Enhancing vision: a project on audio-visual decision-making, and a project on strengthening the visual system by short-term monocular patching

Siu Fung Andrew Chua

A thesis submitted for the degree of PhD
at the
University of St Andrews

2023

Identifier to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/sta/392

Candidate's declaration

I, Siu Fung Andrew Chua, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 79,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2018.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date **4th April 2023** Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date **4th April 2023** Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Siu Fung Andrew Chua, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date **4th April 2023** Signature of candidate

Date **4th April 2023** Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Siu Fung Andrew Chua, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date **4th April 2023** Signature of candidate

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

No embargo on underpinning research data or digital outputs.

Date **4th April 2023** Signature of candidate

Date **4th April 2023** Signature of supervisor

General Acknowledgements

My work was only possible with the support of many people, within and outside of the University of St Andrews. I am eternally grateful to the people that I have met during this time, and have guided and supported me, at work and also in my development.

In particular, I would like to thank my supervisors Dr. Thomas Otto and Professor Julie Harris. Dr. Otto taught me computer programming, was crucial to my project on sensory decision-making, and provided all-round support during my time as a student at the University of St Andrews. Professor Harris first spotted me as an undergraduate student, and has since given me many opportunities, including this postgraduate position, and has supported me on many vision experiments. I would also like to thank Dr. Justin Ales, who went out of his way to provide immense support in setting up the vision equipment for the monocular patching project.

Funding

This work was supported by the St Leonards Research Scholarship. The earliest experiments were supported by the Experimental Psychology Society's Undergraduate Research Bursary.

Research Data / Digital Outputs access statements

Research data underpinning this thesis are available at https://doi.org/10.17630/99f30a82-803f-43ed-8bea-e05485549d89

Ethical Statement

All research involving human participants comply with the *Code of Human Research Ethics* (The British Psychological Society, 2014), and was reviewed and approved by the University of St Andrews' *University Teaching and Research Ethics Committee* (UTREC; approval code PS12994 for the audio-visual looming project, Appendix A for ethical approval letter; approval code PS14148 for the short-term monocular patching project, Appendix D for ethical approval letter).

Abstract

Vision is important; two projects were conducted to explore the enhancement of vision (multisensory combination, novel procedure to improve visual functioning). In the first project, audition was combined with vision, to examine if the conjunction of two sensory modalities benefits sensory decision-making in the context of looming – a potentially dangerous motion requiring quick responses (e.g., Neuhoff, 2001). Four behavioural experiments were conducted, testing responses towards auditory, visual, or audio-visual motion-in-depth signals (i.e., looming or receding). From behavioural data, the decision-making mechanism was explored in two separate analyses: 1) a comparative approach (e.g., Innes & Otto, 2019) which compared empirical data to predictions made using probability summation (Raab, 1962), and 2) a computational modelling approach which determined the best permutation of an interactive probability summation rule (Otto & Mamassian, 2012) to fit the empirical data. Altogether, the first project revealed that decision-making had a speed benefit when auditory and visual signals were both present, compared to either modality alone, and this benefit could be explained by a probability summation rule with simple audio-visual interactions. Not found was evidence of especially quick processing uniquely towards audio-visual looming signals. For the second research direction, a pilot study was conducted to test a novel short-term monocular patching technique which purportedly induces eye dominance via latent neuroplasticity (e.g., Lunghi, Burr, & Morrone, 2011), with the idea that it could become a procedure for improving amblyopic visual functioning. In one small-scale experiment (n=4), a significant dominance effect on the patched eye was found, but the effect was short-lived. Two single-participant experiments indicated that patching durations should not be too short (10 minutes) and stereopsis sensitivity may slightly change after patching. This pilot study offered a preliminary look into patching effects, on the back of which further experiments are suggested.

Table of Contents

Chapter 1: General introduction

1.1 Preface

It is morning and it is time to wake up, seeing sunlight filtering through the curtains into the dark room. One looks around in the kitchen to see what to have for breakfast. Then navigating the roads out of the house, one sees challenging, potentially dangerous stretches of road, like stairs, and look to make sure every step is safe. There are people around, and one recognises people one knows, and greets them. There may be beautiful landscapes on the way; one sees the scenery and derive pleasure from the views. Finally, one sees somewhere comfortable to sit, and make their way there to relax. Then (perhaps) one might take out this thesis to read...

The above could be a mundane routine on any given day, yet all of the above actions benefit greatly from, if not require, vision in one form or another. Vision is a very important sense. It was the realisation that vision is so immediately useful, and used so often in every moment awake, that inspired this PhD on investigating how vision can be enhanced, in some way beneficial in the real world. But how exactly can vision be enhanced? This PhD approached the question using two projects: first taking an external view of maximising the sensory environment, the second taking an internal view of improving the visual system itself.

In more detail, the first project asks the question of how the other navigational and spatial sense, audition, can pair with vision to produce faster and more accurate decisions. Because looking around, one may realise the problem of relying on vision alone. Additionally, one will also sense that the world is not only visual. Consider the following real life example. Imagine standing on the pavement, about to cross the road, but the road has a bend to it, and it is lined with tall bush. As a result, there is a limited view of the road, left or right. There is a speeding car approaching, and it is not far away, but it is only somewhat visible through the bush. Deciding to cross the road, based solely on this incomplete visual information, could have been disastrous. However, if one also used their sense of hearing, then one may hear the tyre roar and engine noise getting louder, and deduce that there is a car approaching. One waits on the pavement, and indeed sees the car passing by moments later. Two points can be made from this everyday scenario. First, visual signals are not always complete, and sometimes not available at all (e.g., in the dark). However, as the second point, the environment is multisensory, and one uses all available senses to successfully interact with the environment. Thus, putting all these ideas together, the first project of my PhD explores how vision works together with audition to determine decision-making towards approaching objects. Approaching motion is also known as looming, hence this project is titled as the audio-visual looming project.

The second project asks if the visual system itself can be improved. Not everyone has the perfect visual system, which given how useful vision is, it is important to find techniques that improve visual functioning. Thus, the second project tests a novel eye patching technique which shows promise by purportedly tapping into latent visual system neuroplasticity in adults (Lunghi, Berchicci, Morrone, & Di Russo, 2015a; Lunghi et al., 2011; Lunghi, Emir, Morrone, & Bridge, 2015b). Visual system neuroplasticity, conventionally thought of to be non-existent by the time one reaches adulthood, could be useful in inducing positive visual system changes. This second project is titled as the short-term monocular patching project.

As this thesis is a large document, comprising two distinct projects with their own perspectives, methodologies and results, the general layout of the thesis is presented here, before proceeding further. In the introduction (Chapter 1), the literature review and rationale for both projects are presented side-by-side, for easy comparison, and to highlight that although the projects are distinct in focus, they both answer to the same interest in enhancing vision. Then, the thesis splits into Section 1, for the audio-visual looming project, within which is a general method chapter (Chapter 2), three empirical chapters (Chapters 3 to 5), and an overall discussion of the project (Chapter 6). Next, Section 2 is for the short-term monocular patching project, consisting of a general method chapter (Chapter 7), three empirical chapters (Chapters 8 to 10), and an overall discussion of the project (Chapter 11). Finally, Chapter 12 concludes, pulling everything learnt from the two projects, to answer how vision can be enhanced.

1.2 Introduction to the audio-visual looming project

Looming, as a motion towards oneself, is potentially dangerous, and we can detect looming by sound and sight. In Chapter 1.2.1 the processing of auditory and visual looming signals are explored in more detail, revealing that humans have biased responses towards looming signals but not receding signals. However, the real world is multisensory, so in Chapter 1.2.2, the case of human multisensory perception is explored, looking into the benefits that multisensory perception brings, and the literature about multisensory processing of audio-visual looming signals. In reviewing the literature about multisensory processing of audio-visual looming signals, it is revealed that there are still avenues in which this topic can be explored further. Thus, in Chapter 1.2.3, the knowledge gaps are formulated, and these knowledge gaps serves to guide my project on audio-visual looming.

1.2.1 Looming

Looming is the motion towards oneself, and in the real world, it can be associated with an impending collision or attack (e.g., approaching car, dangerous animal charging towards you). Thus, looming motion, as a sign of a dangerous situation, must be responded to with timely evasive action, for survival. Along such lines, there have been postulations that humans are especially quick to attend to, and ultimately respond to such dangerous motions that is looming (e.g., behavioural urgency hypothesis; Franconeri & Simons, 2003). More specifically, if one can quickly attend and respond to looming motion, then one leaves more room between oneself and the looming object, increasing the margin of safety and hence the likelihood of coming out unharmed – survival (Neuhoff, 2001). Cues to looming are available in the two sensory modalities that allow for navigation in the environment: audition and vision. A wide range of auditory and visual studies have found that humans do indeed show enhanced attention and responses towards looming motion in particular, and the phenomenon is termed looming bias in this thesis. Studies showing the looming bias are reviewed below, starting with studies on auditory looming, then with studies on visual looming.

Studies on Auditory Looming

Looming can be presented in the auditory modality, and in studies, it is often the rising intensity cue that represents auditory looming (e.g., Bach, Neuhoff, Perrig, & Seifritz, 2009; Bach et al., 2008; Neuhoff, 1998, 2001; Zahorik, Brungart, & Bronkhorst, 2005). In basic terms, if a sound source is far away from the listener, then the sound energy dissipates into the environment as it travels to the listener, and the listener only hears a quiet sound. But if the sound source is near to the listener, overall there would be less energy dissipation due to the smaller distance travelled, and the listener hears the sound louder. In fact, for a sound source over one metre away and in an auditory free-field, the distance to the sound source and the sound intensity heard follows a regular relationship: a six decibel decrease for every doubling of the distance to the sound source (Zahorik et al., 2005), i.e., if a speaker 50 metres away can be heard playing music at 60 decibels, then positioning the speaker 100 metres away, the same music will be heard at 54 decibels. Hence, tapping into this regular relationship between source distance and heard intensity, an increase in heard intensity can be associated with looming motion, while a decrease in heard intensity can be associated with receding motion.

In reviewing the literature, humans seem to respond differently to auditory looming signals than auditory receding signals, and such a looming bias broadly falls into three categories: perceptual, neurological and physiological. Perceptually, on listening to a sound increasing (or decreasing) in intensity, people do seem to perceive a movement in the sound source, much more so than for a sound of constant intensity (Seifritz et al., 2002), so the intensity manipulation appears to be a valid approximation to a looming or receding sound source – it is not perceived as simply an intensity change from a stationary source. However, the perceived sound source movement is not the same with an intensity increase versus an intensity decrease. Given symmetrical changes in intensity to simulate looming (70dB increasing to 85dB) or receding (85dB decreasing to 70dB), the perceived size of the movement is greater for the intensity increase than with the intensity decrease (Seifritz et al., 2002).

The judgement of greater illusory movement associated with intensity increases links with the finding that people also tend to perceptually exaggerate the loudness change in intensity increases (Neuhoff, 1998, 2001). When presented with a symmetrical pair of sounds, an intensity increase (40dB to 70dB) and an intensity decrease (70dB to 40dB), people tended to perceive bigger loudness changes with the intensity increase than with the decrease, despite both being a 30dB change (Neuhoff, 2001). The perceptual exaggeration of intensity increase was even more pronounced when the sounds were louder overall (60db to 90dB to loom, 90dB to 60dB to recede, maintaining the 30dB change; (Neuhoff, 2001)). The perceptual exaggeration of intensity increases may be adaptive, because if the sound seems loud, then presumably the sound source is nearby and dangerous (especially with the 60-90dB set), so there is an impetus for quick evasion, quicker than that strictly necessary, such that there is a margin of safety to the looming object, thus increasing survival odds (Neuhoff, 2001). The link between perceptual exaggeration of looming signals and an illusory percept of closeness was backed up in a further experiment which used a physically looming or receding speaker playing tones (Neuhoff, 2001). Blindfolded participants judged the speaker to stop closer to themselves when it physically loomed compared to when it physically receded, even though the speaker always travelled at the same speed, and in fact stopped at the same point for both looming and receding motions (Neuhoff, 2001).

Perceptual biases towards auditory looming signals were also demonstrated using the time-to-arrival paradigm, which presents segments of a looming signal, and examines if the participant correctly guesses when the moving object suggested by the signal will arrive at the participant's position; underestimates of the time-to-arrival suggest a propensity to err on the side of caution, thus likely to respond quicker than necessary to safely avoid the impending collision (Rosenblum, Wuestefeld, & Saldaña, 1993; for an audio-visual study, see Schiff & Oldak, 1990). In Rosenblum and colleagues' study, which used audio recordings of a car approaching from the side, 84% of the time-to-arrival estimates were underestimates (Rosenblum et al., 1993). Similarly, Schiff and Oldak (1990) made audio and video recordings of a car approaching head-on, or of a talking person approaching, and found underestimates of the time-to-arrival across both modalities. Underestimates of the time-to-arrival are proposed to be adaptive, because it means that people can make evasive actions earlier than needed, thus creating a margin of safety which increases survival odds (Rosenblum et al., 1993).

Neurologically, effects associated with auditory looming signals have also been observed. When presented with tones increasing in intensity, but not for tones decreasing in intensity, greater blood oxygen level dependent (BOLD) responses were observed in brain regions that are associated with auditory attention and spatial perception (Seifritz et al., 2002). In a similar vein, for rhesus monkeys, neural activity in the auditory cortex was larger in magnitude in response to auditory signals with intensity increases, but not intensity decreases, and that the common need to quickly evade looming objects could mean there are similar biases in humans as well (Maier & Ghazanfar, 2007). Touching on both neurological and physiological effects, auditory signals with intensity increases are associated with increased activity in the amygdala, which functions as a detector of relevant events, and sets off a series of physiological responses: orienting reflex, enhanced skin conductance, early heart rate deceleration, and speeding up of responses to further auditory stimulation (Bach et al., 2008). Altogether, the raft of neurological and physiological effects appear to be in preparation for quickly detecting and responding to further auditory signals of danger, and is likely adaptive (Bach et al., 2008). Finally, on a behavioural note, it has been found that infants under the age of one would lean back in avoidance when presented with a sound increasing in intensity, but not for a sound decreasing in intensity (Freiberg, Tually, & Crassini, 2001).

In summary, there is an abundance of studies to show that humans have particular responses towards auditory signals with intensity increases, and these responses follow a common theme of effecting quick and timely avoidance responses. Taking that auditory signals with intensity increases approximate a looming sound source, and looming is a dangerous motion, then it seems quite evident that there is an adaptive bias towards looming signals which works to quickly put oneself out of the dangerous looming path.

Studies on Visual Looming

Looming motion can also be presented in the visual modality. There are several visual cues to an object's movement in the depth dimension, but commonly, visual size change is used – looming as an increase in visual size, and receding as a decrease in visual size, as projected on the observer's retina. Conveniently as well, for experimentation, visual size increases (or decreases for that matter) are straightforward to implement in laboratory setups, e.g., computer displays or translucent screens, as they are in practice a size change on the object.

In studies on visual looming signals, biased responses have been found, similar to those found in studies on auditory looming signals. Broadly categorised, the visual biases fall into behavioural, attentional and motoric realms. Starting with behavioural biases towards visual looming signals, infants several weeks old produced protective avoidance responses, such as leaning backwards and bringing arms up in front of the face, towards physically looming objects as well as towards two-dimensional representations of visual looming (i.e., visual object increases in size to simulate looming), but never towards receding motions or its two-dimensional representation, or in the absence of the stimulus (Ball & Tronick, 1971). Interestingly, infants also seem capable of distinguishing between collision looming and non-collision looming: infants only turned slowly to look at the motion of the object if it was approaching askew, without a clear avoidance response (Ball & Tronick, 1971). It appears that at an early age, humans can visually recognise dangerous looming motion, and respond specifically to the motion using appropriate protective actions (Ball & Tronick, 1971).

If one has question marks around the first paragraph of this section, where experiments equate two-dimensional size increases on a flat plane to three-dimensional motions in space, then Ball and Tronick's (1971) study is also perfect to assure the validity of this practice in stimulus design. In Ball and Tronick (1971), the use of a physically looming stimulus produced the protective avoidance responses in infants, but interestingly, Ball and Tronick (1971) also tested the two-dimensional analogue of physical looming: a projecting light shining onto an object, thus projecting a shadow onto a translucent screen which the infant views, and by moving the object closer or further away from the projecting light, the shadow cast onto the screen changes size. Crucially, with this two-dimensional analogue to physical looming motion, infants also produced the same protective avoidance behaviours specific to looming, real or two-dimensional (Ball & Tronick, 1971), thus validating the common practice of using visual size changes to represent looming motion. Even more interesting was that infants could use the two-dimensional imagery to distinguish between head-on collision looming (symmetrical size increase of the shadow on the screen) and non-collision looming (asymmetrical size increase of the shadow on the screen) – they showed avoidance only to collision looming (Ball & Tronick, 1971). For further validation of two-dimensional visual size changes as representations of looming and receding, in a study using a similar setup of projecting a changeable shadow onto a translucent screen, infant and adult rhesus monkeys showed avoidance behaviours (cries of alarm, physical avoidance, leaping to the back of the cage) on seeing the shadow on the screen increase in size to 'loom', but not when it decreased in size to 'recede' (Schiff, Caviness, & Gibson, 1962). The fear and avoidance response was the same for infant and adult rhesus monkeys, and one interpretation is that these responses to looming are learnt very quickly early in life, as these responses are protective and crucial for survival (Schiff et al., 1962). Looming's relevance to survival for all beings might explain the parallels in behavioural responses towards looming, between rhesus monkeys (Schiff et al., 1962) and humans (Ball & Tronick, 1971).

Studies on visual looming signals have also claimed heightened attention towards looming signals but not to receding signals. In a landmark study using the visual search paradigm, participants were quick to find the target when cued with a masking object increasing in size (i.e., looming), thus suggesting attention capture by the looming masking object, but this efficient search was not found when the masking object decreased in size to simulate receding motion (Franconeri & Simons, 2003). In more detail, Franconeri and Simons (2003) used an experiment paradigm where on each trial, there would be a circular search array of letters surrounding the central fixation point, but before the letters were shown, each letter was masked by an object. Shortly before the letter masks were removed to reveal the search array, one of the letter masks would increase in size to simulate looming, or decrease in size to simulate receding; after the motion, the search array was then revealed and the participant had to look for a target letter in the array (Franconeri & Simons, 2003). If the motion on the masking object captures attention, then there would be two effects on search performance: first, if the moving masking object was where the target letter would be, then the target letter would be quickly found because captured attention is already in the target area, and second, if the moving masking object was on a distractor letter, then the target letter would be found slowly because captured attention has been brought away from where the target letter would appear (Franconeri & Simons, 2003). Empirically, search performance exactly followed the above predictions – quick search when the looming masking objects were on the target letter, slow search when the looming masking objects were on the distractors (Franconeri & Simons, 2003). Receding masking objects apparently did not capture attention, as search speeds were slow regardless of where the receding masking objects were in the array (Franconeri & Simons, 2003). The finding that looming (but not receding) motion captures attention must be placed in the wider context: in the same study, motions that similarly captured attention were the appearance of new objects, or sudden movement of objects (Franconeri & Simons, 2003). These three motion types all relate to danger in the real world, because taking a wildlife example, the appearance of a dangerous animal from hiding, or sudden movements from the animal, or the animal looming towards oneself, are all events that one should attend to and be prepared to evade; the behavioural urgency hypothesis is the formulation that survival-relevant motions capture attention, and evidently looming is one of those motions (Franconeri & Simons, 2003).

Extending Franconeri and Simons’ (2003) experiment paradigm and the behavioural urgency hypothesis, a further study experimented with looming motion in the central versus peripheral visual field, and collision looming versus non-collision looming motion (Lin, Franconeri, & Enns, 2008). If the behavioural urgency hypothesis is true, i.e., attention is captured only by survival-relevant motions, then looming in the peripheral visual field should capture attention more strongly than looming in the central visual field, because the periphery is not as well processed as the centre, so looming in the periphery needs attention more urgently than looming in the centre, and secondly, collision looming would capture attention more than non-collision looming (Lin et al., 2008). The findings from experimentation exactly borne out these two predictions made from the behavioural urgency hypothesis. In one experiment, on each trial, masking objects were presented ahead of the visual search array, and in cases where the looming masking object was on a distractor, then a looming mask on a peripheral distractor slowed down visual search more than a looming mask on a distractor in the central area (Lin et al., 2008). In other words, looming in the periphery captures attention more strongly than looming in the centre, and in this context, manifests as being more distracting (Lin et al., 2008). In a second experiment, again with masking objects over the visual search array on each trial, in cases where the looming masking object was on a target, then a masking object looming on a collision path (symmetrical visual size increase) sped up visual search more than a masking object looming on a non-collision path (asymmetric visual size increase, suggesting looming above or to the participant's sides), so collision looming appears to capture attention more strongly than non-collision looming (Lin et al., 2008). In all, Lin et al.'s (2008) findings of enhanced attention capture to more urgent forms of looming (looming in the visual periphery, collision looming) fits with the behavioural urgency hypothesis (Franconeri & Simons, 2003): looming biases are driven by survival needs.

