd aan de proefpersonen gevraagd hoe moe ze zich voelden, hoe ze het experiment ervaarden en hoe ze te werk gingen tijdens het schatten van elk interval.

De 270 experimentele trials werden ingedeeld in 3 blokken (zie figuur 2.1) van 90 trials. Bij groep 1 was het presentatie-schema 'UHU': het eerste blok heeft een uniforme verdeling (U), het tweede blok een oplopende verdeling (H) en het laatste blok weer een uniforme verdeling. Bij groep 2 was het presentatie-schema 'ULU': het eerste blok heeft een uniforme verdeling, het tweede blok een aflopende verdeling (L) en het laatste blok weer een uniforme verdeling.

Bij een uniform blok werden alle negen tijdsintervallen 10 maal gepresenteerd. Bij een oplopend blok werden de stimuli 1, 2, ..., 9 respectievelijk 18, 16, 14, ..., 2 maal gepresenteerd. Bij een aflopende blok werd het tegenovergestelde gedaan. Hierbij moet worden opgemerkt dat de tijdsintervallen in alle soorten blokken in een willekeurige volgorde werden getoond.

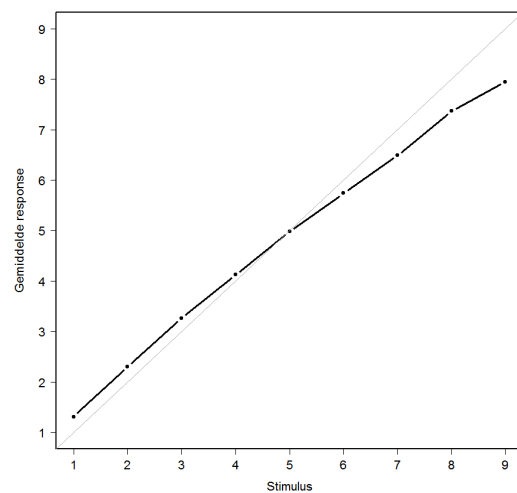
3. Resultaten

Bij het analyseren van de empirische data is vooral gelet op de stimulus en de daarbijbehorende respons van proefpersonen. Naast deze data zijn ook reactietijden bijgehouden. Er is gemiddeld over alle proefpersonen, per groep of per blok waar nodig. De gehele dataset bestaat uit 6480 trials, waaruit dus ook 6480 stimulus-respons paren zijn te extraheren. Er zijn verschillende fenomenen in de data onderzocht, naar inspiratie van Petrov & Anderson (2005), voor elk

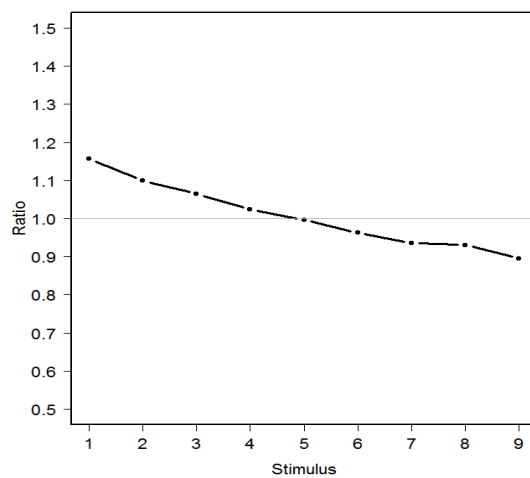
fenomeen is een verschillende analyse-benadering gebruikt.

3.1. Contexteffecten per stimulus

Omdat de stimuli-schaal relatief klein is kunnen we voor elke stimulus de gemiddelde bijbehorende responsen analyseren over alle proefpersonen. Elke stimulus is in totaal 720 keer aangeboden. In figuur 3.1 is een grafiek geplott die de gemiddelde respons per stimulus voor alle proefpersonen en alle blokken weergeeft. De grijze lijn geeft de hypothetische grafiek weer waarbij de gemiddelde respons gelijk is aan de stimulus.



Figuur 3.1: Gemiddelde respons per stimulus (categorie).



Figuur 3.2: Ratio per stimulus

In figuur 3.2 staat de gemiddelde ratio (Bobko et al, 1977) per stimulus-lengte geplotted. Deze ratio voor een trial i wordt als volgt berekend aan de hand van stimulus S en respons R , met tijdsinterval t in milliseconden:

$$ratio_i = t(R_i) / t(S_i) \quad (3.1)$$

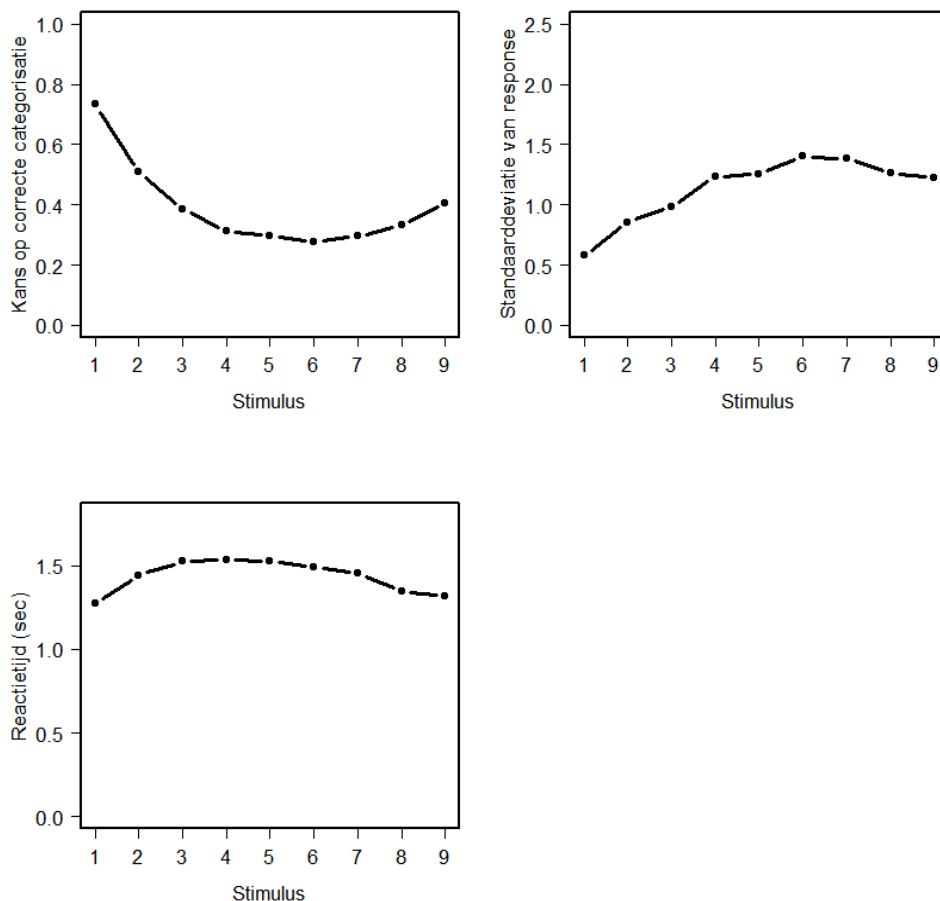
De ratio is dus de afwijking van het interval behorende bij de respons-categorie relatief aan het stimulus-interval. Een ratio van 1.00 betekent een perfecte schatting, een ratio onder de 1.00 staat voor een onderschatting, een ratio boven de 1.00 staat voor overschatting. Vooral in figuur 3.2 is een duidelijk een schuine verdeling te zien, met ratio's boven de 1.00 voor intervallen kleiner dan categorie 5, en ratio's onder de 1.00 voor intervallen groter dan categorie 5. Opvallend in beide grafieken is de snijding van de grafiek met

de rechte lijn op stimulus 5. De ratio voor stimulus 5 is $S(5) = 0.996$.

3.2. Edge effecten

Naast het kwantitatief analyseren van de responses per stimulus kunnen we ook kwalitatief gaan kijken naar de stimulus-respons paren. We willen weten wat de invloed van de seriële positie van een stimulus op de schaal is op de prestatie van proefpersonen.