Franconeri and Simons (2003) and Lin et al. (2008) have focused on attentional biases towards visual looming signals, but it could be said that in real-world responses towards looming, attentional processes are only the first part. The subsequent motoric response to evade the looming object is also necessary for survival (Moher, Sit, & Song, 2015). The unanswered question of biased motoric responses towards looming motion was the foundation of Moher et al.'s (2015) study. Taking the basic experimental paradigm of Franconeri and Simons (2003), again there were trials of visual search with pre-search masking objects, and one of those masking objects would increase in size to simulate looming before the search array was revealed, but unique to Moher et al. (2015), the experimental task was for participants to respond by reaching out with their hands to the target element on the display, and the measurements were primarily concerned with the motoric aspects of this hand movement. Moher et al. (2015) used motion-tracking on the participant's response hand to determine both the response's timing (delay from stimulus onset to the initiation of response hand movement, duration of the response hand movement itself) and the motion of the response (the size of deviation from the most direct path from the response movement's start and finish). When the masking object loomed on the position of a target element, as opposed to the position of a distractor, then participants tended to start their motoric action with a smaller delay, the motoric action itself took less time to complete, and hand travelled along a more direct path towards the target (Moher et al., 2015). To the contrary, on trials where the looming masking object was on a distractor, the participant's hand movement was found to follow a path that curved towards where the distraction looming had occurred (Moher et al., 2015), perhaps also explaining why the timing metrics on distractor looming trials were slower than those on target looming trials. In further tests, the looming path was also varied, and on distractor elements, collision looming seemed more 'distracting' than non-collision looming (Moher et al., 2015). Nonetheless, on distractor elements, non-collision looming was still distracting, as participants' hand movements tended to follow the veer on non-collision looming (Moher et al., 2015). Altogether, Moher et al. (2015) provides evidence that visual looming signals automatically and strongly capture motoric actions, whether or not it is relevant to the experimental task. If the looming motion was on a target, then motoric actions to the target is quick to initiate, fast to complete, and direct in its motion path, but if the looming motion was on a distractor, then the looming motion distracts and even influences the hand motion to the target (Moher et al., 2015). Moher et al. (2015) presents direct evidence that, in addition to the attentional biases towards visual looming signals (Franconeri & Simons, 2003; Lin et al., 2008), there are also motoric biases towards visual looming signals.

Moher et al. (2015) is also interesting on a theoretical level. Implicitly, evasion is the 'correct' response to looming, so why would the Moher et al. (2015) paradigm of reaching out to the looming object (i.e., a non-evasive response) be valid? An evasive response is likely the best survival response towards the instance of an approaching car or dangerous animal. However, in, for example, a ball game, then standing ground and interacting with the looming ball is the more appropriate response. Against a looming ball, one could also raise one's arms to deflect the ball away from oneself (see also the infant arm-raise response to visual looming motion in Ball & Tronick, 1971). This hand motion to interact or deflect the ball was cleverly integrated into Moher et al.'s (2015) study as the arm-reach in order to respond. Moher et al. (2015) shows that the response to looming need not only be evasive (a further discussion of this topic is in Chapter 6.3).

Concluding remarks

All in all, in both auditory and visual modalities, there is a wide variety of studies which show that humans appear to respond in a multitude of protective ways towards looming signals (e.g., perceptually, behaviourally, neurologically, physiologically), but not to receding signals. This asymmetry of responses is termed the looming bias, and it appears to be robust, given the multitude of supporting studies, but also in its presence in infants and adults, humans and primates, thus reflecting the universal importance to survival of appropriately responding to looming. However, the studies reviewed have examined the looming bias in unisensory terms, i.e., separately for audition and vision. The real world is multisensory, and a single looming object can have both auditory and visual looming signals (e.g., a car, one can hear and see it approaching). The prototypical human has both senses of hearing and sight, so it is a question if humans can use both their sense of hearing and sight to enhance their percept of looming. In other words, given unisensory biases towards looming, is there also a multisensory bias towards looming?

1.2.2 Multisensory processing and the case of audio-visual looming

Real-world environments are complex; singular sensory signals may not contain enough information to successfully interact with and navigate in challenging environments (Ernst & Bülthoff, 2004). Yet, the environment often contains many sensory signals, and crucially, the same object or physical event can produce multiple sensory signals (Ernst & Bülthoff, 2004). Hence, one strategy to better perceive the environment is to combine relevant sensory signals into a coherent percept, and as the combination pulls together more information, the percept is more robust and leads to faster decisions than that possible with singular signals (Ernst & Bülthoff, 2004). Note, multisensory processing is not a simple matter of amassing as much sensory information as possible. Breaking down the environment as properties to judge, for accurate and reliable multisensory perception, the variance on the multisensory estimates of the properties must also be minimised; one proposal which minimises multisensory variance is that the contribution of each modality is weighted by its inverse variance, giving less weight to modalities producing less reliable estimates in a given situation (Alais, Newell, & Mamassian, 2010; Burr & Alais, 2006; Calcagno, Abregú, Eguía, & Vergara, 2012; Ernst & Bülthoff, 2004; as an example, think of television, the overall multisensory percept is of a person talking, even though auditorily the speech is actually coming from the speakers - vision is given more weight in this situation).

To put all the above ideas on multisensory processing into another concrete example, consider again the example in the preface (Chapter 1.1). In a complex, ambiguous environment such as crossing a difficult road with limited view, auditory perception fills the gaps in visual perception, aiding detection of approaching cars and making the crossing safer. Along similar lines, emergency vehicles have flashing lights and sirens, so with both auditory and visual signals, other road users have an abundance of sensory information to robustly and promptly detect the emergency vehicle, and make way for them (Gondan & Minakata, 2016). As either the auditory or visual signal is sufficient to make the same response (e.g., waiting on the pavement as one hears or sees the car, making way as one hears the siren or sees the flashing lights), the signals can also be termed 'redundant'. The speeded and more accurate responses towards redundant signals compared to unisensory signals are crucial for this project, and this speedup effect is termed the redundant signals effect (RSE) and has been repeatedly found in past studies (e.g., Giard & Peronnet, 1999; Hershenson, 1962; Kinchla, 1974; Miller, 1982; Todd, 1912). Given the unisensory biases towards auditory or visual looming signals, and that the cross-modal combination of sensory signals is generally advantageous for percept robustness and response speed, is there also a bias in the multisensory processing of congruent audio-visual looming signals?

To begin with, and as a check of the basic tenets, one question is whether humans are able to combine auditory and visual looming signals (or their experimental approximates) into a coherent percept to act on. In a study on human newborns only a few hours old (Orioli, Bremner, & Farroni, 2018), there were two competing displays, one showing visual size decreases (to simulate receding), the other showing visual size increases (to simulate looming), and in the centre was a speaker that played tones either decreasing in intensity (to simulate receding) or increasing in intensity (to simulate looming). Crucially, when the tone increasing in intensity was played, newborns spent more time looking at the imagery increasing in visual size, but there was otherwise no preference for either visual display when the tone decreased in intensity (Orioli et al., 2018). On a primary level, the findings in Orioli et al. (2018) show that humans do seem able to bring together auditory and visual representations of looming, and this ability appears to have been picked up early given that the newborns were only a few hours old. Second, the longer time spent attending to looming signals in the looking time paradigm of (Orioli et al., 2018) do seem to suggest an attentional preference for looming motion, which would mirror the attentional biases for looming found in auditory (e.g., Seifritz et al., 2002) and visual (e.g., Franconeri & Simons, 2003; Lin et al., 2008) looming studies. Third, (Orioli et al., 2018) used only representations of looming motion, such as the increase in visual size, or the increase in tone intensity, rather than naturalistic looming signals, yet the newborns responded to these simple representations of looming motions in a seemingly adaptive manner, just as one would towards real looming motion. Hence, there seems to be a sound basis for the use of such simple stimuli in studying looming, as other looming studies have also shown (e.g., Ball & Tronick, 1971).

Having established that, on a basic level, humans do combine auditory and visual looming signals and respond to the combination in a seemingly adaptive manner (Orioli et al., 2018), the next step is to examine the multisensory process itself, to see if there is a bias in the multisensory processing of audio-visual looming signals. The oft-cited studies that addresses a potential multisensory processing bias in audio-visual looming signals are Cappe, Thut, Romei, and Murray (2009) and Cappe, Thelen, Romei, Thut, and Murray (2012). First, Cappe et al. (2009) examined the behavioural responses towards audio-visual motion-in-depth (looming or receding), using a redundant signals paradigm. As an outline of the redundant signals paradigm, the same response (e.g., a button press) is required after detecting the auditory signal, the visual signal, or the 'redundant' combination of audio-visual signals, 'redundant' in the sense that either the auditory or visual component in the combination would have been sufficient to make a response. The metric of interest in the basic redundant signals paradigm is the response time (RT), because the indicator of a multisensory process is the RT speedup, between the RTs towards redundant conditions versus the RTs towards the unisensory components, a phenomenon known as the redundant signals effect (RSE; e.g., Giard & Peronnet, 1999; Hershenson, 1962; Kinchla, 1974; Miller, 1982; Todd, 1912). To test for the multisensory processing of motion-in-depth, Cappe et al. (2009) expanded the basic redundant signals paradigm of three sensory conditions (auditory, visual, redundant audio-visual) to include different motion conditions, such that in the unisensory conditions, the motion can be looming, receding or static (non-moving i.e., catch), while in the redundant audio-visual condition, all possible combinations of auditory and visual motions are used. Auditory motion-in-depth was simulated by intensity changes of 10 dB SPL from a start value of 77 dB SPL, while visual motion-in-depth was simulated by visual size changes of 6° from a start value of 7° on a circular object presented onscreen (Cappe et al., 2009). There were two main results: average RTs towards redundant audio-visual signals were faster compared to the average RTs towards their unisensory components, and crucially, the congruent audio-visual looming condition had the fastest RTs of all the redundant audio-visual conditions (Cappe et al., 2009). Moreover, the congruent audio-visual looming condition was rated as having the most amount of perceived motion (Cappe et al., 2009). As RTs were fastest and motion ratings were highest for the congruent audio-visual looming condition, a claim was made for a selective integrative process, ‘selective’ in the sense that the process only applies to the processing of congruent audio-visual looming signals, ‘integrative’ in reference to a speedy method of combining cross-modal signals (Cappe et al., 2009).

In a follow-up study (Cappe et al., 2012) using electroencephalography (EEG), certain waveforms had super-additive effects (the effect in redundant audio-visual conditions is greater than the sum of unisensory auditory and visual effects) unique to the congruent audio-visual looming condition. As super-addition fits with an integrative account of multisensory processing (see also Chapter 4 where sensory integration is discussed in detail), the findings of Cappe et al. (2012) were taken as further evidence of a selective integration process towards audio-visual looming signals. Altogether, the works of Cappe et al. (2009) and Cappe et al. (2012) claim a unique multisensory mechanism special to processing audio-visual looming signals, to the effect of producing especially fast responses that is presumably adaptive, in effect a looming bias on the multisensory level.

1.2.3 Knowledge gaps in the processing of audio-visual looming

From the literature review (Chapters 1.2.1 and 1.2.2), it may seem as though there is a straightforward narrative of there being unisensory looming biases, and when there are redundant audio-visual looming signals, then there is special multisensory processing that produces especially fast responses, all of which is adaptive because quick evasive responses improves survival odds (e.g., Cappe et al., 2009). However, one should cast a critical eye before accepting these claims. In reality, there are in fact shortcomings in past studies, and avenues in which the topic of multisensory looming can be taken further.

Take, for example, Cappe et al. (2009), the main study that put forward the claim of a selective integration mechanism specific to processing audio-visual looming signals. Putting their claim of mechanism aside for the moment, there are several odd points to the methodology and results of Cappe et al. (2009). One of the most striking aspects of the RT data in Cappe et al. (2009) is that they are generally quite slow, taking over 600 milliseconds on average for auditory conditions, almost 500 milliseconds in visual conditions, and above 400 milliseconds in the redundant audio-visual conditions. The slowness of RTs in Cappe et al. (2009) may raise questions about their validity – the study is about responses to looming that should be fast for survival purposes, yet the responses elicited from participants were not fast. On speculation, one possible explanation for the slow RTs is Cappe et al.'s (2009) use of 3780 trials for each participant – the sheer number of trials may have been tiring and contributed to the slow RTs. Additionally, to represent auditory looming and receding Cappe et al. (2009) used a small auditory intensity change of 10 dB SPL at very loud levels, which might have been difficult to discern from the auditory catch, and it might explain why the RTs in auditory conditions were especially slow. Apart from the RT metrics, Cappe et al. (2009) also report a response accuracy of 88% ±1.1% without an operational definition of this metric; in any case, this level of accuracy is not exemplary for the detection of quite simple stimuli. Two further odd points in Cappe et al.'s (2009) methodology are perhaps their small sample size of 16 participants, and the use of a 2x3 ANOVA to analyse the RTs, which does not cover the redundant conditions with incongruent audio-visual motion. A more pressing issue with Cappe et al. (2009) is the large gaps in RT performance between the unisensory auditory and visual conditions – at least 100 milliseconds. With such large gaps between the unisensory auditory and visual RTs, their combination in the audio-visual conditions will lead to sub-optimal multisensory processing – the auditory processing is too slow to temporally overlap with the visual processing, and activate multisensory processing (Otto, Dassy, & Mamassian, 2013). Indeed, the basis of Cappe et al.'s (2009) claim of selective integration is that RTs under the congruent audio-visual looming condition is the fastest of all the redundant conditions, but upon closer inspection, the advantage is only by a few milliseconds on absolute terms (certainly smaller than the 25 millisecond increments on their Figure 2B).

Critically, Cappe et al. (2009) has mistaken the true quantifier of multisensory processing performance, the RSE, instead using the absolute speed of the RTs as a measure of multisensory processing performance. The RSE is defined as the *speedup* of the RT in the redundant condition, versus the RT in the fastest unisensory component. If the RTs towards the unisensory components are fast, i.e., auditory looming and visual looming, then it is not surprising that the RTs towards their combination, congruent audio-visual looming, are also fast (Otto et al., 2013). As long as the RSE, the speedup, for congruent audio-visual looming is on par with the RSEs for other redundant conditions, then there cannot be a claim of special multisensory processing specific to congruent audio-visual looming conditions. However, the true RSEs were not provided in Cappe et al. (2009).

While the redundant signals paradigm as used by Cappe et al. (2009) is fundamentally sound, its implementation and interpretation by Cappe et al. (2009) seems incomplete, with unanswered questions and unknowns around their findings (e.g., what were their RSEs, what does their accuracy statistic of 88% actually refer to?). There needs to be a strong experimental foundation from which to build further experiments that answer the knowledge gaps. Hence, as the first step, a replication of Cappe et al. (2009) was performed, but with crucial modifications (e.g., quieter auditory stimuli with larger intensity changes, fewer conditions and trials per participant) to address what I think are critical basic issues of the Cappe et al. (2009) experimental design (e.g., the loud and difficult stimuli, very lengthy experiment) that could have produced problems in their data (e.g., slow RTs, disparate auditory and visual RTs). While these methodological issues are not knowledge gaps as such, the replication is important as a way to obtain a handle, an understanding of the experimental paradigm used by Cappe et al. (2009), and to see if there are novel findings to be gained from the correct analysis on the replication data. The replication acts as a baseline, and a launch point for further experiments to explore multisensory processing of looming signals.

Knowledge gaps

When it comes to knowledge gaps, the first I identified was that the auditory and visual stimuli typically used in past studies were highly simplified and unlike naturalistic looming motion. Typically, auditory motion-in-depth is represented by intensity changes, while for visual motion-in-depth, it is visual size change (e.g., Cappe et al., 2009; Franconeri & Simons, 2003; Neuhoff, 1998). Such simple motion-in-depth representations have elicited responses that seem appropriate to looming i.e., evasion, preparation for looming motion (e.g., Bach et al., 2008; Ball & Tronick, 1971; Seifritz et al., 2002). Yet, there is actually a wealth of cues to motion-in-depth that are concurrent to the main cues of auditory intensity and visual size change (see Paquier, Côté, Devillers, & Koehl, 2016; Zahorik et al., 2005 for a review of auditory and visual cues to distance - motion-in-depth is signified by the change in these distance cues). Auditory cues to motion-in-depth include changes on the direct-reverberant sound energy ratio, changes on the auditory spectrum, and the Doppler effect itself (Bach et al., 2009; Baumgartner et al., 2017; Paquier et al., 2016; Zahorik et al., 2005). Briefly, the direct-reverberant sound energy ratio works in a space with sound-reflective surfaces, e.g., a room, because in such an environment, a sound source emits sound that either reaches the listener directly, or indirectly by first bouncing off surfaces (reverberation); the closer the sound source is to the listener, the stronger the direct component of the sound is heard against the background of reverberation (Paquier et al., 2016; Zahorik et al., 2005). For the auditory spectrum cue, as the air filters high frequencies more than low frequencies, so the closer the sound source is to the listener, the more high frequency content is heard in the sound (Paquier et al., 2016; Zahorik et al., 2005). Finally, there is also the Doppler effect (Zahorik et al., 2005), which is a phenomenon observed with moving sound sources: to a listener standing still, the sound from a looming sound source is shifted into a higher pitch because the source is catching up with the sound waves it emits, in effect compressing the wavelength (i.e., higher frequency / pitch to the listener), whilst the sound from a receding sound source is shifted into a lower pitch because the source is pulling away from the sound waves it emits, in effect stretching out the wavelength (i.e., lower frequency / pitch to the listener). Visual cues to visual depth include occlusion (the occluding object is perceived to be closer), atmospheric filtering (objects appear clearer when near due to less filtering by the atmosphere), perspective cues (e.g., convergent lines, repeating patterns), proprioceptive cues (greater eye convergence angle the nearer the object is, eyes focus on near or far), and stereopsis which is the perception of three-dimensions by using both eyes together (Paquier et al., 2016).

Although there have been studies which used ‘alternative’ cues to looming motion, such as Baumgartner et al. (2017) which found looming biases using auditory spectral cues to looming rather than the conventional intensity change, it is still rather atypical to use *multiple* cues to looming (one study which did explore the use of multiple auditory cues to looming is Bach et al., 2009). Real-world looming motion contains multiple cues in both auditory and visual modalities, but the study which searched for a multisensory bias in processing audio-visual looming signals (Cappe et al., 2009) only used basic stimuli. Hence, the first knowledge gap (Chapter 3): **does the use of more realistic auditory and visual motion-in-depth stimuli produce stronger multisensory looming effects?**

The second knowledge gap is linked to the claim of selective integration towards audio-visual looming signals (Cappe et al., 2012; Cappe et al., 2009), which is a direct assertion of a multisensory processing mechanism. In terms of theories on multisensory processing mechanism, there are two main schools, integration and race, each with their own predictions of multisensory processing, but since Miller (1982) and the finding of multisensory processing in excess of that predicted by a race mechanism, viewpoints have been in favour of integration. However, more recently, there has been a realisation that the mechanisation of multisensory processing is not so clear-cut, and that the evidence nominally in support of integrative mechanisms may not be well-founded (Gondan & Minakata, 2016; Otto & Mamassian, 2017). For example, Cappe et al. (2009) applied Miller's (1982) mechanism test on their data and deduced that an integrative process, yet their application of Miller's (1982) test looked unconventional, so it was not clear if the test was indeed applied correctly – a not uncommon issue (Gondan & Minakata, 2016). Furthermore, another avenue in which integration mechanisms is supported is neuro-imaging, specifically in the findings of super-additive effects (more neural activity in audio-visual looming conditions compared to the sum of unisensory neural activities) because it mirrors integrative mechanisms (Alais et al., 2010; an example of such super-additive findings is Cappe et al., 2012). My project does not touch on neuro-imaging, but it has been noted that certain areas in the brain are known to have super-additive effects towards audio-visual stimuli, and that perhaps these findings have been over-extended as evidence of integration, when other brain regions which also handle multisensory signals do not show super-addition (Alais et al., 2010). In short, race mechanisms have historically been neglected, and have not been considered for use in understanding behaviours towards audio-visual looming signals. Thus, the second knowledge gap (Chapter 4): **to what extent can the race architecture be used to explain human decision-making towards redundant audio-visual looming signals?**

The third knowledge gap also relates to Cappe et al.'s (2009) claim of selective integration towards congruent audio-visual looming signals. It seems incomplete to claim an integrative mechanism by automatic default based only on Miller's (1982) test, and then not instantiate the proposed mechanism in a specific integrative model to test the claim (see also Gondan & Minakata, 2016 for a meta-analysis in which only a small portion of studies support claims of integrative mechanisms with actual models of integration). The recent direction in studying cognitive processes is to instantiate a hypothesis of cognitive processing into a model which specifies all the assumptions of the proposed cognitive process, thereby making the hypothesis exact and testable, and not vague like hypotheses derived only on a verbal level (Farrell & Lewandowsky, 2015). In any case, a single hypothesis of cognitive processes should not be taken as a one-size-fits-all affair, because there can be other correct hypotheses or models of a cognitive process, as the cognitive process can depend on the situation and the individual (Farrell & Lewandowsky, 2015). There has not been a computational modelling study on the multisensory processing of audio-visual looming signals, and the modelling process would definitely give insights into the existence of a multisensory looming bias. Hence, the third and final knowledge gap of this project (Chapter 5): **what can a computational modelling analysis show about the multisensory processing of audio-visual looming signals?**

1.3 Introduction to the monocular patching pilot study

This project examines how visual functioning can be improved using an eye patching technique. First, in Chapter 1.3.1, there will be an overview of the development of visual functioning during the human critical period, and how abnormal experiences can shape and form abnormal visual functioning, such as in amblyopia. Next, in Chapter 1.3.2, a new line of research into short-term monocular patching is examined, as it has been claimed that in briefly patching one eye, the patched eye becomes more dominant, and this is reflective of latent neuroplasticity in adult visual systems, which would be key to re-shaping abnormal visual functioning post-critical period (e.g., Lunghi et al., 2011). Finally, in reviewing the existing literature on short-term monocular patching, there were still unanswered questions, so in Chapter 1.3.3, the knowledge gaps for this project are formulated. These knowledge gaps guide my project on monocular patching.

1.3.1 Imperfect visual function: the critical period and amblyopia

One is not born with a visual system in its mature, fully-functional form (Knudsen, 2004). Instead, one is born with the anatomy for visual function, but there is also the need for visual experiences, such that visual abilities can be picked up (e.g., stereovision, or the ability to use both eyes to perceive depth, see also Crawford, Harwerth, Smith, & von Noorden, 1996 for a study on infant primates that became stereo-blind from being deprived of binocular stimulation), and all the components of the visual system can be calibrated and fine-tuned to the individual (Knudsen, 2004). Visual experiences shape the visual system, but crucially, there is only a limited period in which experiences can influence and shape the visual system (and other parts of the brain, more generally), and such a period is known as a sensitive period (Knudsen, 2004). Once the sensitive period has lapsed, then further experiences have a much smaller effect on influencing and shaping the brain (Knudsen, 2004). More pressingly, in a specific type of sensitive period known as 'critical period', once the critical period has lapsed, the brain is no longer receptive to further experiences and cannot be further shaped, meaning the person or animal is permanently stuck with what was formed previously (Knudsen, 2004). Problems arise when the previous experiences were abnormal, as those abnormal experiences would have shaped the brain to function abnormally, and after the critical period, the abnormal functioning would remain for life – no amount of normal experiences after the critical period can remedy the abnormal functioning (Knudsen, 2004).

The critical period can be seen in the development of visual systems, and one direction of study to demonstrate such visual system critical periods was to intentionally induce abnormal visual experiences on infant animals, and examine their visual functioning after the abnormal visual experiences (e.g., Crawford et al., 1996; Wiesel & Hubel, 1965). In a study using newly-born kittens, eyelids were sutured either monocularly or binocularly for three months, thereby preventing normal visual experiences during the kittens’ sensitive period for developing visual function (Wiesel & Hubel, 1965). It was found that no amount of visual experience after the suturing episode could remedy the kittens’ visual functioning – the damage from abnormal visual experiences had been permanently set (Wiesel & Hubel, 1965). On a behavioural level, after being relieved of the initial eyelid suturing, kittens showed behaviours that suggested impaired visual functioning when using the previously-sutured eye, such as colliding with large objects and feeling their way around the environment, and these deficits lasted through the 18 months of the study (Wiesel & Hubel, 1965). Subsequent cell recordings found much reduced neural activity on cells connected to the previously-sutured eye, and was further backed up by findings of reduced brain matter in brain sections, with no evidence of brain matter gains from receiving normal visual experiences after the initial three months of visual deprivation (Wiesel & Hubel, 1965).

In another study, infant macaque monkeys wore prism goggles which angled the view between both eyes in a way that prevented stereovision, and stereovision is what allows the perception of visual depth (Crawford et al., 1996). It was found that if the infant macaque monkey received an extended episode of prism goggles from birth, they would never have measurable stereopsis (the percept of visual depth) afterwards, but if the episode of prism goggles were given 60 days from birth, then the monkey did still have stereopsis (Crawford et al., 1996). Crawford et al. (1996) proposes that stereovision is a visual function that needs to be developed and calibrated through visual experiences, and evidently, the first 60 days of a macaque monkey’s life is formative for stereovision, such that disruptive, anti-stereo visual experiences during this period prevents normal stereovision development, thus permanently impairing the individual from seeing in depth. From another perspective, after the first 60 days, during which presumably stereovision was developed normally and locked in, then disruptive visual experiences does not appear to influence and cause damage to the already-formed stereovision capabilities (Crawford et al., 1996).

Together, the study of abnormal visual experiences in infancy, whether in kittens (Wiesel & Hubel, 1965) or in macaque monkeys (Crawford et al., 1996), provide evidence of critical periods in the development of visual functioning. Visual functioning, normal or abnormal, is a result of the visual experiences since birth, and the functioning is unchangeable after a certain age has been reached (Morishita & Hensch, 2008).

Critical periods in visual system development also apply to humans, which can be observed through the condition of amblyopia. Amblyopia is a condition that is clinically defined as reduced visual acuity in one eye, as tested by Snellen or logMAR visual tests, although there are no obvious anatomical defects in the eyes themselves (Holmes & Clarke, 2006; Maconachie & Gottlob, 2015). Typically, apart from visual acuity deficits, other visual functions such as contrast sensitivity and stereopsis are also affected in amblyopia (e.g., Chadnova, Reynaud, Clavagnier, & Hess, 2017; Daw, 1998; Ellemberg, Lewis, Maurer, & Brent, 2000; Li, Ngo, Nguyen, & Levi, 2011). Instead of anatomical defects in the eyes, amblyopia is caused by exposure to abnormal visual experiences during the critical periods for developing visual functions (Holmes & Clarke, 2006; Webber & Wood, 2005). There are three common causes for abnormal visual experiences, that without prompt treatment during the critical period, eventually leads to amblyopia: visual deprivation such as in the condition of congenital cataracts (born with an opaque lens in the eye), strabismus which is the misalignment between the two eyes, and anisometropia which is where the two eyes are very different in their optical properties (Holmes & Clarke, 2006). Common to the three causes is that they all produce binocular asymmetry in the visual experience, such as the absence of visual stimulation to one eye in the case of monocular cataracts, or the incompatibility between both eyes' views due to conflicting perspectives in the case of strabismus, or unequal clarity between both eyes in the case of anisometropia (Holmes & Clarke, 2006; Webber & Wood, 2005). Without correction of these causes, the visual system eventually has to prioritise one of the two binocularly-incompatible sides to develop and use, producing one good eye (fellow eye) which dominates the other eye (amblyopic eye), to the point of reducing visual functioning in the amblyopic eye (Holmes & Clarke, 2006; Webber & Wood, 2005).

Conventionally, the critical period for visual functioning is thought to end around age eight, so there is a need to treat amblyopia early (Daw, 1998; Webber & Wood, 2005). The general strategy of amblyopia treatment is to address the causes, then train up the amblyopic eye by forcing it to work (Holmes & Clarke, 2006; Webber & Wood, 2005). The amblyopic eye can be forced to work by patching the fellow eye, or applying an atropine eyedrop to blur the fellow eye, so the individual must use the amblyopic eye to see (Holmes & Clarke, 2006; Webber & Wood, 2005). While there is some evidence to suggest that these amblyopia treatments do improve visual acuity on the amblyopic eye (Holmes & Clarke, 2006; see also Vedamurthy et al., 2015 which found visual acuity gains in the monocularly patched control group), complete recovery of visual acuity in the amblyopic eye is rare (Maconachie & Gottlob, 2015). Furthermore, there are practical challenges to administering patching treatments: there are no standardised practices in terms of patching durations, so patching prescriptions can be several minutes each day, or all hours awake, and people may not follow through with the patching prescription in its entirety, which is not to mention the distress that patching can cause, particularly as the amblyopic patient undergoing treatment is typically young (Holmes & Clarke, 2006; Maconachie & Gottlob, 2015; Webber & Wood, 2005).

1.3.2 New directions: short-term monocular patching

Given the topics mentioned in Chapter 1.3.1, from the critical period of visual system development, whereby the brain is receptive and changing to visual experiences, to the practical example of amblyopia caused by abnormal visual experiences during the critical period, it was intriguing to read about the claims of Lunghi and colleagues (e.g., Lunghi et al., 2011). First, in the original study, adult humans had one eye occluded for 150 minutes with a translucent eyepatch, and once the eyepatch was removed, these individuals were found to be strongly dominant in the eye that was patched, and for a considerable duration (Lunghi et al., 2011). The finding of dominance on the patched eye was postulated to be a compensatory strengthening process in response to the deprivation, claimed to be a sign of visual system neuroplasticity in adults, which defies the convention of non-plasticity post-critical period, and this discovery of latent neuroplasticity could offer hope in treating amblyopia (Lunghi et al., 2011).

Further experimentation on short-term monocular patching yielded further support to the above postulations. The dominance effects found in Lunghi et al. (2011) were repeated and extended in Lunghi, Burr, and Morrone (2013), whereby dominance effects after short-term monocular patching were stronger on binocular rivalry tests using chromatic gratings instead of luminance gratings, and this could be down to the different characteristics of the colour and luminance pathways in visual processing, resulting in differing responses towards short-term monocular patching. Short-term monocular patching was found to produce increases in visual evoked potentials (VEP) towards only the patched side of the visual system (Lunghi et al., 2015a; Zhou, Baker, Simard, Saint-Amour, & Hess, 2015), and these VEPs were associated with early visual processing in areas such as V1 (Lunghi et al., 2015a) and could reflect contrast-based compensatory processes strengthening the patched eye (Zhou et al., 2015), which could be interpreted as neuroplasticity (Lunghi et al., 2015a). In an MRI study, GABA decreases were noted during short-term monocular patching, and the GABA decreases appear correlated with plasticity, which the authors argue makes GABA a mechanism for visual system plasticity as demonstrated in short-term monocular patching (Lunghi et al., 2015b). In a curious finding, the dominance effects of short-term monocular patching were apparently stronger when the person was physically active during monocular patching, compared to being still during patching (Lunghi & Sale, 2015).

Altogether, though admittedly mostly from one lead author (e.g., Lunghi et al., 2015a; Lunghi et al., 2011; Lunghi et al., 2013; Lunghi et al., 2015b; Lunghi & Sale, 2015), it appears that short-term monocular patching can quite reliably produce dominance effects on the patched eye, and the proposal is that these dominance effects reflect a process that compensates for the contrast reduction during patching by uplifting the contrast-based processes, a process that Lunghi and colleagues argue is neuroplasticity because these are neural changes in response to visual experiences. This body of research is a foundation from which more research could be done, because it raises questions – there are knowledge gaps.

1.3.3 Knowledge gaps in short-term monocular patching

The visual system develops through visual experience during the critical period, and a condition known as amblyopia may occur if the visual experience in the critical period was abnormal (Holmes & Clarke, 2006; Webber & Wood, 2005). Past the critical period, the brain is no longer receptive to new experiences (Knudsen, 2004), so for amblyopia, there is a risk of permanent deficits in visual functioning if the condition is left late (Webber & Wood, 2005). Yet, there seems to be interesting research developing on short-term monocular patching, whereby a patched eye would later undergo a compensatory increase in contrast processing, which might suggest latent neuroplasticity in adult visual systems, and could be tapped into for amblyopia treatment (e.g., Lunghi et al., 2011). However, the link between the supposed neuroplasticity effects of short-term monocular patching, and the real-world problems of treating amblyopia has not been established. There are knowledge gaps, which this project will explore, in an attempt to answer the question of short-term monocular patching as a therapeutic procedure.

First, if short-term monocular patching were to be a therapeutic procedure, then at a minimum, it must bring about positive effects to the visual system of the person receiving patching. Lunghi and colleagues (e.g., Lunghi et al., 2011) have largely focused on the immediate dominance effects of short-term monocular patching, perhaps more so as a scientific phenomenon, and not so much in actual suggestions for patching's practical considerations or therapeutic usage, such as the durability of effects in the longer-term, or effects other than dominance. Moreover, the dominance effects found by Lunghi and colleagues (e.g., Lunghi et al., 2011) were measured by the singular technique of binocular rivalry, in which conflicting gratings are presented between the two eyes of the observer, and the observer indicates which grating was seen. Binocular rivalry is not the only technique for testing eye dominance (see also Bossi, Hamm, Dahlmann-Noor, & Dakin, 2018), but crucially, eye dominance is not the only aspect of a well-functioning visual system. There needs to be an exploration on whether there are other effects of short-term monocular patching, and the longevity of the effects. Additionally, as the start of the project, I need to first see if I can replicate the basic patching effects, such as that found in Lunghi et al. (2011), using my experiment setup. Hence, the fourth knowledge gap of this PhD (Chapter 8): **are there other positive effects of short-term monocular patching?**

Next, this thesis refers to Lunghi et al.'s (e.g., 2011) patching procedures as short-term monocular patching, but the typical patching duration of existing studies is for two hours or more (e.g., Lunghi et al., 2015a; Lunghi et al., 2011; Lunghi et al., 2015b; Lunghi & Sale, 2015). In comparison, prescriptions for conventional amblyopia patching treatments range from a few minutes to several hours per day (Webber & Wood, 2005), but it has also been suggested that effective amblyopia treatment only requires one to two hours of patching daily, or even just two sessions weekly of atropine eyedrop-induced eye blurring (Holmes & Clarke, 2006). If the challenge of adhering to conventional amblyopia treatments is its lengthy duration, discomfort and inconvenience, then there is only value in looking for new therapeutic procedures if the new procedure is useful, *and* it improves on these practical challenges. On one level, short-term monocular patching is still patching, in that it does not solve the practical issues of patching discomfort and inconvenience, so perhaps the key to short-term monocular patching is in its short-duration. It has been suggested recently that visual deprivation for a few minutes can already elicit the dominance effect on the deprived eye (Kim, Kim, & Blake, 2017). In terms of patching, the current status of 'short-term monocular patching' is, in fact, not shorter (i.e., quicker to do) than effective prescriptions for conventional amblyopia patching treatments. Hence, the fifth knowledge gap of this PhD (Chapter 9): **how are the effects of short-term monocular patching when short-term monocular patching is very short-term?**

Finally, existing experiments on short-term monocular patching have only tried patching one eye (e.g., Lunghi et al., 2011). The effects of patching both eyes have not been tested before, and are therefore unknown. The thinking here is that if the dominance effects on the patched eye is actually an improvement of the patched eye, then would a procedure that involves alternately patching both eyes confer improvements to both eyes, and ultimately improve overall visual functioning? The percept of visual depth, stereopsis, is a visual function that benefits from the use of both eyes, but it is often absent in amblyopic vision (Webber & Wood, 2005); a procedure that could improve stereopsis is enticing. Hence, the sixth and final knowledge gap of this PhD (Chapter 10): **what are the effects of alternately patching both eyes, and does it improve visual functions that require both eyes, such as stereopsis?**

Section 1: Audio-visual looming project

Chapter 2: General methods of the audio-visual looming project

All four experiments conducted for this project utilised similar methods, so in this chapter, methodology common to all four experiments are described. This includes technical details and concepts of the apparatus I used. Details specific to each experiment are described in the relevant chapters.

2.1 Participants

In each of the four experiments, 20 participants were recruited by means of paper and electronic advertisements, and by word-of-mouth. Advertisements were posted within the University community, though recruitment was not limited to University staff or students – the general public may participate as well. The inclusion criteria were normal or corrected-to-normal hearing and eyesight. These criteria were enacted on the University's online experiment sign-up system (SONA), where it is possible to limit experiment sign-up only to participants who have declared normal or corrected-to-normal hearing and eyesight.

To achieve informed consent, several steps were taken. First, the advertisements describe the tasks, duration and recompense involved in the experiment (standard rate of £5 per hour, experiment scheduled for two hours, meaning £10 in recompense). This information was also available on SONA. On experiment day, before the experiment began, the researcher first gave the information sheet for the participant to read, and then also gave a verbal explanation of what the experiment entails. A quick demonstration run of the experiment task was also given so that the participant had a chance to see and try the experiment task for themselves. Throughout, the participant had opportunities to ask questions. Once the participant received all the information, the researcher solicited the participant's consent to participate, via the consent form. When the experiment was completed, the debrief form was given to the participant, and the researcher also gave a verbal explanation of the experiment, going into more detail such as explaining the experiment's purpose, past findings, and answering any questions the participant may have. The procedures comply with the *Code of Human Research Ethics* (The British Psychological Society, 2014), and have been approved internally by the *University Teaching and Research Ethics Committee* (UTREC, University of St Andrews; approval code PS12994, Appendix A).

2.2 Apparatus

The experiment setup involved multiple pieces of equipment (see Figure 2.1). Details of the apparatus are given in the following categories: computer, audio equipment, visual equipment, response and timing equipment.

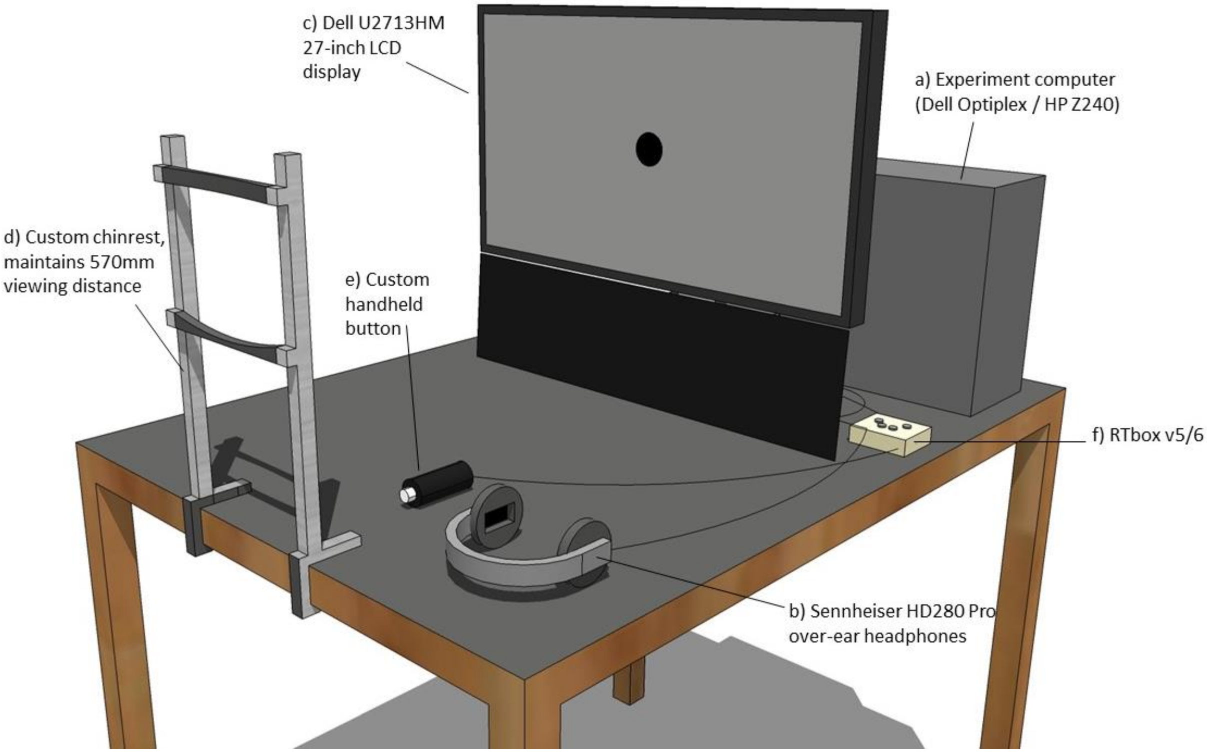


Figure 2.1. Rendering of the experiment apparatus and its setup. A desktop workstation (a) was used to run the experiment program, which outputs audio and visual stimuli via the headphones (b), and display (c), respectively. Participants were kept 570 millimetres away from the display using a custom chinrest (d). Responses were made using the custom handheld button (e), connected to a RTbox v5/6 (f) for ultimate timing accuracy and precision.

2.2.1 Computer

A desktop workstation running Matlab was used to conduct the four experiments (Figure 2.1a). Experiments 1 to 3 (Chapters 3.1, 3.2, and 4.2) were conducted on a Dell Optiplex workstation, running Windows 7 (Microsoft Corporation), Matlab 2015b (The MathWorks, Inc.) and Psychophysics Toolbox Extension version 3.0.13 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Experiment 4 (Chapter 4.3) was conducted on a HP Z240 workstation, running Windows 10 (Microsoft Corporation), Matlab 2019a (The MathWorks, Inc.) and Psychophysics

Toolbox Extension version 3.0.15 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Both computers had dedicated sound and graphics cards.

The change in computer was due to a regular lab equipment update. Calibrations were made specific to each computer setup, such that although there was a hardware change, the functions and end output (i.e., the stimulus delivery and timing) were the same for all setups. More details on the specific calibrations are provided below.

2.2.2 Audio equipment

A key part of the audio system was the sound card. Experiments 1 to 3 were conducted on the Dell Windows 7 workstation, which had a soundcard and ASIO driver combination that required a sampling frequency of 44.1 kHz. Experiment 4 was conducted on the HP workstation, which for its dedicated soundcard under Windows 10, Windows WASAPI was the soundcard driver that achieved optimal audio signal timing, and this setup required a sampling frequency of 48 kHz. Both computers were connected to a pair of Sennheiser HD280 Pro headphones (Figure 2.1b).

Calibrating the audio system is crucial, because it ensures that sounds are played at the intended intensity, i.e., a sound meant to be played at 60 dB SPL is indeed played at that value. The key point to grasp is that to generate an auditory signal in Matlab, a ‘magnitude’ number is specified and this number directly relates to the output intensity, but how this number maps onto the actual intensity as measured from the headphones depends on the hardware, from the soundcard to the headphones. It is somewhat like the volume knob on a stereo HiFi unit – it can be turned to ‘5’, but how ‘loud’ would it be at the speakers to the listener? The ‘loudness’ (or more correctly, the intensity, which is a measurable value, as opposed to ‘loudness’ which is a subjective experience) depends on the hardware involved, and so the purpose of the calibration is to obtain the magnitude number that produces 60 dB SPL at the headphones, or whatever intensity is needed.

First, to understand how calibration is performed, let us look at the calibration apparatus and physical setup. I used an artificial ear (Brüel & Kjær Type 2250, Figure 2.2a), which is a sensitive mechanical analogue of the human ear, connected to a sound level meter (Brüel & Kjær Type 4153, Figure 2.2b) which measures the sound pressure level pushed onto the artificial ear. The artificial ear is placed on a tabletop, but not the table that holds the computer tower as the minute mechanical vibrations from the computer interferes with measurement. One side of the headphone (Figure 2.2c, cut-out perspective shown) is positioned approximately centrally on top of the artificial ear, and thus the other side is used to clamp the whole setup to the tabletop (Figure 2.2). This headphone positioning on the artificial ear largely mimics how the headphone sits in relation to a human listener's ears. Only one side of the headphones was measured because the same signal is played to both sides, and it was assumed that both sides of the headphones are equal.