In figuur 3.3 staan vier verschillende analyses geplotted. In eerste oogopslag is in alle vier de grafieken een verschil te zien tussen de randen van de schaal en het midden. Met randen bedoelen we hierbij stimulus 1 en stimulus 2 aan het begin en stimulus 8 en stimulus 9 aan het eind van de schaal.



Figuur 3.3: Edge effecten. Linksboven: Kans op correcte identificatie per stimulus. Rechtsboven: Standaarddeviatie van responses per stimulus. Linksonder: Gemiddelde reactietijd per stimulus.

Voor de grafiek linksboven is het aantal correcte responses gedeeld door het totaal aantal responses van alle proefpersonen. Merk hierbij op dat alle stimuli in totaal 720 keer zijn aangeboden. De grafiek voor de kans op correcte categorisatie is assymetrisch. Er is een indicatie voor een significant betere prestatie op de uiteinden van de schaal, vooral bij stimulus 1 en 2.

Voor de grafiek rechtsboven is de standaarddeviatie van de respons voor elke stimulus berekend voor alle proefpersonen. Deze data wijzen op een verminderde discriminabiliteit in het midden van de stimulus-schaal. Net als bij de proportie-grafiek is deze grafiek assymetrisch. Er is een indicatie op een sterk verbeterde discriminabiliteit voor stimulus 1 en 2, in mindere mate geldt dit voor stimulus 8 en 9.

Als laatste is de reactietijd gemeten per stimulus. Deze vertoont net als de twee bovenste grafieken een boog. De reactietijden voor stimuli aan de randen zijn lager in vergelijking met de reactietijden in het midden.

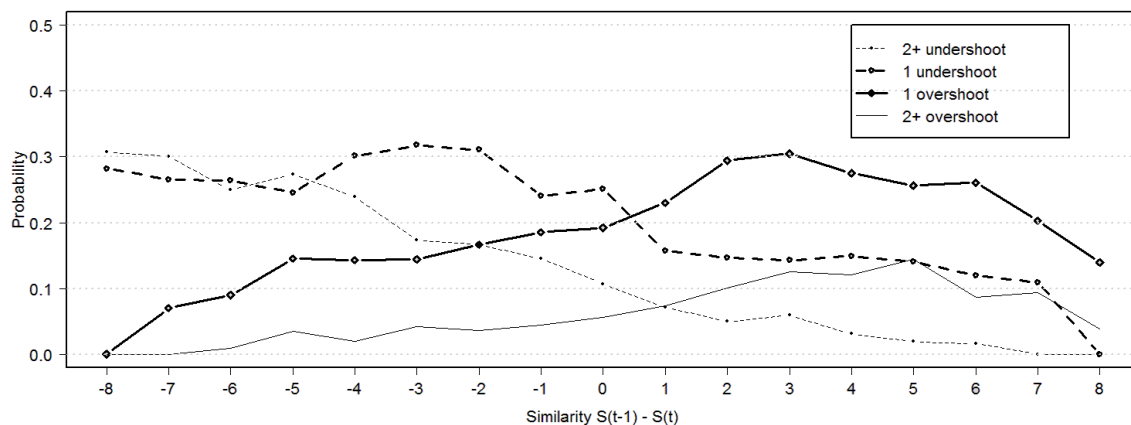
3.3. Invloed van vorige stimulus op huidige respons

Aan de hand van de methode van Luce *et al.* (1982) zijn verschillende soorten schattingsfouten bekeken per verschil tussen de huidige en vorige stimulus. Er is gekeken naar twee soorten fouten. *Undershoot 1 errors* zijn fouten waarbij respons 1 lager is dan de juiste categorie. *Undershoot 2+ errors* zijn de rest van de onderschattingsfouten.

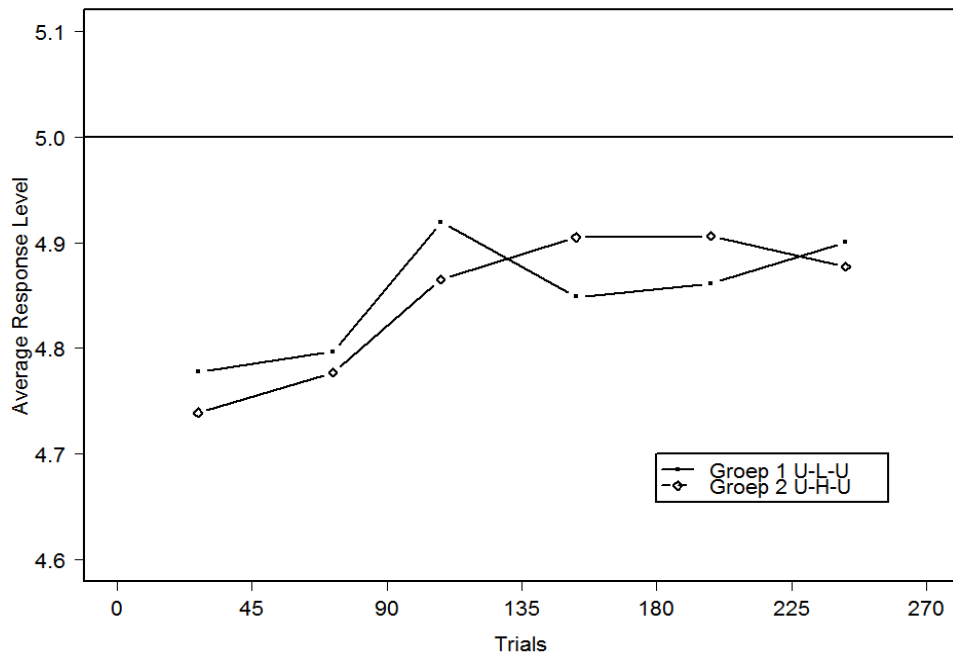
Voor *overshoot errors* geldt hetzelfde, maar dan voor gevallen waarbij de respons hoger is dan de juiste categorie. De grafieken worden uitgezet tegen het verschil $\Delta S = S_{(t-1)} - S_t$ tussen trials t-1 en t.

Figuur 3.4 laat zien dat alle vier de curves positief evenredig zijn met het verschil van het vorige aangeboden tijdsinterval. De kans op een overshoot error zit tussen 0 en 0.2 als de vorige stimulus kleiner is dan de huidige stimulus. Als de vorige stimulus groter is dan de huidige stimulus, is de kans op overshoot groter, namelijk tussen ongeveer 0.1 en 0.3. We zien verder een duidelijke afname in kans op overshoot als $\Delta S > 7$ (merk hierbij op dat $\Delta S > 7$ alleen voorkomt als de huidige stimulus 1 of 2 is).

Als het vorige interval kleiner is dan het huidige interval resulteert dit in een grotere kans op een undershoot error. Er is een duidelijke stijging in de kans op een undershoot error zichtbaar wanneer $S(t)$ groter wordt dan $S(t-1)$. Verder zien we dat de 2+ undershoot curve blijft toenemen in tegenstelling tot de 2- undershoot curve. Als $S(t-1)$ veel kleiner is dan $S(t)$ ($\Delta S < -3$), is te zien dat de kans op grote fouten snel toeneemt (merk hierbij op dat $\Delta S < -3$ alleen voorkomt voor grote langere intervallen (>4)).



Figuur 3.4: Sequentiële effecten: Kansen op verschillende soorten fouten uitgezet tegen de gelijkheid van de huidige stimulus met de vorige stimulus.



Figuur 3.5: Average response level over 6 delen van het gehele experiment.

3.4. Context-effect: Average response level

Om de invloed van context op de geheugentoestand en responses te meten is een methode uit Petrov & Anderson (2005) gebruikt *average response levels* (ARL) te berekenen voor responses in een bepaalde tijdperiodes. De ARL kan worden gezien als de gemiddelde respons op de gemiddelde stimulus van de data die worden geanalyseerd en wordt als volgt berekend:

$$\begin{aligned}
 ARL &= R_0 + a(S_{min} + S_{max})/2 \\
 &= R_0 + 5 \cdot a
 \end{aligned}
 \quad (3.2)$$

Door lineaire regressie uit te voeren op de deel van de data waar de ARL voor wordt berekend, kunnen de intercept R_0 en helling a worden berekend. Voor onze data hebben we de experimentele data in 6 stukken verdeeld, elk stuk is 45 trials lang. De zes ARLs zijn in figuur 3.5 voor beide groepen afzonderlijk afgebeeld.

Over de gehele data gezien valt op dat de ARL onder de 5.0 blijft. Dit betekent dat voor het gehele experiment geldt dat de gemiddelde respons altijd lager is dan de gemiddelde stimulus. Hieruit valt af te leiden dat proefpersonen de neiging hebben om vaker

stimuli te onderschatten dan te overschatten of juist te schatten.

Verder is een duidelijk stijgende lijn te zien voor beide groepen, waardoor valt af te leiden dat, naarmate het experiment vordert, proefpersonen stimuli langer gaan inschatten. De toename van de ARL tussen trials 0-45 en trials 225-270 is van 4.78 naar 4.90 voor groep 1 en van 4.74 naar 4.88 voor groep 2

Verder is te zien dat in het tweede deel van het tweede blok (trials 135-180) de grafieken elkaar snijden. Voor groep 1 is dit het lage (L) blok; de ARL neemt af van 4.92 naar 4.85. Voor groep 2 is dit het hoge (H) blok; de ARL neemt toe van 4.87 naar 4.90. De ARL voor beide groepen blijft aan het begin van het derde (uniforme) blok op hetzelfde niveau, waarna de ARLs weer convergeren.

4. Model

4.1. ANCHOR

Bij de implementatie van het model is gewerkt aan de hand van het ANCHOR-model (Petrov & Anderson, 2005) voor het categoriseren van lijnstukken. Hieronder een korte uiteenzetting van de werking van dit model.

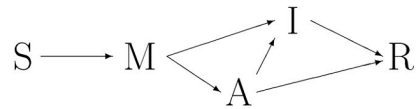
ANCHOR bestaat uit twee subsystemen die communiceren aan de hand van interne magnitudes (zie figuur 4.1). Dit systeem is te beschrijven aan de hand van drie computationele mechanismen wat betreft het absolute identificatie-experiment.

- Perceptuele module
- Anchor-selectiemechanisme
- Correctiemechanisme

Een stimulus S wordt aangeboden aan het model (zie figuur 4.1) De perceptuele module maakt stochastisch een magnitude M aan. Het selectiemechanisme kiest stochastisch de winnende anchor A aan de hand van vergelijking van M met aanwezige anchors in het declaratief geheugen, alsook de activatie van de anchors. Het declaratief geheugen bestaat uit evenveel anchors als er categorieën zijn, in dit geval negen. Het correctiemechanisme vergelijkt hierna anchor-magnitude A met de doel-magnitude M, om te bepalen of een correctie I nodig is. Als M en A te veel verschillen wordt M naar boven of beneden bijgesteld en de uiteindelijke respons R geproduceerd. Als laatste wordt de interne toestand van het model bijgewerkt, door de activatie en de locatie van de anchors aan te passen aan de hand van twee leermechanismen.

4.2. Aanpassingen

Om dit model te laten werken met tijdsintervallen zijn drie belangrijke aanpassingen gedaan. Allereerst is de (visuele) perceptiemodule vervangen door een cognitief plausibele temporele module. Deze bestaande temporele module is exact overgenomen van de in ACT-R geïntegreerde module van Taatgen, Van Rijn & Anderson (2004; Taatgen & Van Rijn & Anderson, in press) en werkt aan de hand van



Figuur 4.1: De wisselwerking tussen de belangrijkste ANCHOR-variabelen (Petrov & Anderson, 2005).

een pacemaker-gedreven interne klok die tijd codeert als ticks.

De tweede aanpassing is cognitief plausibelere versie van anchor-opslag in het declaratief geheugen. In plaats van een vast aantal anchors hebben we gekozen voor een dynamische benadering. Aan het begin van het experiment is het declaratief geheugen leeg. Voor elke nieuwe combinatie van categorie en ticks wordt een anchor aangemaakt. Anchors hebben dus een vaste waarde met een variabele activatie. Hierdoor wordt geheugenruis automatisch gegenereerd en kunnen we een parameter weglaten in vergelijking met de oorspronkelijke geheugenbenadering van ANCHOR.

De laatste aanpassing is het weglaten van het correctiemechanisme. Tijdsintervallen zijn in tegenstelling tot ruimtelijke objecten niet op te slaan als een mentale visuele weergave. Hierdoor is het ook niet mogelijk om terug te kijken naar de lengte van een tijdsinterval (vergelijking van M en A), omdat het alleen bestaat op het moment dat het wordt aangeboden, waarna gelijk codering in de hersenen met ruis plaatsvindt. Als we een tijdsinterval willen reproduceren, zal dit altijd ruis opleveren. De consequentie hiervan is dat het correctiemechanisme niet van toepassing is voor een model dat met tijdsintervallen werkt.

4.3. Temporele module

De structuur van de temporele module is gebaseerd op het idee dat tijd wordt bijgehouden aan de hand van de toestand van bepaalde neuronen (Matell & Meck, 2000). Aan het begin van elk interval kan een interne timer worden gestart, door de toestand van de neuronen te resetten. Elk punt in de tijd wordt gepresenteerd door een daarmee geassocieerde subset van vurende neuronen. De neuronen onderling verschillen van periodiciteit, waardoor overlap tussen subsets voorkomt (Taatgen, Van Rijn & Anderson, 2004). Voor langere intervallen geldt dat de schatting steeds minder precies wordt.

Deze interne-klok-theorie wordt computationeel als volgt gemodelleerd. Als een startsignaal wordt gegeven aan het begin van een tijdsinterval begint de *pacemaker* te lopen. De pacemaker vuurt ticks af:

$$t_0 = \text{starttick} \quad (4.1)$$

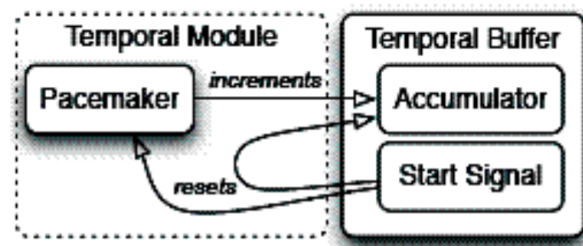
Deze starttick heeft een waarde van 0.02 aan het begin van elk interval. De ticks zijn van oplopende lengte (zie figuur 4.3); de lengte van de huidige tick t_{n+1} wordt bepaald aan de hand van de lengte van de vorige tick t_n :

$$t_{(n+1)} = a t_n + \text{noise}(\text{mean} = 0, \text{sd} = b \cdot a t_n) \quad (4.2)$$

Hierbij zijn a en b parameters met respectievelijk waarden van 1.2 en 0.2. Ruis wordt stochastisch toegevoegd door de *noise*-functie, waarvan de standaarddeviatie toeneemt met het aantal ticks. Gevolg hiervan is dat de ticks toenemen (zie figuur 4.3) en dit ook op een steeds grilliger wijze doen naarmate de tijd vordert. Ticks worden opgeteld en het resultaat wordt opgeslagen in een buffer (zie figuur 4.2) tot de pacemaker wordt gereset aan het einde van een tijdsinterval, waarna de inhoud van de temporele buffer als output wordt gegeven, dit is de magnitude M van het tijdsinterval.