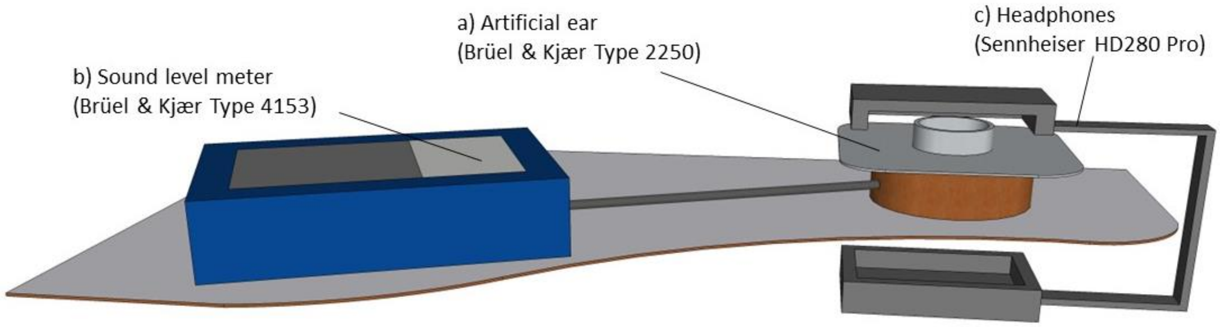


Figure 2.2. Hardware and setup for audio calibration. a) An artificial ear is the mechanical analogue of the human ear, and takes the sound pressure detected to b) the sound level meter which gives a reading of sound intensity in decibels SPL. c) One side of the headphones were placed on the artificial ear (cut-out view shown in Figure), in a similar distance and positioning as one would place headphones on their ears, and the auditory stimulus were played through the headphones for measurement and ultimately, calibration of the auditory stimulus' intensity.

The next component to audio calibration is the software. I created a simple Matlab script which plays the experiment's soundwave, a 1000Hz pure tone, at a constant intensity for 10 seconds. The sound was played for 10 seconds so that there was enough time to have a stable measurement of the actual intensity on the sound level meter. In order to make the 'magnitude' number easier to handle, in the Matlab code I specify the actual intensity I want, e.g. 60 decibels, and then convert it to magnitude using the *db2mag* function. The problem with *db2mag(60)* is that it produces a magnitude number of 1000 – much too loud. The solution was to multiply *db2mag(60)* by a reduction factor. Therefore, the goal of the calibration software is to find the reduction factor that maps a *db2mag([desired decibel number])* input directly onto the correct intensity output as measured at the headphones.

With this setup, it was a process of playing the experiments' 1000Hz pure tone and aiming for 60 dB SPL, then taking the reading from the sound level meter, and adjusting the reduction factor to cancel the difference between actual and intended output intensity. It was determined that for a 1000Hz pure tone, the reduction factor should be 2.6608 x 10^{-6} for the Dell workstation used in Experiments 1 to 3, while it should be 1.9953 x 10^{-6} for the HP workstation used in Experiment 4.

One may note that the experiments actually played the 1000Hz pure tone at a start value of 40 dB SPL, with endpoints at either 20 dB SPL, 40 dB SPL, or 60 dB SPL, so why was calibration performed at 60 dB SPL? The issue is that it is quite difficult to have an environment with zero background noise. Despite the laboratory's location in a quiet place, background noise was still detected and would interfere with, if not mask, the calibration sound if it was played at 20 dB SPL – it would be uncertain what exactly was measured by the sound level meter, and the calibration would most likely be incorrect. Hence, calibration was performed at 60 dB SPL, because that level is clearly above the background noise, so it is certain that the calibration equipment is measuring and calibrating the signal. To be sure, I checked that the reduction factor obtained from 60 dB SPL also correctly maps *db2mag(40)* to a true 40 decibels SPL at the headphones. By extension, the same calibration should work in mapping *db2mag(20)* to 20 dB SPL.

2.2.3 Visual equipment

Visual equipment consists of a 27-inch LCD display (Dell U2713HM, 60Hz refresh, set at 1920 x 1080 resolution, Figure 2.1c) and a custom-made chinrest (Figure 2.1d). The chinrest kept all participants centred to and 570mm away from the display. This particular chinrest had chin and head cushions separately adjustable for height. The chin and head cushions were fixed in place so that when one rests their head there, one's field of view naturally centres with the centre of the display. As the chinrest was fixed, for participant comfort, the participant's seat was height-adjustable. Before the experiment began, participants were encouraged to find a comfortable seat height to reach the chinrest.

2.2.4 Response and timing equipment

A custom-made handheld button (Figure 2.1e) connected to a RTbox v5/6 (Li, Liang, Kleiner, & Lu, 2010; Figure 2.1f) made up the response and timing equipment. The handheld button is a small cylindrical device which one holds on to for the duration of the experiment. To make a response in the experiment, one presses the small button on the end of the cylinder (the white protruding part on the end of the cylinder, Figure 2.1e). The button actuation has a tactile and audible click, so the participant knows they have made the response and can release the button. This handheld button is the only device the participant uses throughout the experiment.

The handheld button is an accessory to the RTbox v5/6 (Figure 2.1f). The RTbox itself is a very important piece of equipment, and it was used for two main functions: collect response times, and calibrate the audio-visual equipment. But first, what exactly is an RTbox?

The raison d'être of the RTbox becomes clear when one takes a closer look at standard computer input devices, such as the keyboard. A keyboard can be programmed to act as a response time collection device, e.g., stimuli is presented, the participant presses a certain key on the keyboard to respond, and the response time is the difference between stimulus onset time and key press time. The problem is that keyboards are meant for typing, i.e., word processing, where precise timing of key actuations is not paramount, thus they are built without precise timing in mind and would return variable key press times (Li et al., 2010). There are five ways in which keyboards are compromised in their timing accuracy. First, there is lag in the keyboard mechanicals, such that there is a slight delay between the physical button press and that press activating the key's switch on the keyboard's electrical circuit, and therefore registering as a keypress on the computer (Li et al., 2010). Second, after a single key press, the key mechanism produces excess alternations on the key's switch, known as bouncing, so debouncing software is needed to prevent nuisance bounces from registering as multiple keypresses, but this solution introduces delays in registering the real keypress (Li et al., 2010). Third, a keyboard can contain over a hundred keys, and they are not individually and directly wired to the keyboard's encoder chip, because this would be expensive and the timing benefits unnecessary for typing (Li et al., 2010). Instead, the keys' wiring is organised as a grid, so the encoder chip serially scans the single electrical switch at the end of each column and row of the grid, to determine which key, if any, was pressed; the keypress' timing however is variable because it depends on where the encoder chip was scanning at the time of the keypress (Li et al., 2010). Fourth, the computer only periodically retrieves information from the keyboard, so there is timing variability dependent on when the keypress was in relation to the computer's retrieval cycle (Li et al., 2010). Fifth and finally, the computer does not necessarily process the keyboard press immediately – there may be a slight delay if the computer is busy with other tasks (Li et al., 2010). Altogether, these imperfections can amount to a timing variability of up to 70 milliseconds (Li et al., 2010), which would be problematic for the purposes of examining small differences in response times between experiment conditions. This project requires precise timing to determine how looming motion is processed in the auditory and visual modalities.

The ideal device for collecting response times should therefore have good button mechanicals which reduces mechanical lag and obviates the need for heavy-handed debouncing software. It should have only the necessary number of buttons so that each button is directly wired to the controller chip, and avoids the problem of serially scanning electrical switches on keyboards. It should also have its own logic board, complete with microprocessor, accurate and precise clock, and storage, so that button press events are recorded and stored locally on the device, with accurate and precise timing synchronised to the computer's clock from the start. This ready-made and time-accurate data can then be retrieved by the computer when it is ready, avoiding the need for the computer to handle raw button press data live, and with variable delays added to the data. The RTbox is precisely such a device (Li et al., 2010). With the RTbox as a dedicated response time collection device, the response times collected in the experiments were accurate and precise, giving the best technical basis for examining how motion processing differs between sensory modalities.

With such a precise and accurate clock, the RTbox also lends itself well to timing the computer's audio and visual equipment (see above), such that any timing discrepancy between the two systems can be addressed, with the aim of presenting auditory and visual signals synchronously in the experiments. A key feature of the RTbox v5/6 I used is that it has an audio input port, and an input port for connecting to a photodiode. The function of these input ports is to allow an auditory or visual signal (e.g., an audio signal carried by an audio line, or a burst of light hitting the photodiode) to produce a 'button press' event on the RTbox, instead of a person pressing on the RTbox button. From the audio and visual equipment 'button presses' on the RTbox, as produced by the experiment's Matlab code, latencies of the computer's audio and visual equipment while running the experiment were obtained. The aim is for the latency of the computer's audio system to at least be similar to the latency of the computer's visual system.

In practice, this is how the synchronisation check was done: the experiment's Matlab script was adapted to have a synchronisation check mode, where the basic experiment runs as-is, but with the audio and photodiode lines connected to the RTbox. The photodiode was placed on the display exactly where the stimuli would appear (in this synchronisation check mode, the visual stimulus is white to trigger the photodiode). The sounding and appearing of the audio-visual stimuli were recorded on the RTbox as events. The synchronisation check mode checks the audio and visual equipment 60 times in one go. From the RTbox, the onset latencies of the auditory and visual events were graphed (Figure 2.3). Examining the graph, one can see if the audio and visual equipment are centred to each other (synchronous) over the 60 checks. If asynchronous, the experiment settings were adjusted. For example, if the visual equipment was slower than the audio equipment, then a line-execution timing check was done to see if there were slow to execute lines of code relating to visual equipment, and improve on them. Or perhaps the audio system latency was large and variable, in which case a check was done to see if other soundcard drivers may produce a more stable audio system latency. For the new computer (for the final experiment, Chapter 4.3), best audio-visual synchronisation was achieved using the Windows WASAPI soundcard driver, and these settings were used. These settings achieved an onset jitter not exceeding 3 milliseconds in either direction (Figure 2.3).

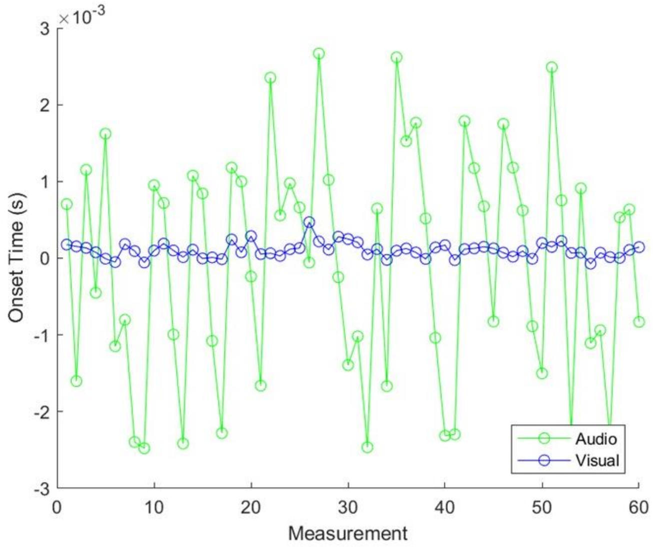


Figure 2.3. Audio-visual synchronisation check results, from the final experiment (Chapter 4.3). The earlier experiments used an older computer which had been calibrated. After hours of tweaking, the best results were obtained by using the Windows WASAPI soundcard driver. Synchronisation is seen by the latencies of the auditory (green) and visual (blue) systems largely centred with each other near zero onset delay, over the course of 60 measurements.

2.3 Experiment design

2.3.1 Redundant signals paradigm as the basis

The aim of this project is to investigate whether there is combined audio-visual processing of looming motion, which purportedly is an especially quick form of processing (Cappe et al., 2009). To investigate this proposition, the experiment design should compare response times in audio-only, visual-only, and audio-visual conditions. Such an experiment paradigm is known as the redundant signals paradigm (Giard & Peronnet, 1999; Hershenson, 1962; Kinchla, 1974; Miller, 1982; Todd, 1912). To recap, the basis of the redundant signals paradigm is that both audio and visual signals confer the same information, so either one is sufficient to elicit a response – a logical OR between audio and visual channels (Figure 2.4). When *both* auditory and visual signals are present, then one signal is *redundant* because in terms of eliciting a response, it is not necessary to have both signals. But with both signals available, does one produce a response sooner, compared to having just one of the two signals?

Transferring the paradigm to this project, the experiments have three sensory conditions: auditory-only signals, visual-only signals, or audio-visual signals. The participant responds to the onset of these sensory signals as quickly as possible. Afterwards, response times in the three conditions are compared to see if there are systematic differences, such as faster responses in audio-visual conditions than with the unisensory conditions.

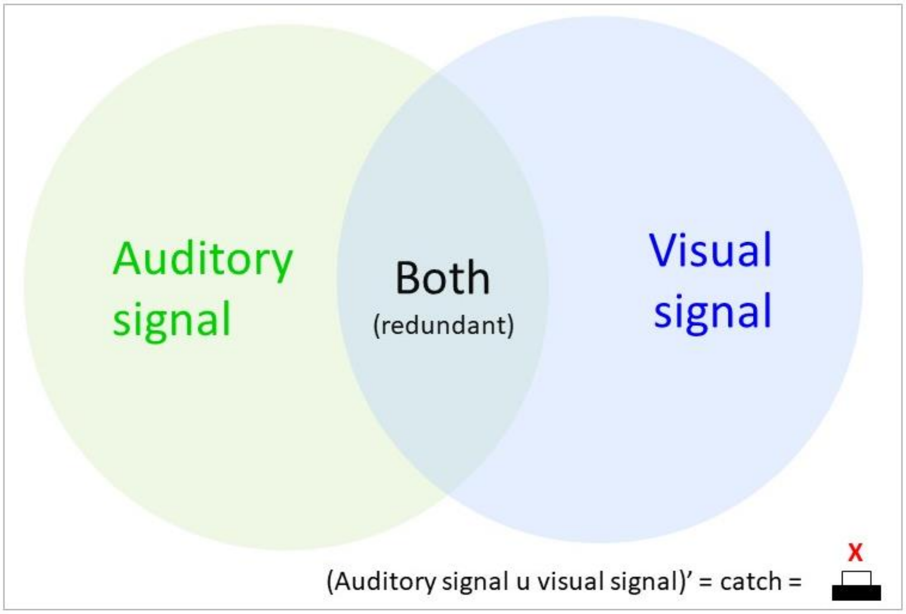


Figure 2.4. The basic redundant signals paradigm, shown as mathematical sets. One should respond to the set of auditory signals, and the set of visual signals, as well as their intersection (i.e., redundant signals). Outside of the union between auditory and visual signals sets are catch trials, which the participant should not respond to.

2.3.2 Extending the basic redundant signals paradigm

The redundant signals paradigm forms the core of the experiments, but the paradigm only addresses the first part of the research question: whether audio-visual processing is different to audio-only or visual-only processing. The second part of the research question looks into whether sensory processing varies with looming and receding motion. Thus, the classic redundant signals paradigm was extended, such that in each modality (audio, visual, audio-visual), there is both a looming variant and a receding variant of the signal (specifics below in Chapter 2.4 *Stimuli*).

In detail, the audio-only condition has two possible signal variants (Figure 2.5a, green area): auditory looming (AL) and auditory receding (AR). Similarly, the visual-only condition has two possible variants (Figure 2.5b, blue area): visual looming (VL) and visual receding (VR).

To investigate audio-visual processing, signals from the audio-only condition (two signal variants) and the visual-only condition (two signal variants) were combined in a 2x2 design, producing four multisensory variants (Figure 2.5c, grey area): auditory looming + visual looming

(ALVL), auditory looming + visual receding (ALVR), auditory receding + visual looming (ARVL), and auditory receding + visual receding (ARVR).

In total, there were eight signals, covering the variations in sensory modality and motion direction (all of Figure 2.5). These eight signals allow for comparisons between sensory modalities and motion directions on a unisensory level, comparisons between unisensory and multisensory conditions, and comparisons within the multisensory conditions, such as congruent multisensory (ALVL, ARVR) versus incongruent (ALVR, ARVL), and in particular whether responses to ALVL is special among the multisensory conditions.

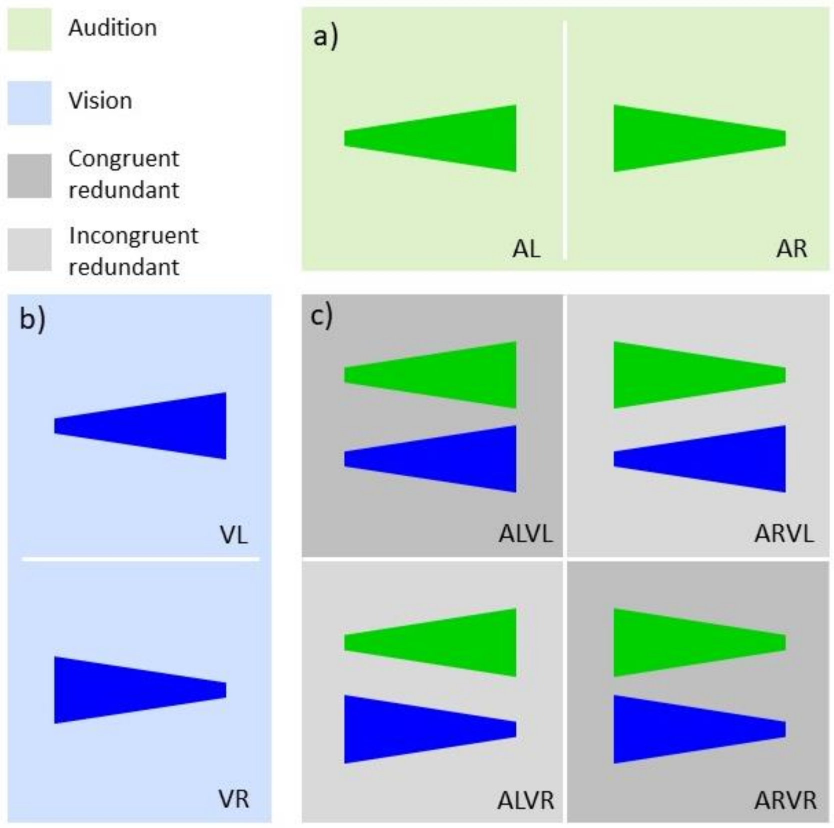


Figure 2.5. The eight signal conditions in the experiment. a) the two auditory signal variants, auditory looming and auditory receding. b) the two visual signal variants, visual looming and visual receding. c) the 2x2 combination of the auditory signal variants and the visual signal variants, making four redundant conditions, consisting of congruent (ALVL, ARVR) and incongruent (ARVL, ALVR) combinations.

In experiments, the participant should respond on trials featuring any of the eight signals described above. However, if every trial were to feature a signal, then one can mindlessly press the response button on every trial, meaning such responses are not about sensory processing. Hence, catch trials were randomly interleaved among the signal trials. Catch trials are trials with the onset of the auditory and/or visual stimuli, but the stimuli do not change after onset, i.e., they do not simulate motion, hence the participant should not respond. There were three types of catch trials, one each for audio-only, visual-only and audio-visual (specifics below in Chapter 2.4 *Stimuli*). Participant performance depends on correctly responding on the signal trials, and not responding on the catch trials (more on this in Chapters 2.5 *Procedures* and 2.6 *Data analysis*).

2.4 Stimuli

In the four experiments, the stimuli revolved around the same core concept. Namely, to simulate looming motion, the auditory stimuli increased in intensity, while the visual stimuli increased in visual size. To simulate receding motion, the auditory stimuli decreased in intensity, while the visual stimuli decreased in visual size. The stimuli in Experiment 2 is a special case (Chapter 3.2), because the experiment is an exploration into more realistic stimuli, so the auditory stimuli featured a frequency change *in addition to* the core intensity change, while the same visual signal was presented against a background with visual perspective cues. Below are the specifics of the core auditory and visual stimuli, which were used as-is for Experiments 1, 3 and 4. Specifics of the Experiment 2 modifications are described in Chapter 3.2.

2.4.1 Auditory signal

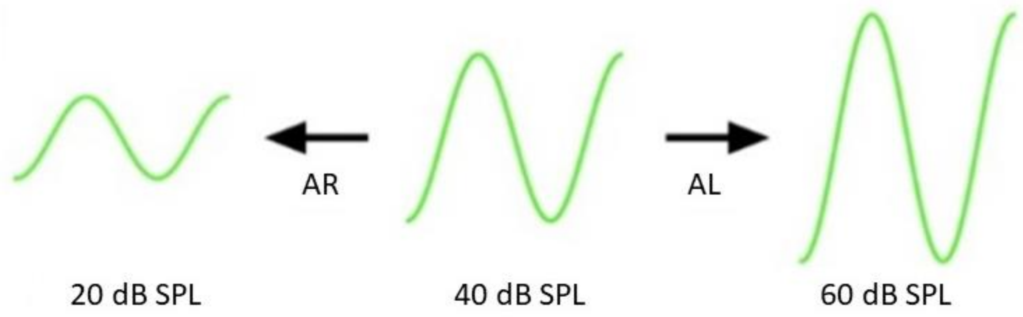


Figure 2.6. The basic auditory signal which was a change of 20 dB SPL from 40 dB SPL to simulate receding (AR) or looming (AL) motion. The stylised waveform has been coloured green to signify an auditory signal, as part of the colour-coding used in Figure 2.5.