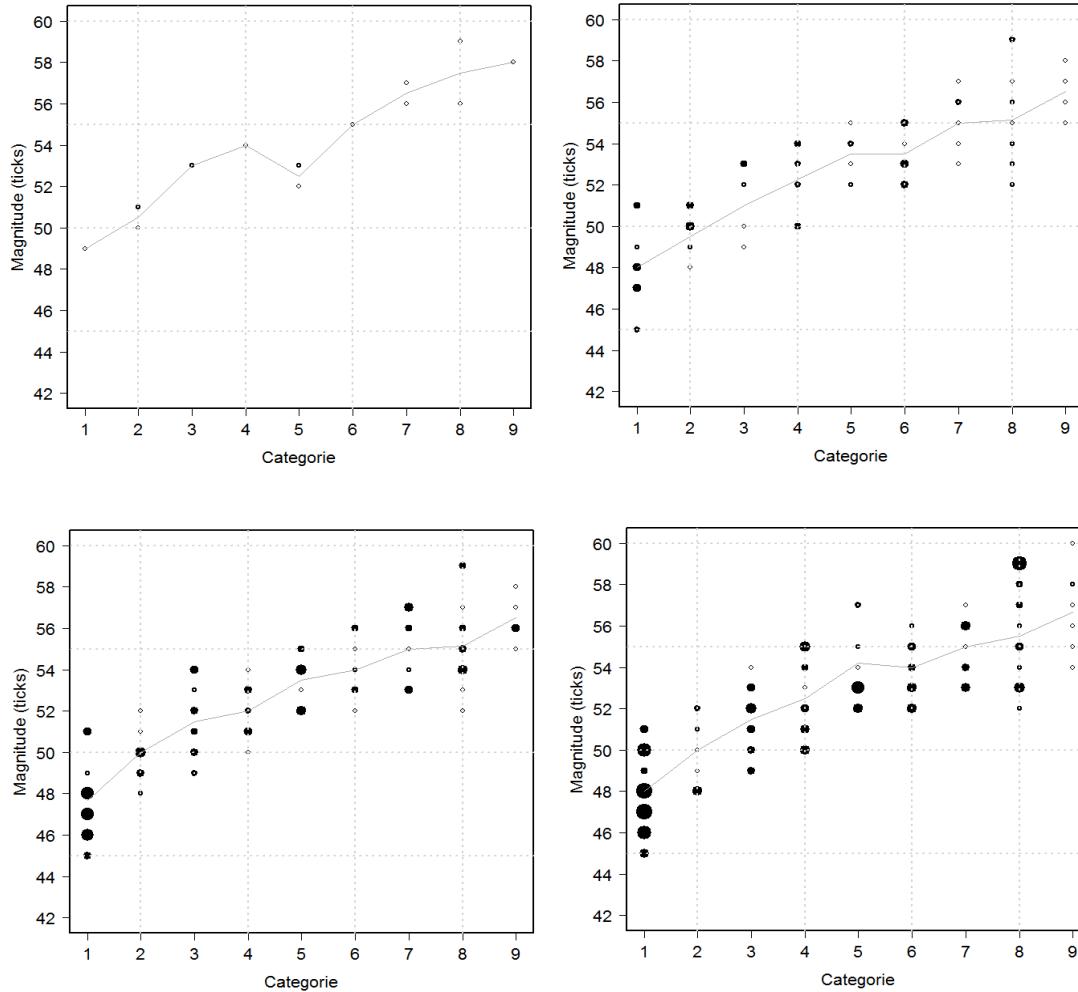
4.4. Activatie en distributie van anchors

De inhoud van het geheugen bestaat uit een set van unieke anchors. Deze anchors zijn gerepresenteerd als chunks in ACT-R. Anchors dienen als basis voor de vergelijking en inschatting van stimuli. Zoals eerder opgemerkt wordt na elke trial feedback in de vorm van de juiste categorie aangeboden. Deze informatie wordt tezamen met de magnitude van het bijbehorende aangeboden tijdsinterval opgeslagen als anchor. Een anchor is dus een associatie tussen de juiste categorie en



Figuur 4.2: Schematische weergave van de temporele module (Taatgen & Van Rijn & Anderson, 2007).

magnitude (in ticks) van een interval. Er kunnen dus meerdere anchors worden geassocieerd met een bepaalde categorie, voor categorie 3 kunnen bijvoorbeeld anchors met magnitudes 50, 51 en 52 in het geheugen voorkomen. Door deze ambigue representatie van anchors in het geheugen wordt automatisch ruis gecreëerd (zie figuur 4.4). Hierbij volgen we de *instance learning*-theorie van Logan (1988) die zegt dat het leren van vaardigheden kan worden verklaard door de optelling van het aantal voorbeelden van die vaardigheid. In het originele ANCHOR-model wordt geheugenruis gegenereerd door een Gaussiaanse ruisvariabele gemodificeerd door een parameter. In ons model komt dus een parameter minder voor.



Figuur 4.4: Verdeling van anchors in het geheugen tijdens een volledige run van één experiment van groep 1 (ULU). Elk punt representeert een anchor. De grootte van de stippen geeft aan hoe vaak de anchors uit het geheugen zijn opgevraagd; groter is hoger. De grijze lijn geeft een indicatie van de gemiddelde anchormagnitude per categorie, zoals in ANCHOR. Linksboven staat de anchorverdeling na het trainingsblok weergegeven, rechtsboven na het eerste blok (U), linksonder na het tweede blok (L) en rechtsonder na het derde blok (U). Omdat hier meer lage dan hoge intervallen worden aangeboden, zijn de stippen bij de lage intervallen groter.

Elke anchor heeft zoals alle chunks in ACT-R een activatiewaarde die de kans bepaalt op selectie. Elke keer als stimulus-categorie paar in het geheugen wordt aangemaakt als anchor of opgevraagd, wordt de activatie hiervan opgehoogd aan de hand van de bekende vergelijking voor de base-level activatie van een chunk (Anderson & Lebiere, 1998), zie vergelijking 3.

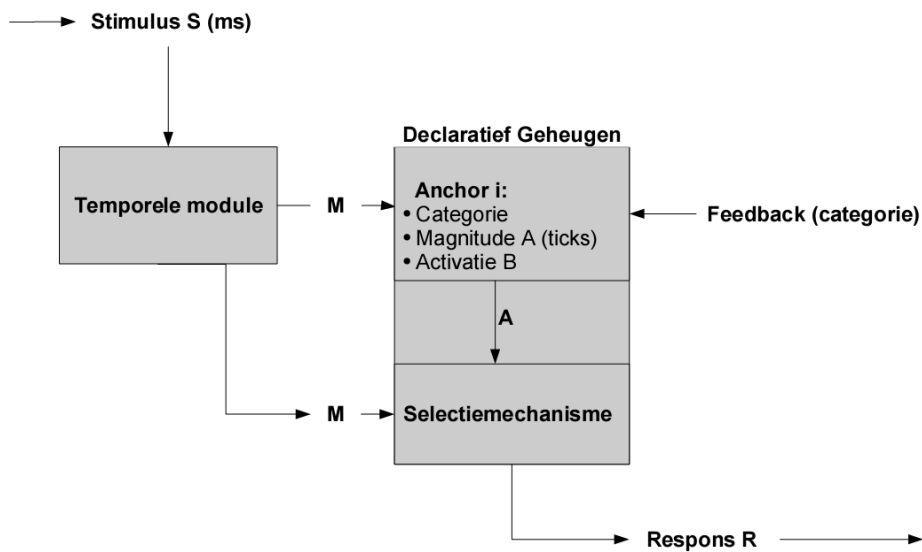
$$B = \log\left(\sum_{t=1}^n t^{-d}\right) \quad (3)$$

De vergelijking is een sommatie over een machtsverheffing van alle tijdstippen t dat een

chunk is opgevraagd uit het geheugen. De activatie vervalt met het verstrijken van tijd. De parameter $d = 0.5$ is de standaardwaarde in ACT-R die het verval van de activatie modificeert.

4.5. Selectiemechanisme

Het selectiemechanisme bepaalt de uiteindelijke categorie-respons op basis van vergelijking van de anchor-magnitudes A met de stimulus-magnitude M en de activatie-eigenschappen van het geheugen. Uit de verzameling anchors in het declaratief wordt een winnaar gekozen, waarna de categorie behorende bij deze anchor direct als respons wordt gegeven.



Figuur 4.5: Procedure van het temporele ANCHOR-model voor één trial.

Het computationele mechanisme is letterlijk overgenomen uit Petrov & Anderson (2005, p. 393). Vergelijking 4 bepaalt de *goodness* G voor elke anchor i in het declaratief geheugen. De anchor met de hoogste *goodness* heeft de meest kans om als winnaar gekozen te worden (zie vergelijking 5). De winnende anchor wordt op stochastische wijze gekozen door een random-number generator.

$$G_i = HB_i - |(M - A_i)| \quad (4)$$

$$P_i = \frac{\exp(G_i/T)}{\sum_j \exp(G_j/T)} \quad (5)$$

De *goodness* bestaat uit twee termen: het verschil tussen de stimulus-magnitude (context van de huidige anchor) en de huidige activatie B (geschiedenis van de huidige anchor). De schalingsfactor H bepaalt de ratio tussen de contextfactor en de geschiedenisfactor (Anderson & Milson, 1989). Deze is op dezelfde waarde gezet als in het oorspronkelijke ANCHOR-model; $H = 0.08$.

De *softmax*-vergelijking 4 zet de *goodness*-score van elke anchor i in het declaratief geheugen om in selectiekans P . Door de normalisatie in

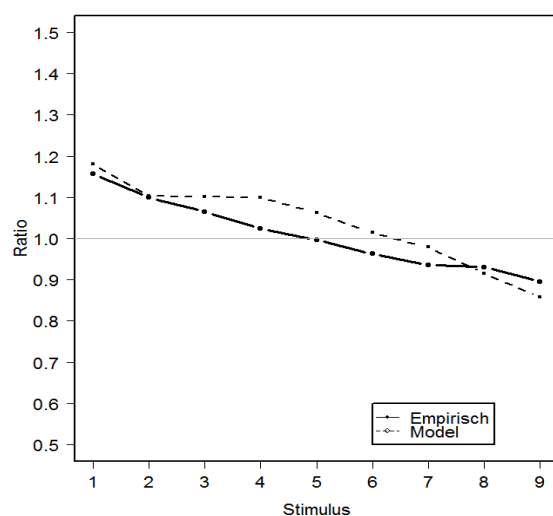
vergelijking 4 hangt de kans op selectie alleen af van de verschillen in *goodness*-scores tussen de anchors. De *temperatuur*-parameter T bepaalt de standaarddeviatie van de ruis in de selectie van de winnende anchor, die impliciet wordt toegevoegd in vergelijking 5. De temperatuur bepaalt dus de willekeurigheid waarmee een anchor wordt gekozen. Hoe hoger de waarde van T , hoe meer ruis wordt toegevoegd, waarden dichter bij nul zorgen voor een deterministischere keuze. De waarde van T is hetzelfde als in het originele ANCHOR-model, namelijk $T = 0.05$. De anchor met de hoogste *goodness*-score zal nu dus lang niet altijd winnen, maar heeft hier wel de grootste kans op. Door de activatie-eigenschappen van het geheugen hebben de vaakstgebruikte anchors (zie figuur 4.4) de grootste kans op selectie, gegeven de context factor. De magnitudes van deze anchors zitten waarschijnlijk het dichtst bij de daadwerkelijke lengte van het tijdsinterval.

In figuur 4.5 staat geschetst hoe het temporele ANCHOR-model te werkt gaat voor één trial. Aan het begin van de trial wordt een stimulus S aangeboden in de vorm van tijdsinterval in milliseconden. Dit tijdsinterval wordt door de temporele module omgezet in een interne

magnitude M. Het selectiemechanisme vergelijkt M met alle magnitudes A van de anchors in het declaratief geheugen en selecteert op basis van context- en geschiedenisfactoren de winnende anchor. De categorie van de winnende anchor wordt als respons gegeven. Het geheugen wordt geupdate door de activatie van alle anchors aan te passen met het de ACT-R leerregel (zie vergelijking 3) en als het declaratief geheugen deze nog niet bevat, een nieuwe anchor aan te maken met magnitude M en de juiste categorie (gegeven als feedback).

4.6. Opbouw interne representatie van stimuluschaal

In figuur 4.4 is goed te zien hoe de opbouw van een interne representatie van de stimuluschaal verloopt. Na het trainingsblok is een globale representatie van alle stimuli opgebouwd, maar de activaties zijn nog niet erg hoog, waardoor de geschiedenis nog geen grote invloed heeft op de selectie van de de winnende anchor (zie vergelijking 5). Na een volledige trial is meer vaststaande, maar ook meer ambigu, schaal opgebouwd. In de representatie na blok 2 (L) is de versterking van activaties van korte tijdsintervallen (vooral interval 1) te zien (voor groep 2 is dit effect tegenovergesteld). Na blok 3 is de verdeling van de activatie van anchors weer wat evenrediger geworden.



Figuur 5.1: Vergelijking ratio's per stimulus. Zie ook figuur 3.1 en 3.2.

5. Evaluatie model

Nu we de mechanismen van het model hebben uiteengezet, kunnen we het model gaan testen aan de hand van de empirische data. Om het experiment te reproduceren, hebben we het model 24 keer door het volledige experiment laten lopen. Uiteindelijk is een dataset verkregen analoog aan de dataset verkregen uit het menselijke experiment, met uitzondering van reactietijden, omdat hiervoor geen implementatie is gemaakt in het model. Voor het model zijn we uitgegaan van een vaststaande reactietijd van 1500 milliseconden, die gelijk is

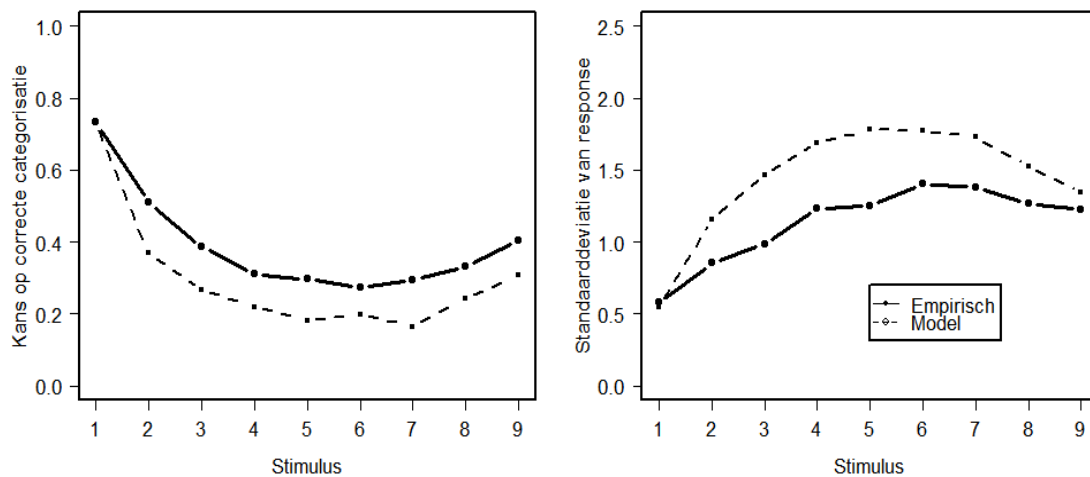
aan de gemiddelde reactietijd van alle proefpersonen over alle trials.

5.1. Gemiddelde respons per stimulus

De gemodelleerde resultaten (zie figuur 5.1) voldoen aan dezelfde kenmerken als de empirische resultaten. Korte stimuli worden overschat, terwijl lange stimuli worden onderschat. Opvallend is dat de modelgrafiek de $Ratio = 1$ -lijn pas na stimulus 6 kruist. Verder is het gesimuleerde overschattingseffect geprononceerder dan het onderschattingseffect. Er zijn geen opvallende verschillen tussen beide groepen voor zowel de empirische resultaten als de gemodelleerde resultaten, daarom zijn deze grafieken weggelaten.