In Cappe et al. (2009), their 1000 Hz auditory stimuli had an initial intensity of 77 dB SPL, and went up 10 dB SPL to simulate looming, went down 10 dB SPL to simulate receding. With my audio apparatus, it was found that already by 65 dB SPL, a 1000 Hz pure tone felt uncomfortable. Therefore, a limit of 60 dB SPL was set as the maximum intensity for whatever auditory stimuli would be played in my own experiments. Thus, for all four experiments, I settled for a 1000 Hz pure tone that initially would be at 40 dB SPL. To simulate looming motion, the tone linearly increased over a period of 0.5 seconds to 60 dB SPL, i.e., the maximum intensity. To simulate receding motion, the tone linearly decreased to 20 dB SPL over a period of 0.5 seconds. These auditory changes are represented in Figure 2.6. An intensity change of ± 20 dB SPL was used instead of Cappe et al.'s (2009) ± 10 dB SPL, because in my tests, I felt that at the lower intensity levels I was using, a 20 dB SPL change was reliably noticeable whereas a 10 dB SPL change was not. Unique to Experiment 2, the auditory stimuli also featured a frequency change in addition to the aforementioned intensity change – more on this in Chapter 3.2.

2.4.2 Visual stimuli

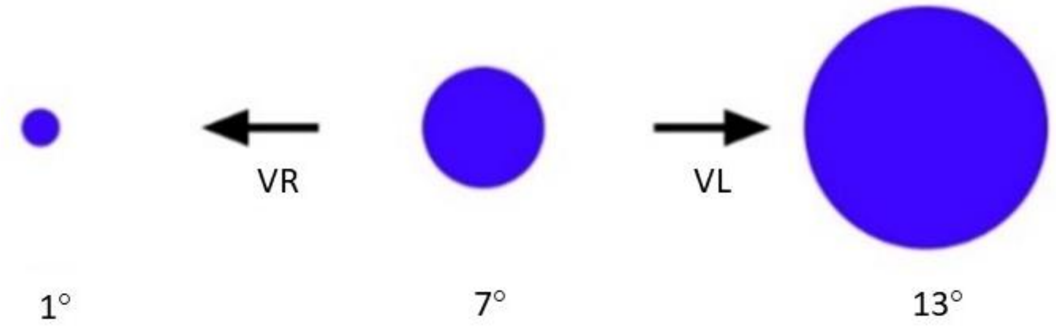


Figure 2.7. The basic visual signal of a circular object changing in size by 6° from the starting size of 7°, to simulate receding (VR) or looming (VL) motion. The circular object has been coloured blue in this figure to signify a visual signal, as part of the colour-coding in Figure 2.5. In the experiment, the visual object was black.

For the visual stimuli, there were also some modifications to Cappe et al.'s (2009) design. Cappe et al. (2009) used the maximum contrast possible, which was either a black visual target against a white background, or a white visual target against a black background. In early testing with my visual equipment, the target, which is a quick-moving shape, appeared to glow (blooming) and leave a brief trail of faint doubles (ghosting) if it was in such high contrast against the background. Thus, the contrast was lowered by using a black visual target against a mid-grey background, but otherwise kept to Cappe et al.'s (2009) design. The visual stimuli was therefore a black disc which appeared at the centre of the display, with an initial diameter of 7°. To simulate looming motion, the disc linearly increased in size to a final diameter of 13°, over a period of 0.5 seconds. To simulate receding motion, the disc linearly decreased to a final diameter of 1°, over a period of 0.5 seconds. These visual changes are represented in Figure 2.7. Unique to Experiment 2 is the use of an image background with perspective cues, instead of a plain background – more on this in Chapter 3.2.

2.4.3 Catch trials

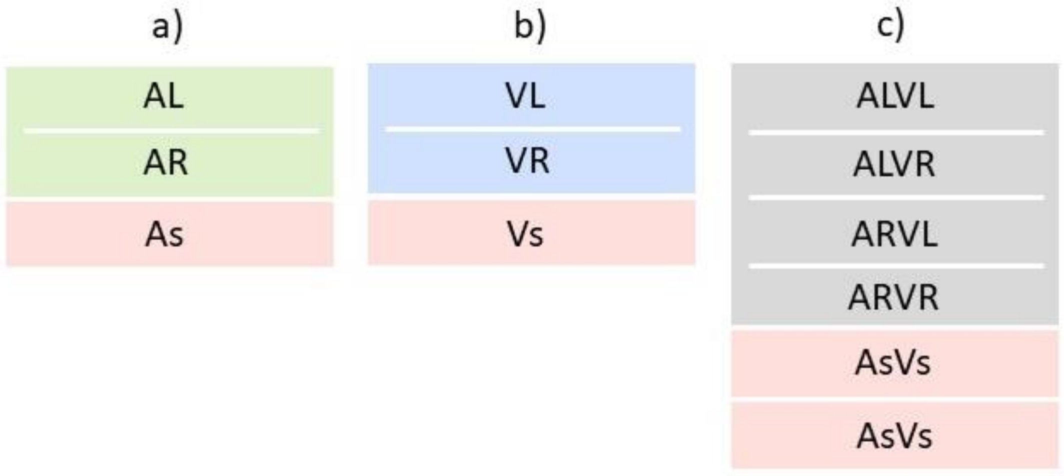


Figure 2.8. The complete set of trial types, featuring signal and catch trials. An experimental block repeats this set of trials five times, the trials in random order. Note that in this set of trials, for each sensory modality, there is the same 2:1 ratio of signal to catch trials. a) In the auditory modality, there is auditory looming (AL), auditory receding (AR), and an auditory catch (As). b) In the visual modality, there is visual looming (VL), visual receding (VR), and a visual catch (Vs). c) In the redundant audio-visual condition, there are four signal trials resulting from the 2x2 combination of the aforementioned auditory and visual signal types, and correspondingly, the audio-visual catch (AsVs) is presented at double frequency to maintain the 2:1 signal-to-catch ratio.

The catch trials were simply the stimulus at the initial audio intensity and visual size, for the duration of the trial. So, the auditory catch trial is the 1000 Hz pure tone played at a constant 40 dB SPL (Auditory static, As). The visual catch trial is the disc shown onscreen at a constant 7° diameter (Visual static, Vs). The audio-visual catch trial is the presentation of both the constant 40 dB SPL 1000 Hz pure tone and the constant 7° disc (Auditory static Visual static, AsVs). Note that these three catch trial types correspond to the three basic sensory conditions of auditory, visual and audio-visual. Thus, there is one auditory catch for the two auditory motion signal variants (Figure 2.8a), and one visual catch for the two visual motion signal variants (Figure 2.8b). There were four multisensory motion signal variants, so to maintain a 2:1 signal to catch ratio, the audio-visual catch was presented at double frequency (Figure 2.8c).

2.5 Procedures

2.5.1 Trial presentation

Once the participant felt comfortable after a demo run of the experiment, and had given their informed consent to participate (see also Chapter 2.1 *Participants* for details), the experiment began. At the beginning of each trial was a foreperiod of random duration, which was necessary because if every trial had the same foreperiod, then the participant very likely will grasp the timing and press the response button by rhythm, and not as a genuine response to the stimuli. Rhythmic button presses would manifest as unreasonably short response times. With a random foreperiod, one must pay attention to the stimuli, because if one were to rely on rhythm, then such anticipatory button presses could come *before* the stimuli were even presented – the response would be marked as incorrect if this were to happen (more on the experiment's response and feedback logic below). Operationally, the random foreperiod duration was defined as 0.75 seconds plus a random component sampled from an exponential distribution with a mean of 0.25 seconds. This meant that the foreperiod had an absolute minimum of 0.75 seconds, with a mean of around one second.

What happens during the foreperiod depends on the experiment. Experiments 1 (Chapter 3.1) and 2 (Chapter 3.2) followed Cappe et al. (2009) in having a foreperiod with no stimulation at all. No sound was played and the participant only saw a blank display for the duration of the foreperiod. In Experiments 3 (Chapter 4.2) and 4 (Chapter 4.3), I changed what happens during the foreperiod: the foreperiod features auditory and/or visual stimulation at the initial intensity (40 dB SPL) / size (7°), which therefore acts as a lead-in to the stimulus that was presented immediately after the foreperiod. The lead-in foreperiod meant the onset of stimulation was dealt with early in the trial where it is non-critical (the onset itself is irrelevant for making a response), thereby making stimulus motion distinct, as it is the stimulus motion which is the signal to make a response. The rationale and effect of having a lead-in foreperiod is explained in more detail in Chapter 4.2, but in brief, having stimulus motion – the signal to make a response – distinct from its irrelevant onset made for much smaller sensory processing artifacts, which in turn led to higher response accuracy, compared to Cappe et al. (2009), and my own Experiments 1 and 2, which all had concurrent stimulus onset and motion.

Immediately after the foreperiod, one of the eight motion signals, or a catch trial, was presented (see Figure 2.5 for the eight motion signals; Figure 2.8 for the complete set of signal and catch trials). The signal lasted for 0.5 seconds, and stayed at the final intensity/size for the remainder of the trial.

2.5.2 Response and feedback logic

When a response was recorded, or at the end of the response period of 1.5 seconds, all stimulation stopped, and feedback was given for 0.5 seconds. The following is the response and feedback logic. On motion signal trials, a response within the response period was considered a hit, which was indicated by showing a green ‘Correct’ on the feedback screen. If there was no response on the motion signal trial, then this was considered a miss, and a red ‘Miss’ was shown on the feedback screen. On catch trials, a green ‘Correct’ was shown if no response was made, while responding resulted in a red ‘False alarm’ being shown. If a response was made during the foreperiod, i.e., before any signal was presented, then a red ‘False alarm’ was shown. After feedback, there was 0.5 seconds of no stimulation (blank screen, no audio) as a brief rest for the participant, and then the next trial would begin.

2.5.3 Data collection and timing

An experiment session was organised into 20 blocks. Each block had 60 trials, made of five runs of the 12 items shown in Figure 2.8, in random order. Thus, there were 40 motion signal trials and 20 catch trials in each block. With 20 blocks in total for each session, there were 800 motion signal trials. Given eight motion types (Figure 2.5), in a full session with 800 motion signal trials, each participant therefore did 100 trials per motion type. A block of trials took approximately three minutes to complete. Participants took breaks of a few minutes typically after every four blocks, though additional breaks were allowed between blocks. A session took approximately 90 minutes to complete. Each experiment had 20 participants. Hence, each experiment has a data set of 24,000 trials, of which 16,000 were motion signal trials. There were four experiments, resulting in four such data sets.

2.6 Data analysis

The core metric across the four experiments is response time (RT), which is the latency from stimulus onset to the button press on the custom-made handheld button connected to the RTbox (see Chapter 2.2.4 *Response and timing equipment*). The RT is a commonly used metric for studying decision-making in sensory research (e.g., Innes & Otto, 2019), because RT is largely composed of the time used in decision-making, the remainder of RT just a small and fixed amount due to neuro-motor latencies (Noorani & Carpenter, 2016). Before analysing the RT, the data set must be inspected and corrected first, in the steps described below in *Data inspection* (2.6.1), *Removal of foreperiod 'responses'* (2.6.2), and *Outlier correction* (2.6.3). To study how audio-visual conditions are processed differently than in unisensory conditions, the RTs under different sensory conditions were compared, but a simple comparison of average RTs would be inadequate (Noorani & Carpenter, 2011; Otto, 2019). Thus, in the latter part of this section, the technique of processing RTs into RT distributions, for geometric comparison across sensory conditions to obtain the redundant signals effect (RSE), is explained.

2.6.1 Data inspection

As a precautionary first step, the data was checked for problems suggesting contradiction with the participant inclusion criterion of normal or corrected-to-normal hearing and eyesight. This inclusion criterion is important because this project is on sensory processing, so functioning sensory systems are required. In practice, participants were asked about the inclusion criteria when recruited via SONA (see also Chapter 2.1 *Participants*), however, some participants may still face difficulties perceiving the stimuli. To account for such difficulties, the experiments were programmed to give accuracy performance by block, which was inspected as the experiment progressed. If the hit rate was under 80% for more than one block, then that participant's data set would be rejected because that level of accuracy for highly salient stimuli may indicate potential hearing or visual difficulties, which are not the focus of this study. Should a participant be rejected, then another participant would be recruited, so that there was always a total of 20 participants in each of the four experiments. Across the four experiments, there were two such replacements.

2.6.2 Removal of foreperiod 'responses'

In each experiment, its complete data set was examined for trials where the participant 'responded' in the foreperiod, which is incorrect as such a 'response' took place before any stimulus was presented. Such occurrences are foreperiod false alarms (termed Foreperiod FA in my data tables). Foreperiod false alarms are unusual because they can occur in both signal and catch trials, whereas a false alarm, strictly, can only occur on catch trials. The highest foreperiod false alarm rate across all four experiments is 0.60% ±0.20%, averaged across participants in the visual looming (VL) condition (Table 2.1, Experiment 4), a small amount. Foreperiod false alarms are identifiable by their negative response times, and they were removed from analysis to avoid incorrectly inflating the false alarm rate.

	Foreperiod FA (%)			
	Experiment 1	**Experiment 2**	**Experiment 3**	**Experiment 4**
AR	0.00 ±0.00	0.00 ±0.00	0.20 ±0.09	0.00 ±0.00
AL	0.10 ±0.07	0.20 ±0.09	0.20 ±0.12	0.25 ±0.10
VR	0.15 ±0.08	0.05 ±0.05	0.00 ±0.00	0.25 ±0.12
VL	0.10 ±0.07	0.05 ±0.05	0.10 ±0.10	0.60 ±0.20
ARVR	0.15 ±0.08	0.05 ±0.05	0.45 ±0.14	0.10 ±0.07
ARVL	0.25 ±0.10	0.10 ±0.07	0.25 ±0.12	0.35 ±0.11
ALVR	0.10 ± 0.07	0.10 ±0.07	0.35 ±0.18	0.20 ±0.09
ALVL	0.05 ±0.05	0.20 ±0.12	0.45 ±0.15	0.20 ±0.09
asvs	0.15 ±0.08	0.08 ±0.04	0.43 ±0.14	0.28 ±0.05
as	0.10 ± 0.07	0.10 ±0.10	0.00 ±0.00	-
vs	0.15 ±0.08	0.05 ±0.05	0.10 ±0.07	-

Table 2.1. The participant-averaged occurrence of foreperiod false alarms (foreperiod FA) as a percentage of the trials, for each condition. The highest foreperiod FA rate was in Experiment 4 for the visual looming (VL) condition, at a participant average of 0.60% ±0.20% of trials (highlighted in pink). This represents a low occurrence of foreperiod FAs.

2.6.3 Outlier correction

With a data set cleaned of foreperiod false alarms, the next step was outlier correction to remove responses that were too fast (likely an early false alarm falling within the valid response window), or too slow (caused, for example, by lapses of attention). An outlier correction was performed separately per condition per participant. The procedure was performed on the reciprocal of RTs (1/RT), using bounds that were defined as ± 1.4826 * 3 median absolute deviations (MAD) around the median (Leys, Ley, Klein, Bernard, & Licata, 2013; Otto, 2019).

In more detail, note that the outlier correction was performed on 1/RT rates, not the raw RT. The problem with raw response times is that typically, RTs are positively skewed in their distribution, meaning there is a propensity for slow RTs (Noorani & Carpenter, 2016; Otto, 2019; Ratcliff, 1979). With an asymmetric, i.e., non-normal, RT distribution, an outlier correction, which works by trimming a set amount either side of the distribution's average, would be incorrect. The solution is to take the reciprocal of raw RT, because 1/RT is approximately normally distributed ('reci-normal'; Noorani & Carpenter, 2016). As there are equal amounts of data either side of the average in a normal distribution, outlier correction is possible and most likely accurate with 1/RT. The use of 1/RT distribution also corresponds with the LATER model (Noorani & Carpenter, 2016) where sensory information accumulates at a linear rate (drift), but the rate varies trial-by-trial according to a normal distribution. The LATER model (Noorani & Carpenter, 2016) forms the basis of the modelling work in this project (more on this in Chapter 5).

The second technical note about outlier correction is the use of median and MAD, instead of mean and standard deviation. In defining the central tendency and hence the bounds of outlier correction, there is the danger that the measure of central tendency is biased, perhaps pulled by outlying values, thereby causing problems in the outlier correction process. The mean is known to be easily pulled by extreme values, whereas the median is more robust in this regard. Hence, as another tool to prevent biased outlier correction, the median was used as the measure of central tendency in the 1/RT distribution, with the associated MAD bounds.

On a final note, outlier correction is in effect data trimming which is controversial. However, I would argue that first, in principle it is necessary to remove invalid data from analysis. Second, the formulation of ± 1.4826 * 3 MAD around the median corresponds to ± 3 standard deviations around the mean if 1/RT is normally distributed, in which case only a conservative amount (0.27%) of data points would be removed (Leys et al., 2013; Otto, 2019).

2.6.4 Accuracy performance

Before proceeding with the main RT-based analyses, a recommended procedure of checking the accuracy performance was performed (Otto, 2019). Ceiling levels of accuracy performance are required, because the modelling stages of this project are based on an adaptation (Otto & Mamassian, 2012; Otto & Mamassian, 2017) of the LATER model of sensory information accumulation to a decision threshold (Noorani & Carpenter, 2016). The LATER model of sensory information accumulation only holds true for simple detection tasks, featuring highly salient stimuli which should produce ceiling levels of accuracy performance (Carpenter, Reddi, & Anderson, 2009; Otto, 2019). A more extensive explanation of the modelling work in this project is in Chapter 5.

2.6.5 Redundant signals effect

The RSE is the RT speedup for the redundant condition (i.e., the audio-visual condition in this study) compared to the unisensory components. The audio-visual condition is termed ‘redundant’ because either the auditory signal or the visual signal alone is sufficient to produce a response (Otto, 2019). The RSE is a well-known phenomenon frequently observed in experiments (Giard & Peronnet, 1999; Hershenson, 1962; Kinchla, 1974; Miller, 1982; Todd, 1912). The size of the RSE reflects the benefit of using multiple senses together, for a large RSE implies much faster responses in the redundant condition compared to its unisensory components. Fast responses are supposedly beneficial, for example in evading impending collisions as signalled by auditory or visual looming, and it has been claimed that the benefit (the RSE) is especially large for audio-visual looming signals (Cappe et al., 2009). Thus, as this project investigates these ideas, it is important to accurately quantify the RSE.

Based on the verbal definition of the RSE, it is tempting to quantify the RSE by simply calculating the difference between the mean of RTs towards the redundant condition, and the mean of RTs towards the faster of the two unisensory component conditions. Indeed, this is the operational definition of the RSE one may occasionally find in literature. However, taking the difference in average RTs as the RSE is arguably not the best practice. To understand the inadequacy of average RTs, allow a quick diversion explaining the intricacies of RT. On a behavioural level, one of the key characteristics of RT is that it varies from instance to instance, even if each instance was a response to the same stimuli (Noorani & Carpenter, 2011). Hence, in an experiment where typically there are multiple trials of an experimental condition, there should be a spread of RTs, even with just one participant in one condition. By reducing RTs to a single value – the mean – one commits multiple mistakes on a statistical and analytical level: the mean cannot accurately represent an entire distribution, especially as the mean assumes normal distribution which RT is not (see also Chapter 2.6.3 *Outlier correction*), and if the RT was reduced so simply and inaccurately into the mean, the resulting vagueness of mean RT could well lead to the selection of a model that incorrectly describes the neurological process at hand (Palmer, Horowitz, Torralba, & Wolfe, 2011). It is therefore important to work on the level of RT distributions (Otto, 2019; Ratcliff, 1979). In this project, RSEs were calculated from RT distributions, using a geometric technique. The task of calculating the RSE was greatly aided by the RSE-box, a Matlab toolbox of custom scripts created by Otto (2019).

To understand what the geometric technique of RSE quantification is, let us begin with a visualisation using the general sensory conditions of auditory, visual, and audio-visual. For each condition let there be a vector of RTs. Each vector of RTs is put through the function *getCP* (Otto, 2019) to obtain a vector of cumulative probabilities, which is then plotted against its RT. Thus, there are three cumulative probability distributions, one each for auditory, visual and audio-visual (see Figure 2.9). The RSE is the shaded area between the cumulative probability distribution of the audio-visual condition, and the cumulative probability distribution of the fastest unisensory component (Figure 2.9, shading performed by the *fillArea* (Otto, 2019) Matlab function). The cumulative probability distribution of the fastest unisensory component is also known as Grice's bound (Grice, Canham, & Gwynne, 1984; Otto, 2019). It can be seen that with this geometric technique, the entire range of RTs is used to calculate the RSE, making it a superior quantification than comparing the mean RTs.