5.2. Edge effecten

De prestatie van het model is minder goed dan de prestatie van proefpersonen bij de kans op correcte identificatie van stimuli (zie figuur 5.2). In de rechter grafiek is een veel grotere toename van de standaarddeviatie van responses te zien in het middelste gedeelte van de schaal. Er is dus een duidelijke structurele afwijking te zien. De grafieken geven ook voor het model een indicatie voor verbeterde prestaties aan de randen van de schaal.



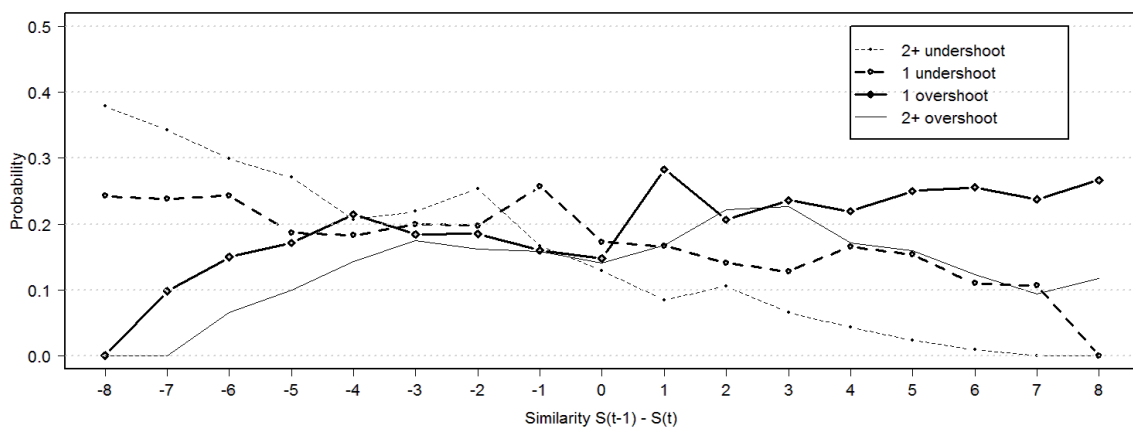
Figuur 5.2: Edge effecten. Links: Kans op correcte identificatie per stimulus. Rechts: Standaarddeviatie van responses per stimulus. De model-data vertoont structurele afwijkingen in vergelijking met de empirische data.

5.3. Sequentiële effecten

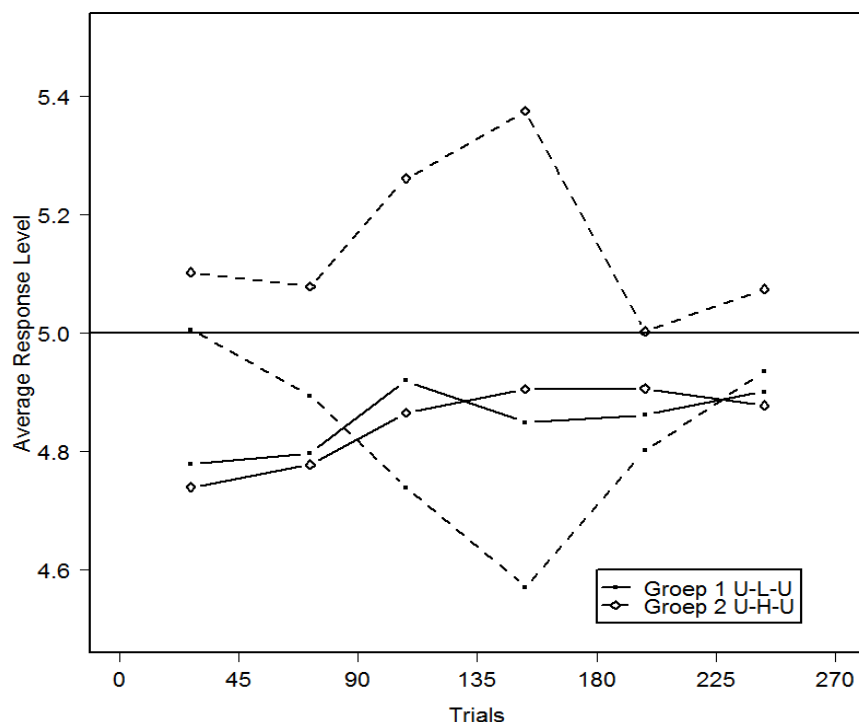
Figuur 5.3 plot de kansen op verschillende soorten fouten bij de uitgezet tegen ΔS over de gesimuleerde data op dezelfde wijze als bij de empirische data gebeurde. De gesimuleerde grafiek vertoont globaal gezien de zelfde kenmerken als de empirische grafiek. Wel is een veel sterkere toename van de undershoot 1 en overshoot 1 grafieken te zien op $\Delta S = \pm 1$. Verder geven de gesimuleerde data een indicatie op meer 2+ fouten dan bij de empirische data.

5.4. Context-effect: Average response level

De average response level (ARL) is op dezelfde wijze uit de gesimuleerde data berekend als bij de empirische data (zie figuur 5.4). Op het eerste gezicht vertoont de gesimuleerde grafiek weinig gelijkenis met de empirische grafiek. De gesimuleerde grafiek vertoont een scherpere toename in het tweede blok (trials 90-180) in vergelijking met de empirische grafiek. De kenmerken van beide grafieken zijn wel gelijk. Voor beide grafieken geldt namelijk dat voor groep 1 de ARL zakt in het scheve (lage) blok, en



Figuur 5.3: Sequentiële effecten: Kansen op verschillende soorten fouten uitgezet tegen de gelijkheid van de huidige stimulus met de vorige stimulus, vergelijk met figuur 4 uit de experimentele resultaten en met Petrov & Anderson (2005, p. 396).



Figuur 5.4: Average response level over 6 delen van het gehele experiment. De gestippelde lijn geeft de ARL van het model, zie ook figuur 5 bij de experimentele resultaten. Vergelijk ook met Petrov & Anderson (2005, p. 397).

voor groep 2 de ARL stijgt in het scheve (hoge) blok.

5.5. Overall prestatie

Tabel 5.1: Prestatie van model vergeleken met empirische data.

	Empirisch	Model
Proportie <i>correct</i>	0.394	.298
Proportie <i>fout</i> 1	0.606	0.702
Proportie <i>fout</i> 2+	0.187	0.308

In tabel 5.1 zijn de gesimuleerde en empirische proporties gegeven gemeten over de totale datasets. We zien dat de proportie correct bij de empirische data duidelijk verschilt van de gesimuleerde data. Verder is het opvallend dat het model 64,7 % meer 2+ fouten maakt dan proefpersonen.

6. Discussie

Nu zowel de gesimuleerde resultaten als de empirische resultaten in beeld zijn gebracht, is er grond voor vergelijking. Zoals gezegd dienen de empirische resultaten als referentiemateriaal voor de evaluatie van het temporele schalingsmodel. Aan de andere kant staan de empirische data ook op zichzelf als bevestiging van bekende (temporele) schalingseffecten, die op hun beurt kunnen worden verklaard door de mechanismen van het model. We bespreken vier effecten die volgen uit de resultaten afzonderlijk.

6.1. Vierordt's law

De analyse van de data per aangeboden stimulus (zie figuur 3.1, 3.2 en 5.1) geeft een duidelijke indicatie voor Vierordt's law. Dit geldt voor de empirische data en in mindere mate voor de gesimuleerde data. Vierordt's law zegt dat relatief korte tijdsintervallen worden overschat en relatief lange intervallen worden onderschat. Deze wet is meerdere malen bevestigd (Bobko et. al, 1977; Brown, 1995; Foley, Michaluk and Thomas, 2004). Een andere manier om Vierordt's

law te beschrijven is door te stellen dat de responses de neiging hebben om naar het midden van de schaal te gaan. Dit kan worden verklaard door te stellen dat proefpersonen bij onzekerheid over de grootte van de aangeboden stimulus eerder een respons in het midden van de schaal kiezen, zodat er toch nog punten worden verdiend.