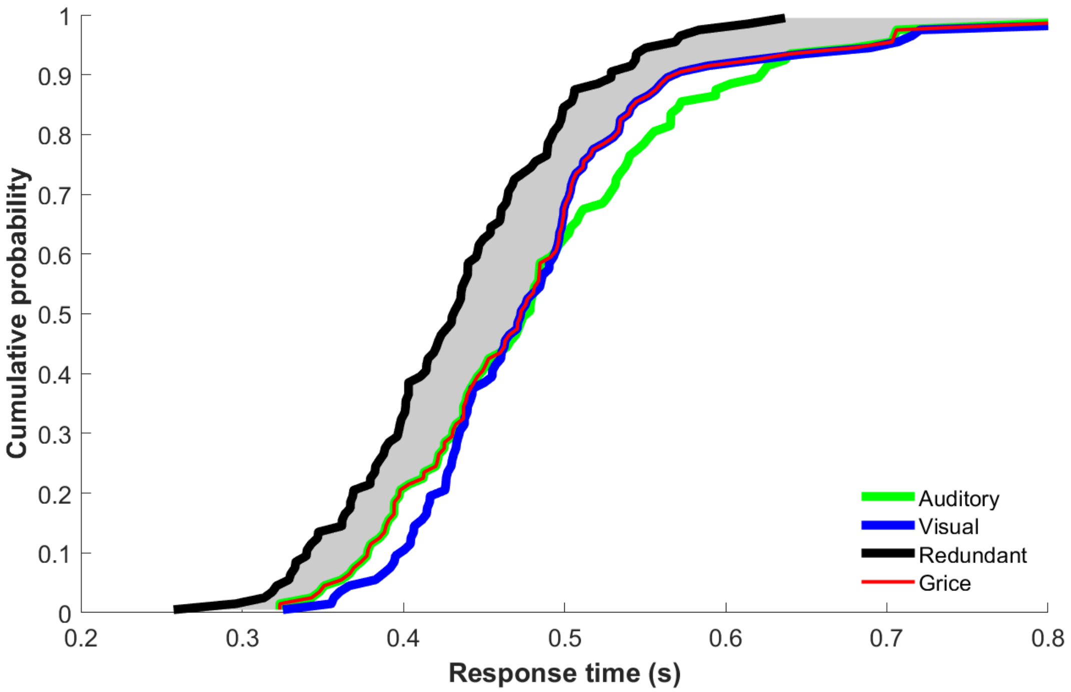


Figure 2.9. Visualisation of the geometric technique for RSE quantification. Cumulative probability distributions are drawn for the auditory (green), visual (blue) and redundant (black) conditions. The cumulative probability distribution of the overall fastest unisensory condition is Grice’s bound (red), which is the lower bound of performance under probability summation using positive processing correlations between the unisensory components (Otto, 2019). The RSE is the shaded area between Grice’s bound and the cumulative probability distribution of the redundant condition.

In practice, the geometric quantification of RSE does not require the drawing of distributions and finding the area between two curves. Instead, the geometric RSE obtained by calculations, using several mathematical shortcuts (Otto, 2019). The prerequisite of these mathematical shortcuts in RSE calculations is an equal number of trials for each condition (Otto, 2019). In the data set, there are 100 signal trials per condition per participant (see Chapter 2.5.3). However, the participant may make occasional mistakes, or some trials were removed in outlier correction. Thus, as a fictional example, there may be 95 valid RTs in the auditory condition, 98 valid RTs in the visual condition, and 99 valid RTs in the redundant condition. To equalise the number of trials, each vector of RTs was put through the function *sampleDown* (Otto, 2019), which uses linear interpolation to produce 50 RT quantiles from the original RT vector. With an equal number of trials between conditions, it is possible to use two calculation shortcuts. First, the calculation of Grice’s bound, which is the lower bound of performance predicted by probability summation assuming positive audio-visual processing correlations (Otto, 2019). With an equal 50 RT quantiles between conditions, Grice’s bound is now simply the minimum value between the auditory and visual vector of RTs at each quantile, the calculation automated in the function *getGrice* (Otto, 2019). The second calculation shortcut is the RSE quantification itself. With equal 50 RT quantiles between conditions, the shaded area representing the RSE (Figure 2.9) can be approximated by taking the difference between Grice’s bound and the audio-visual distribution at each quantile, summing all 50 difference values, and finally dividing the sum by the number of quantiles (Otto, 2019). The entire process of geometrically quantifying the RSE was automated in the function *getGain* (Otto, 2019): one only inputs a matrix containing a down-sampled RT vector for each unisensory component, and a down-sampled RT vector for the redundant condition.

On the level of an entire experiment, RSE quantification using *getGain* (Otto, 2019) was performed per redundant condition per participant. There were four redundant conditions (ALVL, ALVR, ARVL, ARVR) and four unisensory conditions (AL, AR, VL, VR). Thus, to obtain the RSE for ALVL, the input for *getGain* (Otto, 2019) is the down-sampled RT vectors of AL, VL and ALVL. For ALVR, it is the down-sampled RT vectors of AL, VR and ALVR. The same logic applies to ARVL and ARVR. The process was performed for each of the four redundant conditions, per participant. In the end, there were four RSE values (for ALVL, ALVR, ARVL, ARVR) for each of the 20 participants in an experiment. To obtain group-average RSE values, in each redundant condition, the mean of the 20 individual RSEs was taken.

2.6.6 Statistical tests

The main statistical test was the within-subjects ANOVA, which was often in a 2x2 configuration comprising the factors auditory motion direction (receding, looming) and visual motion direction (receding, looming). From this 2x2 ANOVA, it can be determined if a measure, e.g., RSE size, is driven by either unisensory main effect of motion direction. If there is a statistically significant interaction between these two factors, then this may suggest some form of audio-visual interaction, which may corroborate with the Cappe et al. (2009) suggestion of selective integration for audio-visual looming signals. The Greenhouse-Geisser correction was applied if the assumption of sphericity was violated. The alpha level was 0.05 for all statistical tests. The software for statistical analyses was SPSS (IBM).

Chapter 3: Answering knowledge gap 1: does the use of more realistic auditory and visual motion-in-depth stimuli produce stronger multisensory looming effects?

3.1 Experiment 1: Replication

As explained in Chapter 1.2, there is the unanswered question of whether there is a multisensory looming bias in processing redundant audio-visual looming signals. According to Cappe et al. (2009), there is such a multisensory looming bias, in the form of 'selective integration' where there is an especially fast multisensory process only towards audio-visual looming signals. Hence, as a starting point to this project, the basic experimental paradigm of Cappe et al. (2009) was replicated, to see if the basic results of Cappe et al. (2009) could be reproduced, and evaluate the claim of selective integration. Second, the replication gives me a handle on how exactly the experimentation of Cappe et al. (2009) works and whether there is more to the data than what is communicated in Cappe et al. (2009). Third, the replication serves as a baseline, a point of comparison for Experiment 2 which explores if the usage of more realistic stimuli leads to stronger multisensory effects – the first knowledge gap (Chapter 3.2).

3.1.1 Methods for Experiment 1

Participants

20 participants were recruited using the recruitment method described in Chapter 2.1. Several of these participants also participated in Experiment 2 (Chapter 3.2). Regrettably, further demographic details about the participants have been permanently locked and are unrecoverable.

Apparatus, Stimuli, Procedures and Data analysis are all as described in the corresponding sections of Chapter 2.

3.1.2 Results

Accuracy Performance

First, the accuracy performance was examined. In the bigger picture of this project, the aim was to apply computational modelling as a technique for understanding the multisensory processing of audio-visual looming signals. Owing to the simple, salient nature of the stimuli, accuracy performance should be at ceiling level, and is assumed to be as such in the models used for computational modelling (see also Chapter 2.6.4, Chapter 5). Second, the accuracy performance in this replication of Cappe et al. (2009) was of interest because Cappe et al. (2009) gave an accuracy statistic of 88% ±1.1% for their data, which seems low for the detection of salient stimuli, and it is not clear how their statistic was obtained. It was therefore of interest to examine the accuracy performance more closely in my data, and see if there are actually problematic trends already on the level of accuracy, which might also be the case but not mentioned in Cappe et al. (2009). First, an analysis on the hit rates, then an analysis of false alarm rates.

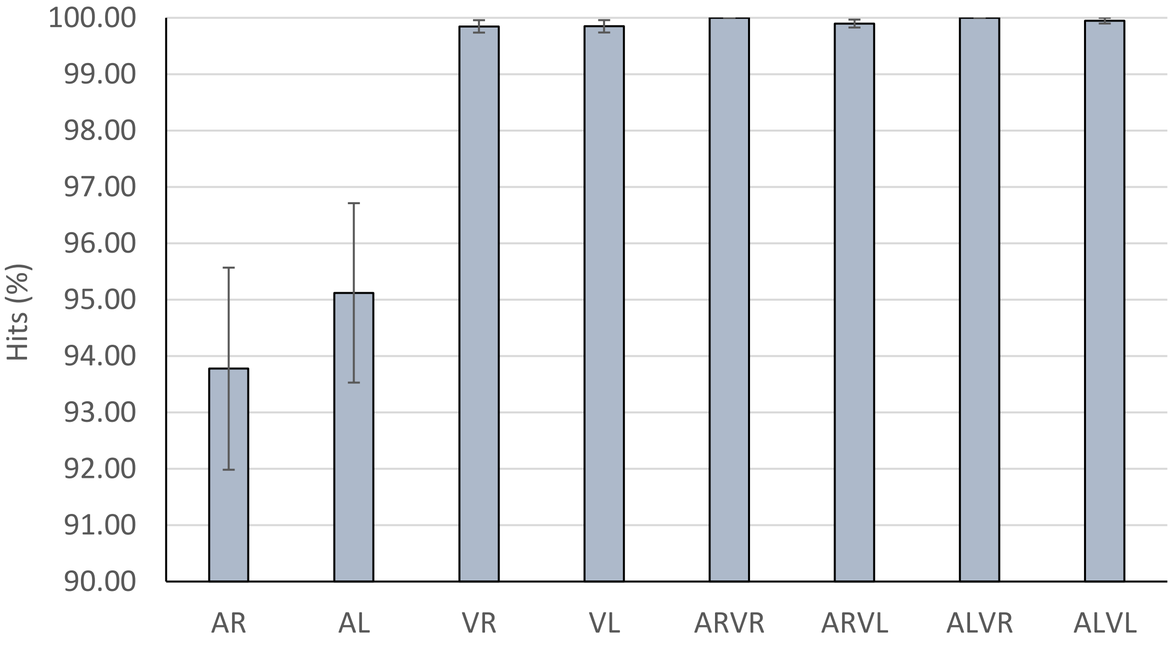


Figure 3.1.1. Hit rates (%) for each condition of the replication experiment, participant-averaged (SEM error bars). The hit rates of the auditory conditions were uniquely lower than all other conditions. However, note that in all conditions, the hit rates are higher than the accuracy statistic reported by Cappe et al. (2009).

On first appearances, the hit rates of the two unisensory auditory conditions (auditory receding, AR and auditory looming, AL) were uniquely poor compared to all the other conditions (Figure 3.1.1). This impression was backed up by statistical analyses. Using multiple paired-samples t-tests, the hit rate of AR (participant average: 93.78% ±1.79%) was compared against the hit rates of the non-auditory conditions, pair by pair. All the paired-samples t-tests against the AR hit rate were significant ($p<0.05$), meaning that the hit rate in the AR condition was statistically lower than the non-auditory conditions (which were assumed to be ceiling level). Using another set of multiple paired-samples t-tests, the hit rate of AL (participant average: 95.12% ±1.59%) was compared against the hit rates of the non-auditory conditions. All the paired-samples t-tests against the AL hit rate were significant ($p<0.05$), meaning that the hit rate in the AL condition was also statistically lower than the non-auditory conditions.

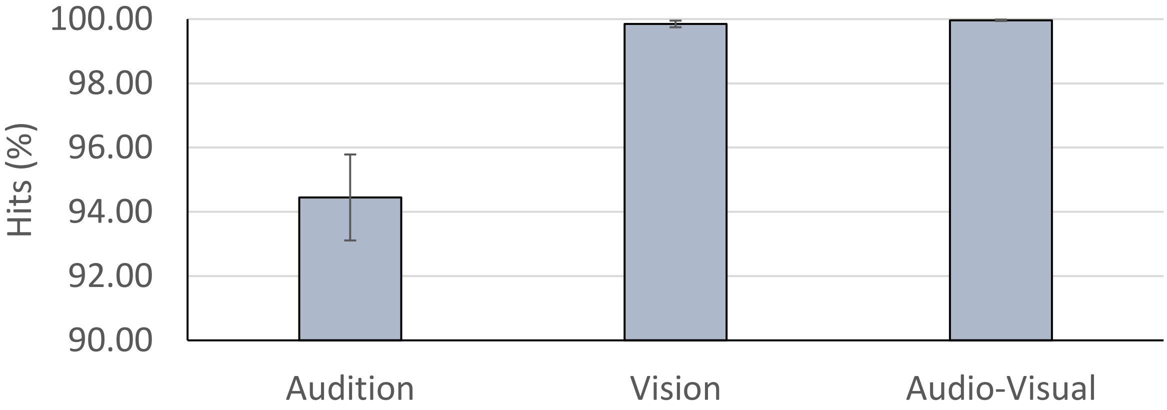


Figure 3.1.2. Hit rates by sensory modality, motion directions averaged (SEM error bars).

Taking t-test results to mean that the auditory conditions had low hit rates, the hit rates were further analysed on the level of sensory modality (motion direction averaged within sensory modality, see Figure 3.1.2). A one-way ANOVA based on modality (audition, vision, audio-visual) was performed, and there was a significant effect of modality on the hit rates ($F_{(1.009, 19.178)} = 16.920$, $p<0.001$, $\eta_p^2 = 0.471$, Greenhouse-Geisser correction applied). Pairwise comparisons between the modalities showed that the auditory hit rate was significantly lower than the visual hit rate (mean difference 5.400% ± 1.311%, $p=0.002$) and the audio-visual hit rate (mean difference: 5.514% ±1.338%, $p=0.002$). The difference between the visual and audio-visual hit rates was non-significant.

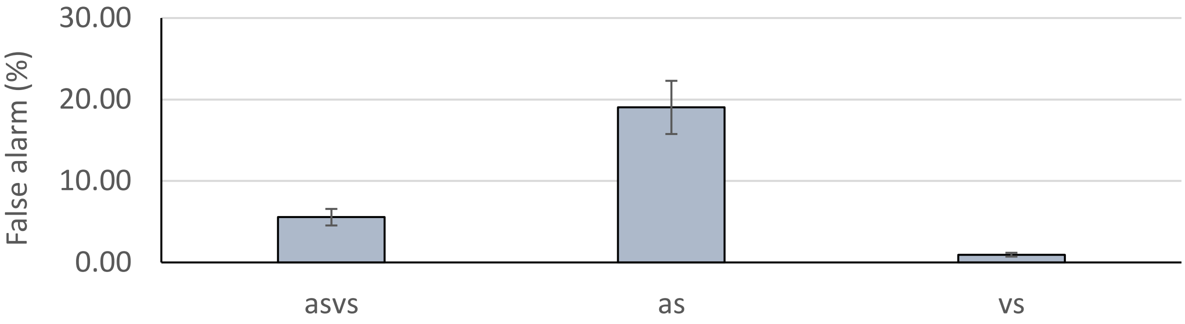


Figure 3.1.3. False alarm rates (%) for each sensory condition of the replication experiment, participant-averaged (SEM error bars). The auditory false alarm rate was significantly higher than in other sensory conditions.

Next, the false alarm rates were examined (see Figure 3.1.3). A one-way ANOVA based on modality (audio-visual catch, auditory catch, visual catch) was performed, revealing a significant effect of modality ($F_{(1.120,\ 21.271)} = 25.541$, $p<0.01$), $\eta_p^2 = 0.573$, Greenhouse-Geisser corrected). Pairwise comparisons showed that the auditory false alarm rate was significantly higher than the audio-visual false alarm rate (mean difference: 13.471% ±3.096%, p=0.001) and the visual false alarm rate (mean difference: 18.075% ±3.216%, p<0.001). The audio-visual false alarm rate was also significantly higher than the visual false alarm rate (mean difference: 4.604% ±0.893%, p<0.001).

Altogether, it appears that the accuracy performance in this replication was uneven, with the worst accuracy in auditory conditions, whether defined by hit rates or false alarm rates. In contrast, the visual and audio-visual conditions were at or near ceiling levels of accuracy. The performance in this replication raises the question if the dataset behind Cappe et al. (2009) also suffers from similar modality-dependent accuracy issues. Interestingly, the overall hit rate for my experiment was 98.56% ±0.91%, or 96.42% ±0.46% if defined as the percentage of hit and correct rejection trials out of all valid trials. Both accuracy statistics for my experiment are higher than that reported by Cappe et al. (2009). Hence, compared to Cappe et al. (2009), my experiment seems to be more conducive to correct responding, and might be attributable to the two modifications: fewer trials, and quieter auditory stimuli.

Response time performance

The next analysis looks into the response times (RT), particularly as Cappe et al. (2009) based their claims on absolute RTs and RT differences between conditions. The claim of selective integration by Cappe et al. (2009) was based on the finding that RTs towards the audio-visual conditions were faster than towards unisensory conditions, and crucially, the congruent audio-visual looming condition had the fastest RTs compared to all other redundant conditions. Hence, for my experimental data, modality effects on RTs, and the RTs between audio-visual conditions are examined.

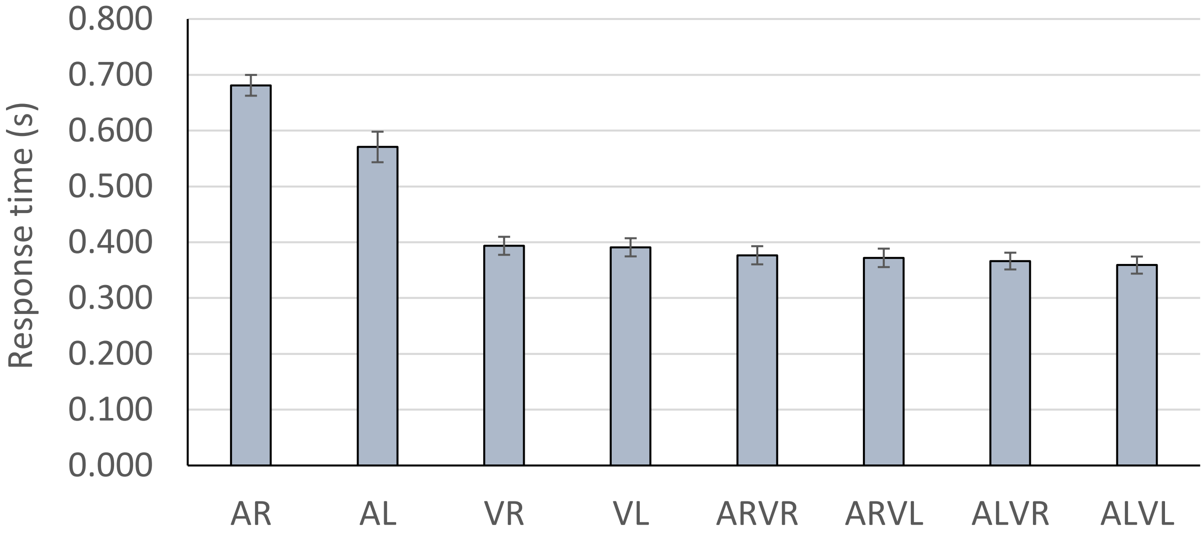


Figure 3.1.4. Participant-averaged RTs for each condition (SEM error bars). Note that the visual and audio-visual conditions have average RTs under 0.4 seconds, which is faster than the RTs obtained by Cappe et al. (2009). However, the RTs in both auditory conditions were uniquely slower than in all other modalities, and appears to be in the same ballpark as the auditory RTs of Cappe et al. (2009; their Figure 2A).

On visual inspection (Figure 3.1.4), the participant-averaged RTs towards visual and audio-visual conditions were below 0.4 seconds. This represents an improvement over the corresponding RTs of Cappe et al. (2009), as their visual RTs were almost 0.5 seconds, and audio-visual RTs in excess of 0.4 seconds, on average. However, it is again the auditory conditions that are problematic. RTs in auditory looming (AL) conditions were approaching 0.6 seconds, while RTs in auditory receding (AR) conditions were even approaching 0.7 seconds (Figure 3.1.4). Yet, these auditory RTs are in the same ballpark as those in Cappe et al. (2009; their Figure 2A).

To check for the unisensory looming biases, paired-samples t-tests were performed on the RTs, comparing looming against receding, for each modality. In audition, RTs towards AL were significantly faster than RTs towards AR ($t_{(19)}$ = 7.301, $p<0.001$, mean difference: 0.110 s ± 0.015 s). Curiously, in vision, RTs towards VL were statistically the same as the RTs towards VR (p = 0.479).

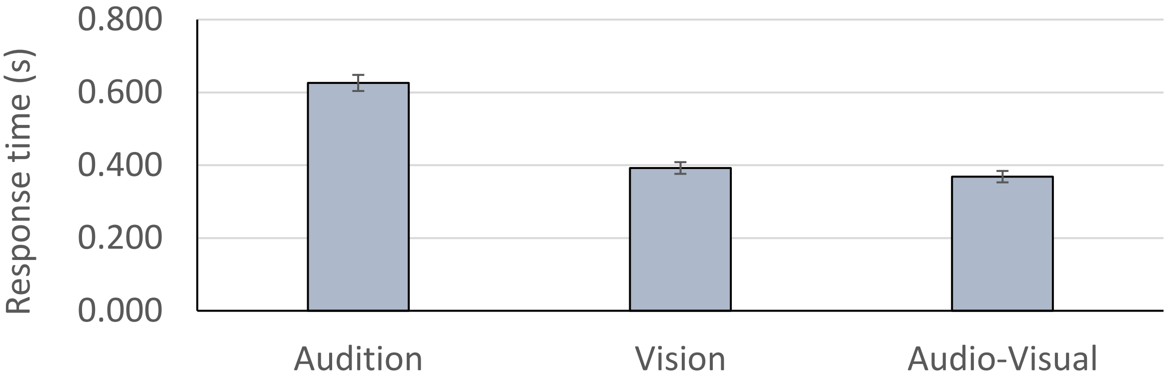


Figure 3.1.5. RTs per sensory modality, calculated from averaging the motion directions within the respective sensory modalities (SEM error bars). Auditory RTs were particularly slow.

The RTs were then grouped by modality (averaging the motion directions; Figure 3.1.5) to determine if there were modality effects, i.e. audio-visual RTs faster than unisensory RTs. A one-way ANOVA on these modality-RTs revealed a main effect of modality ($F_{(1.074, 20.398)}$ = 339.795, $p<0.001$, η_p^2 = 0.947, Greenhouse-Geisser corrected). Pairwise comparisons showed that the audio-visual RTs were significantly faster than the visual RTs (mean difference: 0.024 s ±0.004 s, $p<0.001$) and the auditory RTs (mean difference: 0.257 s ±0.012 s, $p<0.001$). Visual RTs were significantly faster than auditory RTs (mean difference: 0.234 s ±0.014 s, $p<0.001$). Crucially, audio-visual RTs were on average faster than the auditory and visual RTs, which aligns with the findings by Cappe et al. (2009).

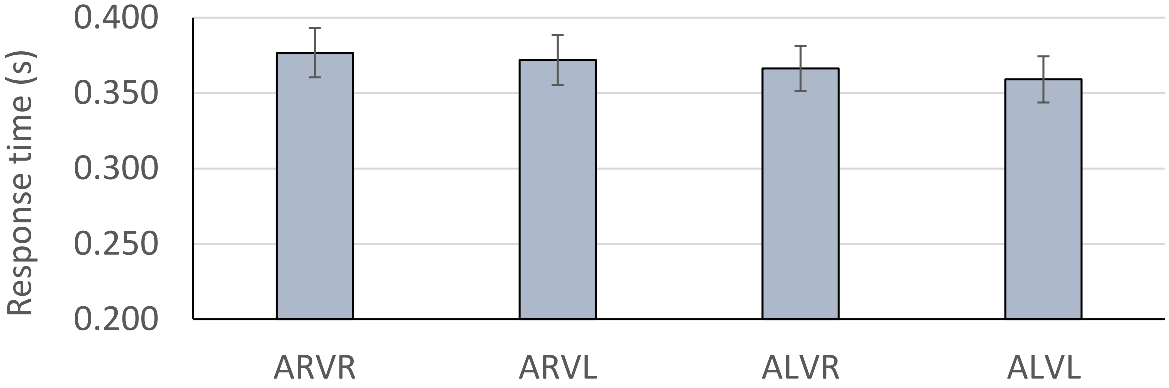


Figure 3.1.6. Participant-averaged RTs for the four audio-visual conditions (SEM error bars).

Next, the audio-visual conditions were examined to see if my data (Figure 3.1.6) concurs with Cappe et al. (2009) in that RTs to ALVL (congruent audio-visual looming) are the fastest of all the audio-visual conditions. Multiple paired-samples t-tests were conducted, comparing each audio-visual condition against ALVL, pair by pair. RTs to ALVL were significantly faster than to ARVR (mean difference: 0.018 s ± 0.015s, $p<0.001$), ARVL (mean difference: 0.013 s ±0.014 s, $p<0.001$) and to ALVR (mean difference: 0.007 s ±0.012s, $p = 0.017$). Thus, it appears that RTs to congruent audio-visual looming conditions are fastest of all the audio-visual conditions, matching the finding by Cappe et al. (2009).

Altogether, the main RT findings of Cappe et al. (2009) seem to have been replicated here (audio-visual RTs faster than unisensory RTs, ALVL RTs fastest of all), despite the two modifications (fewer trials, quieter auditory stimuli) in my experiment. Additionally, my experiment seems to have improved on Cappe et al. (2009), in that the non-auditory RTs here are faster. However, the unisensory visual looming bias of faster RTs towards VL than VR was not found in my dataset. One interpretation for this absence of visual looming bias could be the visual stimuli were not interpreted as three-dimensional motion – danger was not suggested from two-dimensional motion, so fast responses were not called on (cf. behavioural urgency hypothesis; Franconeri & Simons, 2003). Experiment 2 (Chapter 3.2) examines this proposition by introducing more stimulus cues, to investigate if more realistic stimuli might more convincingly represent looming motion, and therefore elicit a stronger looming response.

Redundant signals effect

On the level of RTs, my experiment has successfully replicated the Cappe et al. (2009) findings of faster RTs in audio-visual than unisensory conditions, and fastest RTs in the ALVL condition. However, RTs as such should not be taken as a measure of the multisensory benefit. In fact, it is the speedup of RTs in the redundant condition against its unisensory components that is the correct characterization of the multisensory processing advantage, the redundant signals effect (RSE; see also Chapter 2.6.5). For example, the RT towards ALVL needs to be compared against RTs towards AL and VL, ALVR against AL and VR, and so forth. Thus, the RSEs for each audio-visual condition were calculated from my RT data, using the geometric technique (see Chapter 2.6.5).

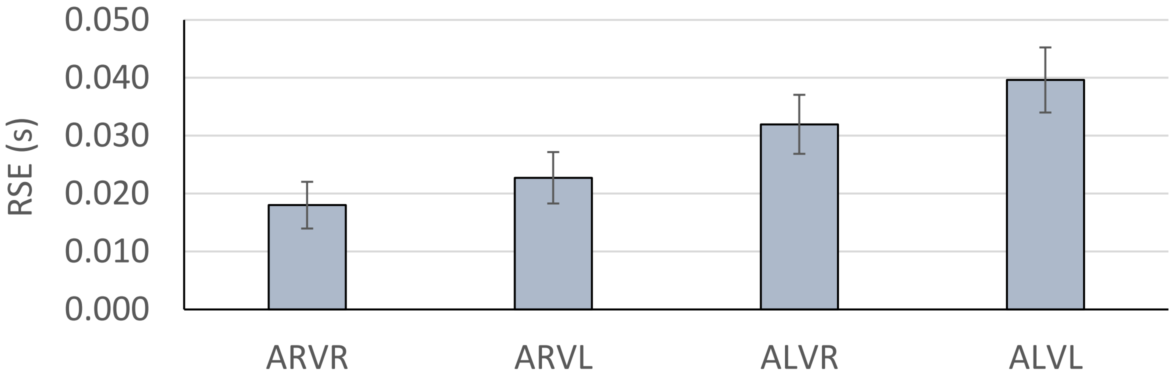


Figure 3.1.7. Participant-averaged RSEs for the four audio-visual conditions (SEM error bars).

A 2x2 ANOVA was performed on the RSEs (see Figure 3.1.7), with the factors auditory motion direction (AR, AL) and visual motion direction (VR, VL). Crucially, there was no significant interaction of the two factors ($F_{(1,19)} = 0.464$, $p = 0.504$, $\eta_p^2 = 0.024$). Instead, there were unisensory main effects, for auditory motion direction ($F_{(1,19)} = 25.776$, $p<0.001$, $\eta_p^2 = 0.576$) and for visual motion direction ($F_{(1,19)} = 7.541$, $p = 0.013$, $\eta_p^2 = 0.284$). Pairwise comparisons within these main effects found that conditions with an AL component had larger RSEs than conditions with an AR component (mean difference: 0.015 s ±0.003 s, p<0.001). Meanwhile, conditions with a VL component had larger RSEs than conditions with a VR component (mean difference: 0.006 s ±0.002 s, p = 0.013). As an extra step, taking that the RSE of ALVL was largest numerically (Figure 3.1.7), a paired-samples t-test was performed to check if the RSE of ALVL was indeed larger than the second largest RSE, that of ALVR. The RSE of ALVL was significantly larger than that of ALVR, but the significance was marginal ($t_{(19)}$ = 2.174, p = 0.043, mean difference: 0.008s ±0.003 s).

In summary, for my replication experiment, I calculated and analyzed the RSE (which was not available in Cappe et al. (2009)), and found the RSE to be present in all four audio-visual conditions, even with incongruent motions. The RSE in ALVL was statistically larger than that of the next largest RSE, that of ALVR, though the significance was marginal. Furthermore, only unisensory factors were important in determining the size of the RSE. Not found was an audio-visual interaction effect specific to looming. Overall, the RSE analysis here does not seem to fully support the idea of special processing just for the ALVL condition.

3.1.3 Experiment 1 summary

At the beginning of this section, three aims were listed for this replication experiment. First, the replication is to see if I could replicate the results of Cappe et al. (2009) and find support for their claim of selective integration. Second, in doing the replication, I obtain first-hand the technique and a dataset, from which I can have a better understanding of how participants respond, and whether there are weaknesses in the methodology shared with Cappe et al. (2009). Third, the replication acts as a baseline for comparison, in preparation for later studies.

To the first aim, the replication was successful in that it appears to have reproduced the data of Cappe et al. (2009), but with my analyses, this data does not seem to support Cappe et al.'s (2009) notion of selective integration processing towards audio-visual looming signals. My replication features two small modifications to the methodology of Cappe et al. (2009): there were fewer trials, and the auditory stimuli were quieter. Importantly, my replication found the same RT results as Cappe et al. (2009): RTs to audio-visual conditions were faster than to unisensory conditions, and RTs in the audio-visual looming condition were the fastest out of all the audio-visual conditions. However, fast RTs per se are not the same as the RT benefit in multisensory conditions, known as the RSE, because the RSE is in fact the RT *speedup* in the multisensory condition versus its unisensory *components*. Here, the size of the RSEs was in fact determined by unisensory factors, namely auditory motion direction and visual motion direction, such that RSEs are larger when there is an AL or VL signal. A motion-specific audio-visual interaction was not found, meaning there was no evidence of a multisensory looming bias, much less a mechanism working by selective integration (Cappe et al., 2009). Altogether, the data had been replicated, but I found no support for the claims of mechanism made by Cappe et al. (2009).

To the second aim, this replication was also successful in gaining an understanding of any intricacies in the experiment data, and perhaps also present in Cappe et al. (2009). First, in this replication, the accuracy performance in auditory conditions were uniquely poor, with low hit rates and high false alarm rates, compared to all other non-auditory conditions. Given that my replication had quite successfully replicated the main findings and general patterns of results in Cappe et al. (2009), it is probable that the data of Cappe et al. (2009) also had these accuracy issues in auditory conditions. Cappe et al. (2009) only communicated an accuracy performance of 88% ±1.1%. Note that in my replication, the overall hit rate was almost 99%, and the overall correct rate (hits and correct rejections) reached 96%, even with the accuracy falloff in auditory conditions. The issues with the auditory conditions were not only in accuracy, but also in RT performance. The auditory conditions had notably slower RTs than the non-auditory conditions, which is also true of the auditory RTs in Cappe et al. (2009). However, my non-auditory RTs were actually faster than those found by Cappe et al. (2009), a result which could be due to the modifications of my replication. The final issue uncovered was the unisensory visual signals seem not to have elicited a visual looming bias: the RTs towards VL were not statistically faster than the RTs towards VR. The equivalence of RTs between VL and VR conditions suggests that participants were only interpreting the motions as harmless two-dimensional size changes, rather than three-dimensional looming and receding which differ in implied threat. In all, my replication revealed a critical weakness of the auditory condition, and the visual stimuli may not be interpreted genuinely as looming. These revelations suggest issues with the auditory and visual stimuli, which are addressed in Experiment 2.

To the third aim, my experiment successfully replicated Cappe et al. (2009), in methodology and overall data, thus there was a basis for comparison in the next experiments. Arguably, my experiment improved on Cappe et al. (2009), producing faster non-auditory RTs and higher accuracy overall, which might be attributable to the two modifications: fewer trials, and quieter auditory stimuli. To speculate, the modifications made my experiment quicker to complete, hence was less taxing on the participant than Cappe et al.'s (2009), so performance was maintainable throughout.

3.2 Experiment 2: Stimulus realism

Real-world looming signals have multiple cues, but typically in looming studies, only a single cue is used, such as an intensity increase for audition, and a visual size change for vision. Experiment 1 used a single cue each for the auditory and visual modalities, and found notably poor performance towards the auditory intensity-change signals, both in terms of accuracy and RT. Moreover, Experiment 1 did not find faster RTs for VL than VR, raising doubts about the participants' interpretations of the supposedly looming and receding signals. Hence, if the stimuli were to be more realistic, by featuring more auditory and visual cues to motion-in-depth, would the aforementioned issues with Experiment 1 be addressed, and furthermore lead to more optimal responses that better answers the question of a multisensory processing bias in favour of looming? In this experiment, further motion-in-depth cues were added onto the basic auditory and visual cues from Experiment 1, to test if stimulus realism is an important factor for looming research.

3.2.1 Methods for Experiment 2

Participants

There were 20 participants in Experiment 2, made up of newly recruited participants (see Chapter 2.1 for recruitment method), and also several participants who had previously participated Experiment 1. Unfortunately, further demographic details about this participant sample are not available as the records have been permanently locked and are unrecoverable.

Stimuli

In Experiment 1, the auditory stimulus was a 1000Hz tone that only changed in intensity to simulate a looming (40 dB SPL to 60 dB SPL) or receding (40 dB SPL to 20 dB SPL) sound source. In this experiment, to increase stimulus realism, another looming cue was added to the basic intensity change: frequency shift. The frequency shift was inspired by the Doppler effect, whereby a stationary listener would perceive a frequency change as the sound source looms or recedes. Thus, to simulate looming, the tone started at 1000Hz and 40 dB SPL, and linearly increased to 1100Hz and 60 dB SPL over the course of 0.5 seconds. To simulate receding, the tone started at 1000Hz and 40 dB SPL, and linearly decreased to 900Hz and 20 dB SPL in 0.5 seconds. Note, these frequency changes were more about increasing the impression of motion-in-depth, rather than accurate representations of the Doppler effect as such. For example, if the sound source is looming at constant speed, then actually, the perceived upwards frequency shift is constant, not increasing. Rather, as used in this experiment, the frequency increase on the auditory looming signal would be more consistent with a looming sound source speeding up. The acceleration suggested by the frequency changes might be desirable: if urgent responses are because looming is a dangerous signal (e.g., behavioural urgency hypothesis; Franconeri & Simons, 2003), then perhaps looming at increasing speed, as suggested by the frequency increase, might lead to even faster responses. On the flip side, the frequency decrease on the auditory receding signal is consistent with a sound source receding at increasing speed, which perhaps is an even more reassuring signal from a survival standpoint, so less urgent responses might be needed.

For this experiment's visual stimuli, the circular object changing in visual size was retained from Experiment 1 (increase from 7° to 13° to simulate looming, decrease from 7° to 1° to simulate receding). However, in Experiment 1, the stimuli was on a featureless grey background, so the size change could be interpreted as three-dimensional motion-in-depth, or simply taken as a two-dimensional size change. If more of the latter, then the responses from Experiment 1 may not be valid representations of the responses to looming signals. Hence, for this experiment on stimulus realism, the featureless grey background was changed to a background with depth cues (Figure 3.2.1). The background was a computer-generated image of a room, with a floor, two side walls, and a back wall, with no ceiling thus revealing the sky. There were three depth cues to this image: the convergent lines from the structure of the room, the brick-like regular repeating patterns on the two side-walls which become finer with simulated depth, and a naturalistic gradient on the sky where the blue becomes lighter closer to the simulated horizon. This background was generated using SketchUp (Trimble, formerly Google at the time of experimentation), a modelling and architecture computer program. The image of the virtual room was taken from the perspective of looking straight down the middle of the room.

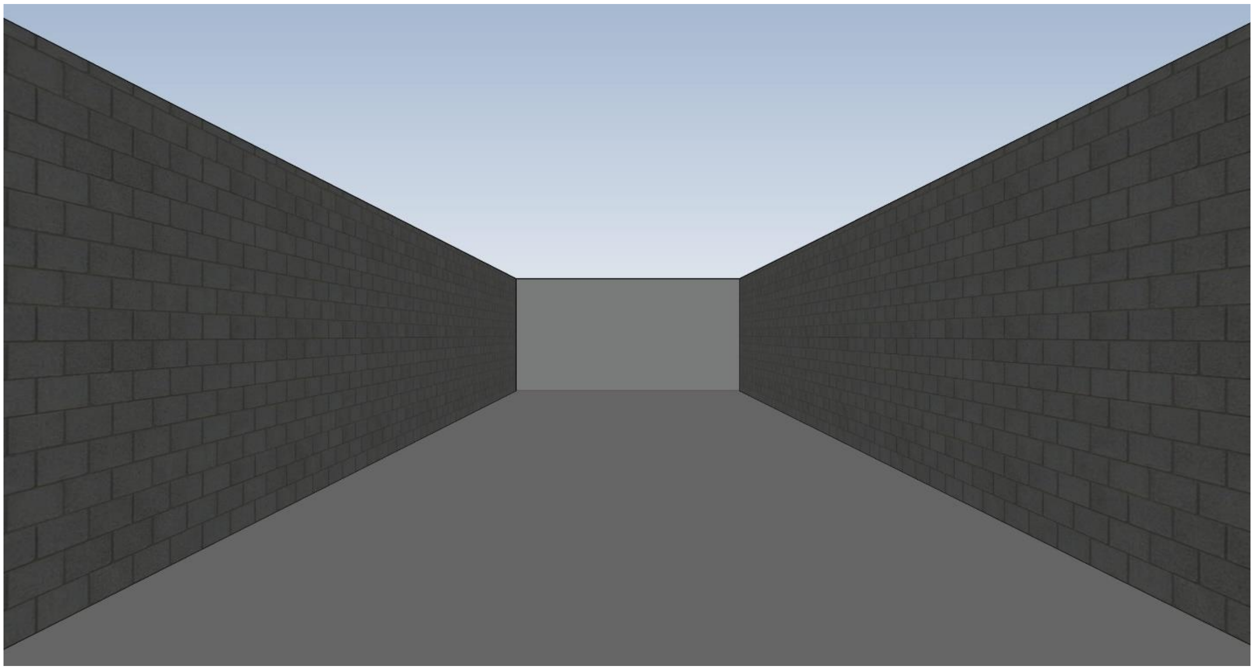


Figure 3.2.1. The computer-generated image of a virtual room which served as the background of this stimulus realism experiment. There are three depth cues in this image background: convergent lines from the structure of the room, repeating and regular brick-like patterns becoming finer with simulated depth, and the colour gradation of the simulated sky. The visual stimulus was again the circular object which changes in size to simulate looming or receding motion, but should be more convincing as three-dimensional motion when presented with the depth cues in this image background.

Apparatus, Procedures and Data analysis are all as described in the corresponding sections of Chapter 2

3.2.2 Results

Accuracy performance

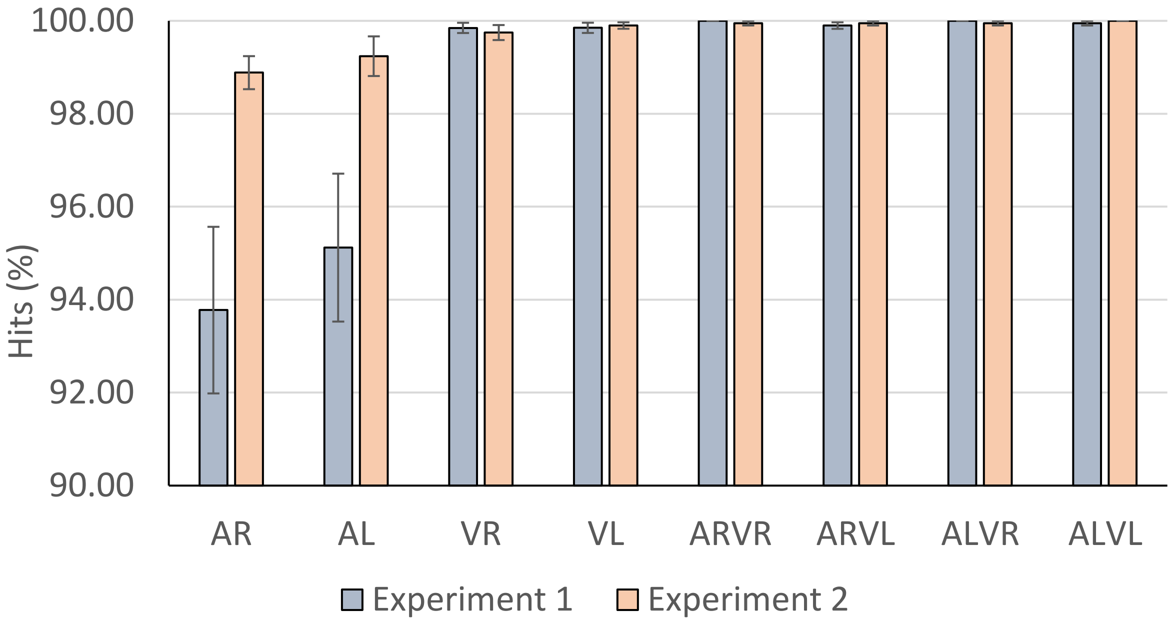


Figure 3.2.2. Participant-averaged hit rates (%, SEM error bars) of the eight conditions, for the replication (Experiment 1, grey bars) and for the current experiment on stimulus realism (Experiment 2, orange bars). More realistic stimuli have improved the hit rate in auditory conditions, and is now near the hit rates of the non-auditory conditions.

Given that the auditory hit rates found in Experiment 1 were low (Chapter 3.1), and the importance of ceiling hit rates for modelling (Chapter 5), the first focus of this analysis was whether stimulus realism has improved the auditory hit rates. An independent-samples T-test was performed, comparing between Experiment 1 (replication, see Figure 3.2.2, grey bars) and Experiment 2 (stimulus realism experiment, see Figure 3.2.2, orange bars), the AR and AL hit rates. AR hit rates with realistic stimuli were significantly higher than AR hit rates with basic stimuli ($t_{(20.490)}$ = 2.796, p = 0.011, equal variances not assumed; mean difference: 5.109% ±1.827%). AL hit rates with realistic stimuli were also significantly higher than AL hit rates with basic stimuli ($t_{(21.720)}$ = 2.500, p = 0.020, equal variances not assumed; mean difference: 4.119% ±1.647%). The addition of frequency change on top of the intensity change seems to have improved the hit rates in auditory conditions.

Next, taking that the non-auditory conditions had reached ceiling level (see Figure 3.2.2, orange bars for Experiment 2), multiple paired-samples t-tests were performed, comparing AR or AL hit rates against each non-auditory condition. The non-auditory conditions all had significantly higher (all p<0.05) hit rates than the AR hit rate. However, the hit rates of non-auditory conditions were not significantly higher than the AL hit rate (all p>0.05).