Er is ook nog een andere verklaring te geven. De fouten aan de uiteinden van de schaal kunnen maar één kant op gaan, hierdoor komen aan het begin van de schaal alleen maar overshoots voor, terwijl aan het eind van de schaal alleen maar undershoots voorkomen. Dit resulteert in een lagere gemiddelde respons voor de uiteinden van de schaal en een verschuiving van de responses naar het midden.

In figuur 5.1 is te zien dat het model voor de stimuli 3-7 structureel een hogere ratio heeft. De overshoot-fouten zijn dus extremer dan bij de empirische data, terwijl de undershoot-fouten minder extreem zijn. Waarschijnlijk staat deze structurele fout in verband met de overall slechtere prestatie van het model.

Voor de overige buitenste stimuli van de schaal lijken de gesimuleerde en empirische data overigens meer op elkaar. Vierordt's law lijkt dus ook op te gaan voor het model.

6.2. Verklaring edge effecten

Uit figuur 3.3 en 5.2 zijn duidelijke indicaties op *accuracy edge effects* (Petrov & Anderson, 2005) af te leiden. Hiermee wordt een toegenomen proportie van correcte responses aan beide zijden van de schaal bedoeld. Dit soort toenames worden waarschijnlijk veroorzaakt door het feit dat er minder mogelijkheden zijn om fouten te maken aan de uiteinden van de schaal.

De standaarddeviatie laat een oplopende trend per oplopende stimulus zien voor zowel de empirische als gesimuleerde data. Hieruit valt af te leiden dat de stimuli in het midden van de schaal minder goed zijn te discrimineren. De linkergrafiek in figuur 5.2 bevestigt de slechtere prestaties in het midden van de schaal.

Door naar de representatie van de anchors (zie figuur 4.4) in het model te kijken, zien we ook dat vooral voor de middelste categorieën een groot aantal anchors met elkaar overlappen, wat

waarschijnlijk zorgt voor verminderde discriminabiliteit.

De verklaring voor de afnemende reactietijden (zie figuur 3.3) is waarschijnlijk dat er minder lang hoeft te worden gezocht in het geheugen. Dit effect is ook te verklaren aan de hand van *competitive latency*. Deze theorie zegt dat de kans kleiner is dat er veel anchors met huidig aantal ticks voorkomen in het geheugen, waardoor sneller een winnaar kan worden gekozen door het selectiemechanisme. Deze theorie is niet geïmplementeerd in ACT-R, maar zou een optie kunnen worden voor eventuele latere temporele schalingsmodellen.

6.3. Sequentiële effecten

Beide datasets laten een duidelijke indicatie zien voor een assimilatief sequentieel effect (zie figuur 3.4 en 5.3). Er zijn duidelijke indicaties dat de keuze voor de huidige respons in grote mate wordt beïnvloed door het verschil tussen de huidige en vorige stimulus.

Als we deze experimentele data vergelijken met de data van Petrov & Anderson (2005, p. 389), zien we vooral gelijkenissen, maar ook kleine afwijkingen. De grafieken voor undershoot vertonen een veel duidelijkere stijgende lijn dan de overshoot grafieken. Als het vorige interval veel lager was dan het huidige interval, dan resulteert dit in een grote kans op undershoot. Uiteindelijk is de kans op een undershoot van 2 of meer even groot als de kans op een undershoot van 1. Bij de overshoot grafieken zien we uiteindelijk weer een dalende lijn bij een $\Delta S > 7$. Hierbij weten we dat het huidige interval alleen 1 of 2 kan zijn. We weten dat deze korte intervallen over het algemeen zeer goed worden gecategoriseerd (zie figuur 3.3 en 5.2). De kans op correcte categorisatie van interval 1 is hoger dan 0.7, de kans op correcte categorisatie van interval 2 is hoger dan 0.5. Hierdoor zijn de zakkende overshoot-curven te verklaren.

Over het algemeen kunnen we zeggen dat de kans op undershoot toeneemt als de vorige stimulus korter is dan de huidige stimulus. De kans op een overshoot neemt toe als het vorige interval langer is dan het huidige interval. Vooral de empirische data laten dit goed zien.

De gesimuleerde curves hebben een grilliger verloop, maar vertonen over het algemeen wel dezelfde kenmerken als hierboven beschreven. Zoals gezegd geven de gesimuleerde data een indicatie op meer 2+ fouten dan bij de empirische data. Uit dit feit kunnen we de hogere standaarddeviatie van de responses bij het model (zie figuur 5.3) verklaren.

6.4. Assimilatieve contexteffecten

De average response levels (zie figuur 3.5 en 5.4) geven een indicatie voor assimilatieve contexteffecten. De ARLs veranderen dus afhankelijk van de context. Als er recentelijk meer korte intervallen worden aangeboden, dan wordt de gemiddelde response lager. Het effect is tegengesteld bij meer lange intervallen. Dit effect kan worden verklaard door de anchoring-hypothese besproken in de inleiding.

De empirische resultaten geven waarschijnlijk geen significante afwijking van de ARLs per blok. Dit is waarschijnlijk te wijten aan het feit dat te weinig verschillende blokken zijn aangeboden. Voor een volgend experiment is dan ook aan te raden om met 5 blokken te werken, zoals wordt gedaan in Petrov & Anderson (2005).

Wat zoals gezegd wel opvalt, is de gemiddelde ARL over het gehele menselijke experiment. We zouden verwachten dat de ARL-waarden rond de 5.0 schommelen, net als bij het experiment van Petrov. Onze waarden schommelen daarentegen rond de 4.85. Deze structurele onderschatting is opvallend, maar is te verklaren uit de andere effecten. Figuur 3.3 laat zien dat de prestatie (% correct) van proefpersonen bij de korte tijdsintervallen (1 en 2) opvallend hoger is dan de prestatie bij lange tijdsintervallen (8 en 9). Dit verschil is groter dan bij de schatting van lijnstukken. Er worden dus meer undershoot-errors dan overshoot-errors gemaakt. Aan de lage kant van de schaal is de kans op een overshoot veel groter dan een undershoot, aan de hoge kant van de schaal is de kans op een undershoot veel groter dan een overshoot. Omdat de kans op fouten aan de hoge kant groter is dan de kans op fouten aan de lage kant, zouden we kunnen afleiden dat de proportie undershoot-fouten groter is. Hierdoor is de lage

ARL te verklaren. Het effect van de plaats op de stimulischaal op de prestatie is dus groter bij schattingen in het temporele domein dan bij schattingen in het ruimtelijke domein.

Verder is geen rekening gehouden met een eventueel *lengthening* effect (Bobko et al, 1977) waarmee eventueel rekening moet worden gehouden tijdens de analyse van de resultaten. Het lengthening effect zegt dat de schatting van tijdsintervallen toeneemt, naarmate het experiment vordert.

6.5. Verklaring kwalitatieve afwijking van gemodelleerde resultaten

Het model maakt structureel te veel fouten (zie figuur 5.2 en tabel 5.1). Hiervoor zijn twee verklaringen te geven. Ten eerste kunnen de parameterinstellingen zorgen voor een te groot aantal fouten. Een tweede verklaring kan het ontbreken van een mechanisme in het temporele schalingsmodel zijn.

In dit onderzoek is het niet mogelijk geweest om alle combinaties van parameterinstellingen te testen. Hierdoor is het mogelijk dat er betere resultaten te bereiken zijn met het huidige temporele model. Ook kan er nog worden gekozen voor het toevoegen van een parameter in de goodness-functie, die kan zorgen voor een beter evenwicht tussen de context-factor en de geschiedenis-factor. Hierdoor kan mogelijk worden gezorgd voor een lagere proportie fouten.

De parameterinstellingen in ons model zijn gebaseerd op het ANCHOR-model. Als het blijkt dat de oorzaak van de fout in ons model bij de parameterinstellingen ligt, is het misschien nodig om ook de parameters in het ANCHOR-model te re-evalueren.

Als blijkt dat de parameterinstellingen niet te optimaliseren zijn, dan wijken de gesimuleerde data kwalitatief af van de empirische data. Hiervoor zijn verschillende verklaringen te geven. In het originele ANCHOR-model zorgt het correctiemechanisme voor een eventuele bijstelling van de respons (aan de hand van een geheugengebaseerde globale schatting) als deze

te veel afwijkt. Dit is zoals gezegd niet mogelijk voor tijdsintervallen. Het correctiemechanisme vergelijkt de magnitude van de stimulus nog eens met de magnitude van de winnende anchor. Voor mensen is dit te zien als het nog eens terugkijken naar de stimulus: 'Wijkt mijn schatting niet te veel af van wat ik zie?'. Dit is goed mogelijk voor lijnstukken, hiervoor kunnen mensen bijvoorbeeld de interne magnitude van de winnende anchor visualiseren en tegelijk naar de stimulus kijken. Voor tijdsintervallen is deze parallelle verwerking niet op deze manier mogelijk, omdat het aangeboden tijdsinterval al voorbij is. Om twee tijdsintervallen te vergelijken moeten mensen ze parallel reproducen, wat leidt tot weer een nieuwe cyclus van geheugengebaseerde schaling. Een correctiemechanisme voor tijdsintervallen levert dus geen voordelen op en is bovendien niet cognitief plausibel.

Vanuit dit perspectief is lokale vergelijking dus niet op dezelfde wijze mogelijk als bij ruimtelijke stimuli. Het model geeft zijn respons dus enkel aan de hand van een globale schatting. Hiermee bedoelen we dat het model een interval categoriseert aan de hand van de positie op de gehele schaal. Omdat het een correctiemechanisme of een analoge vorm hiervan ontbreekt, kunnen responses niet lokaal worden vergeleken. Daardoor kunnen grote fouten (2 of groter) voorkomen. We hypothetiseren dat mensen hebben blijkbaar wel de gave om te grote afwijkingen in hun respons te detecteren en alsnog deze respons bij te stellen. Hieruit zouden we kunnen afleiden dat het model nog een mechanisme mist die mensen wél gebruiken bij temporele schaling.

7. Conclusie

Bij de ontwikkeling van het temporele cognitieve schattingsmodel zijn zoals gezegd drie doelen gesteld. Ten eerste is een correcte implementatie in een al bestaande cognitieve architectuur nagestreefd door twee reeds geteste modellen te combineren. De verwachting was dat dit model een goede eerste opzet is voor een temporele schalingtheorie. Empirische data bevestigde deels de validiteit van het model. Omdat het model een mechanisme voor lokale vergelijking

mist, zijn de prestaties slechter dan bij menselijke proefpersonen.

Het model moet verder zorgen voor meer inzicht in de interne mechanismen en representaties bij temporele schattingstaken. Dit is ook deels gelukt, maar omdat het model nog verbeteringen nodig heeft, kunnen alleen indicaties voor mechanismen worden gegeven.

Het laatste doel was om meer inzicht te verkrijgen in de toepasbaarheid van de mechanismen van het originele ANCHOR-model op het temporele domein. Een correctiemechanisme voor de schatting van tijdsintervallen is niet mogelijk, daarom is dit ook niet geïmplementeerd. Zoals gezegd is gebleken dat de prestaties van de gesimuleerde data achterblijven bij de empirische data. Blijkbaar mist het temporele schalingsmodel een mechanisme dat gebruikt maakt van dezelfde voorkennis-strategieën als mensen. Petrov & Anderson (2005) stelden dat een correctiemechanisme de beste implementatie is van zulks een strategie. Dit gaat voor temporele schattingstaken blijkbaar niet op. Daarom worden twijfels getrokken bij de toepasbaarheid van het ANCHOR-model op het temporele domein.

Toekomstig onderzoek

Het ontwikkelde temporele schalingsmodel is nog verre van compleet, maar kan wel dienen als referentie voor toekomstige onderzoek en verbeteringen van de modellering van temporele cognitieve taken. Hieronder geven we enkele wegwijzers voor toekomstige onderzoek.

De analyse van de empirische data is evenals het model nog incompleet. Er zijn meerdere bestaande effecten die nog geanalyseerd kunnen worden. Er is geen onderzoek gedaan naar leereffecten. Worden proefpersonen beter naarmate het experiment vordert, is dit ook het geval voor het model? Verder is onderzoek naar lengthening effecten nodig, om eventuele biases in de empirische data te verklaren.

In dit onderzoek zijn alleen indicaties voor effecten gegeven, er zijn weinig tot geen statistische maatstaven gebruikt om bijvoorbeeld significante afwijkingen in de data te onderzoeken. Verder is alle analyse in dit

onderzoek gedaan door te middelen over alle proefpersonen. Voor meer inzicht in individuele verschillen is het waarschijnlijk nodig dat de data ook per proefpersoon wordt geanalyseerd.

Het doel van dit onderzoek was niet om de gebruikte temporele module te verbeteren. Hier zijn dan ook geen wijzigingen in aangebracht. Wel kan in latere experimenten en modelleerproeven het huidige temporele mechanisme worden vervangen door een willekeurige andere implementatie, omdat de temporele module volledig losstaat van de rest van het model.

Het blijft nog onduidelijk wat voor mechanisme moet worden gebruikt om tot lokale vergelijking te komen analoog aan het correctiemechanisme. Welke strategie gebruiken mensen bij het inschatten van tijdsintervallen die gebruikmaakt van a priori kennis over onder andere de ordening van de stimuli-schaal?

8. Referenties

- Anderson, J. R. & Milson, R. (1989). Human Memory: An Adaptive Perspective. *Psychological Review*, 96, 703-719.
- Anderson, J. R. & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bobko, D.J., Schiffman, H. R., Castino, R. J., Chiappetta, W. (1977). Contextual Effects in Duration Experience. *The American Journal of Psychology*, 90(4), 577-586.
- Brown, S.W. (1995). Time, change, and motion: the effects of stimulus movement on temporal perception. *Percept Psychophys.*, 57(1), 105-16.
- Foley, A.J., Michaluk L.M., Thomas D.G. (2004). Pace alteration and estimation of time intervals. *Percept Mot Skills*, 98(1), 291-8.
- Logan, G. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-527.
- Luce, R. D., Nosofsky, R. M., Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. *Perception and Psychophysics*, 32, 397-408.
- Matell, M.S. & Meck, W.H. (2000). Neuropsychological mechanisms of interval timing behavior. *BioEssays*, 22, 94-103.
- Petrov, A. (2003). Additive or multiplicative perceptual noise? Two equivalent forms of the ANCHOR model. In R. Alterman & D. Kirsh (Eds.). *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 922-927). Hillsdale, NJ: Erlbaum.
- Petrov, A. (2001). Fitting the ANCHOR model to individual data: A case study in Bayesian Methodology. *Proceedings of the Fourth Intern. Conf. on Cognitive Modeling*. Hillsdale, NJ: LEA.
- Petrov, A., & Anderson, J. R. (2000). ANCHOR: A memory-based model of category rating. *Proceedings of the 22 Annual Conf. of the Cognitive Science Society* (pp. 369-374). Hillsdale, NJ: LEA.
- Postman, L., Miller, G.A. (1945). Anchoring of Temporal Judgments. *The American Journal of Psychology*, 58(1), pp. 43-53.
- Taatgen, N. A., Rijn, H. v., & Anderson, J. R. (2007). An Integrated Theory of Prospective Time Interval Estimation: The Role of Cognition, Attention and Learning. *Psychological Review*, 114(3), 577-598.
- Taatgen, N., Rijn, H. v. & Anderson, J. R. (2004). Time perception: Beyond simple interval estimation. In *Proceedings of the sixth International Conference on Cognitive Modeling* (pp. 296-301). Pittsburgh, PA: Carnegie Mellon University/University of Pittsburgh.
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs*, 77 (13, Whole No. 576).
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1130.
- Vierordt, K. v. (1868). *Der Zeitsinn nach Versuchen*. Tübingen: Laupp