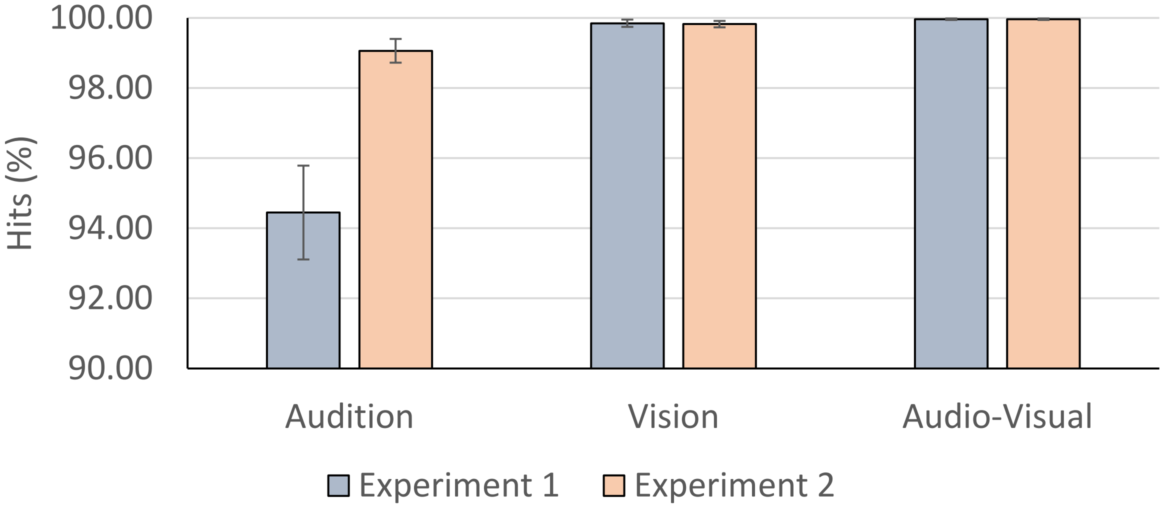


Figure 3.2.3. Hit rates by sensory modality, motion direction averaged (SEM error bars). Grey bars represent Experiment 1 (replication), while orange bars is for Experiment 2 which is the current experiment on stimulus realism. Although the auditory hit rates have been improved by using more realistic stimuli, they are still not at the level of the visual and audio-visual hit rates.

On a sensory modality basis (see Figure 3.2.3, orange bars for Experiment 2), to see if auditory hit rates varied with modality, a one-way ANOVA was performed on the hit rates (motion directions averaged in each sensory modality). The ANOVA revealed a significant main effect of modality ($F_{(1.047,\ 19.901)} = 7.512$, $p = 0.012$, $\eta_p^2 = 0.283$). Pairwise comparisons between the modalities revealed that auditory hit rates were significantly lower than visual hit rates (mean difference: 0.761% ±0.271%, p = 0.034) and audio-visual hit rates (mean difference: 0.900% ±0.328%, p = 0.038). No statistical significance was found in the difference between the visual and audio-visual hit rates (p = 0.255).

Next, a mixed-factors ANOVA was performed on the hit rates (modality as the within-subjects factor, experiment as the between-subjects factor), to compare hit rates between the replication (Experiment 1, Figure 3.2.3, grey bars) and the stimulus realism experiment (Experiment 2, Figure 3.2.3, orange bars). A significant interaction between modality and experiment was found ($F_{(1.013, 38.507)} = 11.545$, $p = 0.002$, $\eta_p^2 = 0.233$). Pairwise comparisons in this ANOVA revealed that the significant interaction was only driven by the auditory modality, as the auditory hit rate in Experiment 2 was significantly higher than in Experiment 1 (mean difference: 4.615% ±1.381%, p = 0.002). Hit rates in other modalities were statistically the same in both experiments (p>0.05).

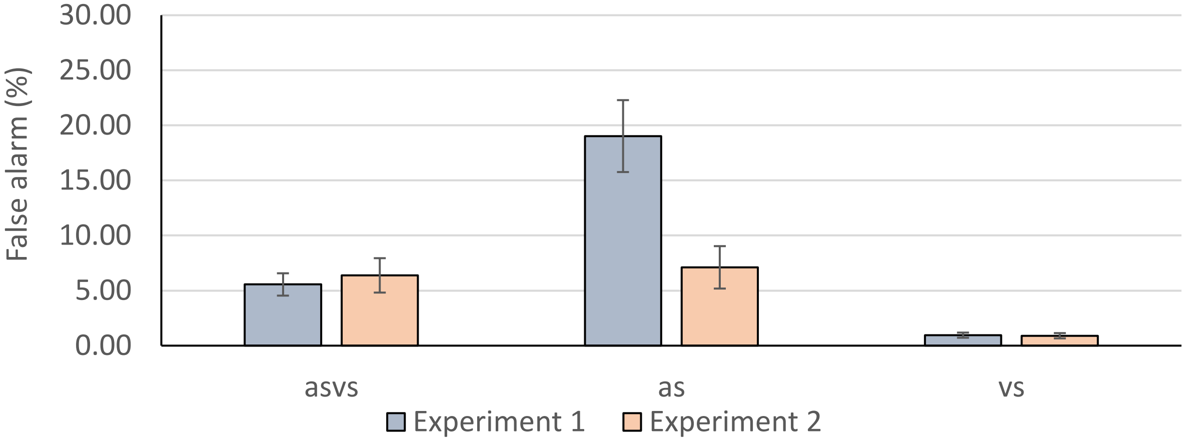


Figure 3.2.4. False alarm rates per sensory modality, participant-averaged (SEM error bars). Grey bars represent Experiment 1, which was the replication experiment (Chapter 3.1). Orange bars represent Experiment 2, the current experiment on stimulus realism. The auditory false alarm rate has been brought down by using more realistic stimuli, and is on par with the hit rates in the audio-visual modality.

For false alarm rates (Figure 3.2.4, orange bars for Experiment 2), a one-way ANOVA revealed a main effect of modality ($F_{(1.162, 22.078)} = 10.743$, $p = 0.002$, $\eta_p^2 = 0.361$). Pairwise comparisons showed that the visual false alarm rate was significantly lower than the audio-visual false alarm rate (mean difference: 5.480% ±1.541%, p = 0.006) and the auditory false alarm rate (mean difference: 6.208% ± 1.895%, p = 0.012). Audio-visual and auditory false alarm rates were not significantly different to each other (p = 0.654).

Comparing the false alarm rates between Experiment 1 (replication, Figure 3.2.4 grey bars) and Experiment 2 (stimulus realism, Figure 3.2.4 orange bars), a mixed-factors ANOVA was performed (modality as the within-subjects factor, experiment as the between-subjects factor). A significant interaction between modality and experiment was found ($F_{(1.349, 51.243)} = 11.192$,

$p<0.001$, $\eta_p^2 = 0.228$), which pairwise comparisons revealed to be driven only by the auditory condition, where the auditory false alarm rate is significantly lower in Experiment 2 than in Experiment 1 (mean difference: 11.918% ± 3.791%, p = 0.003). There was no significant difference in false alarm rates between the experiments on the other modalities ($p>0.05$).

Response time performance

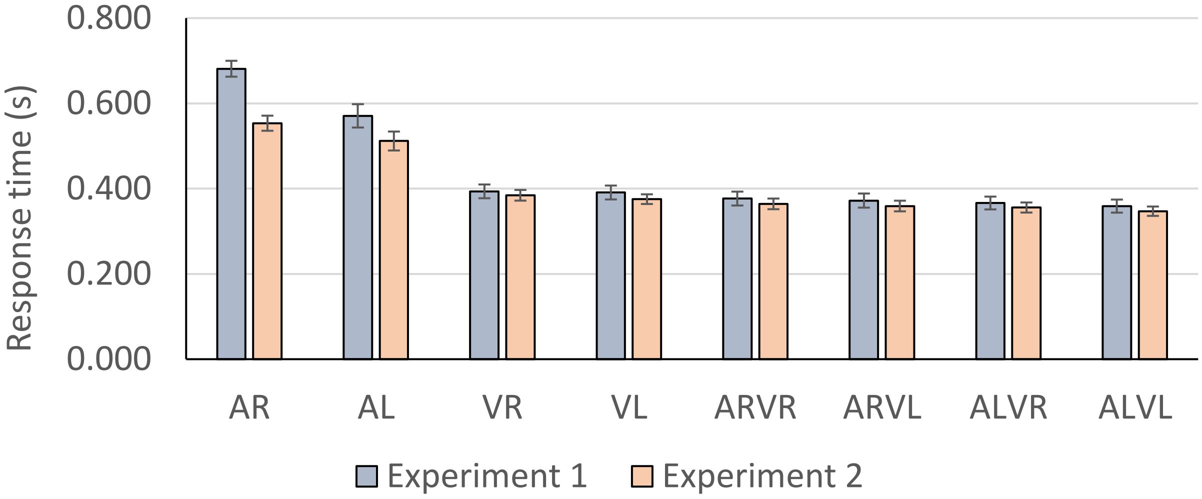


Figure 3.2.5. Participant-averaged RTs across all conditions (SEM error bars). Grey bars represent the RTs of Experiment 1 (Chapter 3.1), orange bars represent the RTs of the current stimulus realism experiment.

Visually (see Figure 3.2.5), RTs appear to be faster with the more realistic stimuli (Experiment 2, orange bars), than with the basic stimuli of the replication (Experiment 1, grey bars), across all conditions. The biggest improvement is in the auditory conditions. To check if these RT improvements were statistically significant, an independent samples t-test was performed, comparing between experiments the RTs in each of the eight conditions. Realistic AR signals (Figure 3.2.5, orange bar) produced RTs that were significantly faster than that of the basic AR signals (Figure 3.2.5, grey bar; $t_{(38)} = 4.971$, $p<0.001$, mean difference: 0.128 s ± 0.026 s). None of the other conditions had RTs that were significantly different between the experiments ($p>0.05$). Thus, except for the AR condition, which had the slowest RTs in Experiment 1, there seems to be a null effect of including more cues to motion-in-depth, at least in terms of RTs.

To check for unisensory looming biases, paired-samples t-tests were performed, comparing the RTs towards looming and receding motions, for each modality. In audition, RTs towards AL were significantly faster than RTs towards AR ($t_{(19)}$ = 4.443, p<0.001, mean difference: 0.042 s ±0.009 s). In vision, RTs towards VL were significantly faster than RTs towards VR ($t_{(19)}$ = 2.741, p = 0.013, mean difference: 0.009 s ±0.003 s). Thus, unisensory looming biases were found in both sensory modalities.

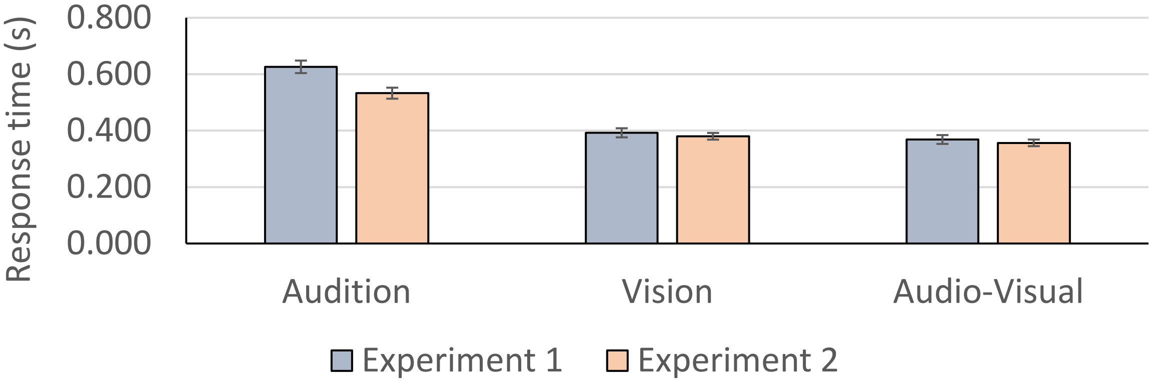


Figure 3.2.6. RTs on the level of sensory modality (motion direction averaged within sensory modality, SEM error bars). Grey bars represent Experiment 1 (replication), orange bars represent Experiment 2, the current experiment on stimulus realism.

Next, RTs of the stimulus realism experiment were analysed on the level of sensory modality (motion directions averaged within sensory modality, see Figure 3.2.6, Experiment 2 in orange bars). A one-way ANOVA on the RTs revealed a main effect of modality ($F_{(1.192, 22.650)}$ = 164.658, p<0.001, η_p^2 = 0.897, Greenhouse-Geisser corrected). Pairwise comparisons between the modalities showed that audio-visual RTs were significantly faster than auditory RTs (mean difference: 0.176 s ± 0.012 s, p<0.001) and visual RTs (mean difference: 0.023 s ± 0.005 s, p<0.001). Auditory RTs were significantly slower than visual RTs (mean difference: 0.153 s ± 0.013 s, p<0.001). Altogether, audio-visual conditions had the fastest RTs, compared to the unisensory conditions. Cappe et al. (2009) also found audio-visual RTs to be fastest out of auditory and visual RTs.

To compare the sensory modality RTs between the replication and stimulus realism experiments (see Figure 3.2.6), a mixed-factors ANOVA was performed, with sensory modality as the within-subjects factor, and experiment as the between-subjects factor. A significant interaction between modality and experiment was found ($F_{(1.133, 43.066)}$ = 19.163, p<0.001, η_p^2 =

0.335, Greenhouse-Geisser corrected). On closer examination, this significant interaction was only driven by the auditory modality. Namely, the auditory RTs in Experiment 2 (Figure 3.2.6, orange bar) was significantly faster than the auditory RTs in Experiment 1 (Figure 3.2.6, grey bar, mean difference: 0.093 s ± 0.030 s, p = 0.003). Thus, the additional cues to motion-in-depth seem to only have had an effect in the auditory modality.

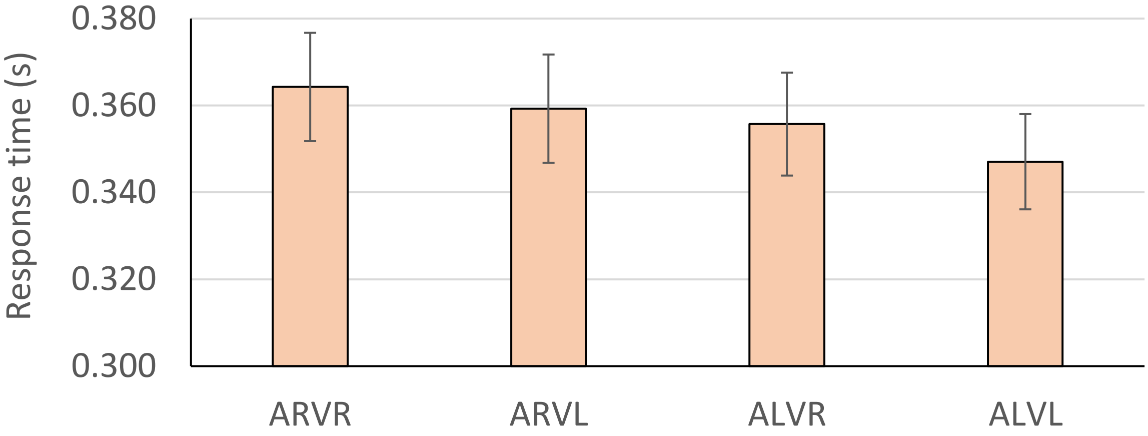


Figure 3.2.7. Participant-averaged RTs in the audio-visual conditions (SEM error bars), Experiment 2.

Finally, the RTs in the audio-visual conditions were examined (Figure 3.2.7). A paired-samples t-test was used between ALVL, and the condition with the next fastest RT, ALVR. RTs in the ALVL condition were significantly faster than RTs in the ALVR condition ($t_{(19)}$ = 2.267, p = 0.035, mean difference: 0.009 s ±0.004 s). Cappe et al. (2009) also found that RTs in the ALVL condition were fastest of all the audio-visual conditions.

A 2x2 ANOVA was performed on the RTs in audio-visual conditions, with factors auditory motion direction (AR, AL) and visual motion direction (VR, VL). Crucially, there was no interaction between auditory motion direction and visual motion direction ($F_{(1,19)}$ = 0.914, p = 0.351, η_p^2 = 0.046). Instead, there was only a main effect of auditory motion direction ($F_{(1,19)}$ = 16.060, p<0.001, η_p^2 = 0.458). The pairwise comparison showed that conditions with AL had faster RTs than conditions with AR (mean difference: 0.010 s ± 0.003 s, p<0.001). There was no main effect of visual motion direction (p = 0.056).

Redundant signals effect

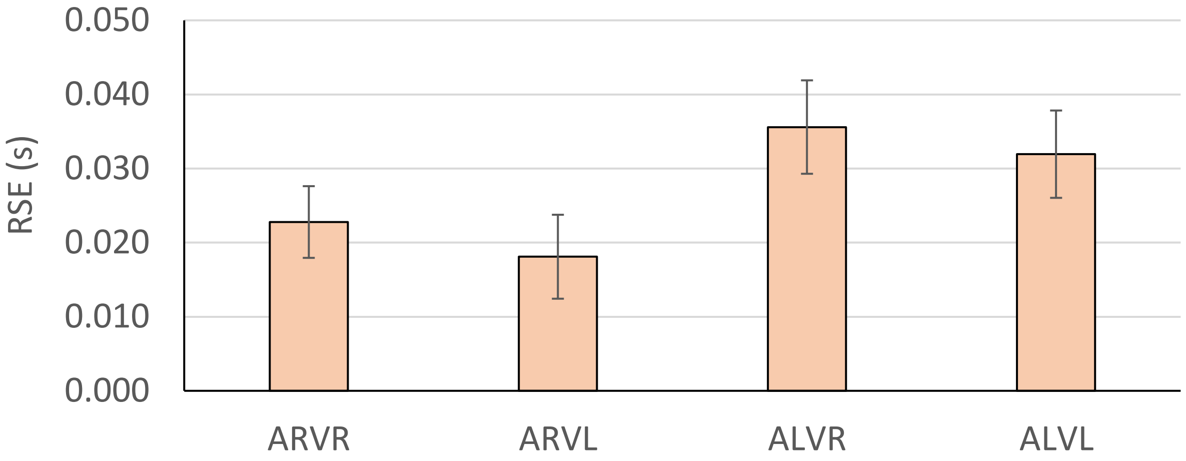


Figure 3.2.8. Participant-averaged RSEs of the four audio-visual conditions (SEM error bars), Experiment 2.

To examine the RSEs (see Figure 3.2.8), a 2x2 ANOVA was conducted, with factors auditory motion direction (AR, AL) and visual motion direction (VR, VL). Crucially, no interaction was found between the auditory and visual motion directions ($F_{(1,19)} = 0.795$, $p = 0.795$, $\eta_p^2 = 0.004$). Only the main effect of auditory motion direction was significant ($F_{(1,19)} = 19.454$, $p<0.001$, $\eta_p^2 = 0.506$). A pairwise comparison within this auditory factor revealed that audio-visual conditions with an AL component have significantly larger RSEs than audio-visual conditions with an AR component (mean difference: 0.013 s ±0.003 s, $p<0.001$). There was no main effect of visual motion direction ($p = 0.177$). This ANOVA result shows that the size of the RSE was only determined by unisensory auditory factors, specifically larger RSEs when the auditory component suggested looming motion as opposed to receding motion. A motion-specific audio-visual interaction, which would be a sign of Cappe et al.'s (2009) selective integration for ALVL signals, was not found.

Interestingly, the participant-averaged RSE of ALVR was larger than that of ALVL, which goes against the notion of an especially speedy or advantageous multisensory process specific to ALVL signals (e.g., Cappe et al., 2009). However, a paired samples t-test between the RSEs of ALVR and ALVL revealed that statistically, the RSE of ALVR was not larger than that of ALVL ($t_{(19)} = 0.954$, $p = 0.352$, mean difference: 0.004 s ± 0.004 s).

3.2.3 Experiment 2 summary

In summary, the additional auditory and visual motion-in-depth cues has improved response performance, in terms of accuracy and speed, albeit primarily in the auditory conditions. Accuracy first, the additional cue of frequency change elevated the AL and AR hit rates (the AL hit rate was statistical indistinguishable from the ceiling-level non-auditory conditions), while false alarms were reduced and equal to that in audio-visual conditions. Overall, the hit rate was 99.70% ±0.15%, which is not significantly higher than the overall hit rate of Experiment 1 (98.56% ±0.91%, independent-samples t-test p = 0.250; Chapter 3.1). The overall correct (proportion of hits and correct rejections) was 98.05% ±0.43%, which is significantly higher than the overall correct of Experiment 1 (96.42% ±0.46%, independent-samples t-test, $t_{(38)}$ = 2.587, p = 0.014, mean difference: 1.531% ±0.630%). In terms of RTs, the additional motion-in-depth cues seem to have benefits for both auditory and visual performance. Compared to the basic auditory stimuli of Experiment 1, the frequency-intensity auditory stimuli here produced faster RTs. For the visual modality, the addition of depth cues via the image background appears to have aided the impression of visual motion-in-depth, because now a visual looming bias was detected (VL responded to faster than to VR), which was absent in Experiment 1. However, despite again reproducing the Cappe et al. (2009) findings of audio-visual RTs being faster than the unisensory RTs, and ALVL RTs being fastest of all the audio-visual conditions, the RSE analysis again found no support for Cappe et al.'s (2009) claim of selective integration: the RSE was not unique for ALVL, and the RSE size was only determined by a unisensory auditory factor. Note that although the RTs towards ALVL were fastest of all the audio-visual conditions, the RSE towards ALVL were smaller than the RSE of ALVR (albeit this difference did not reach significance). This finding illustrates that fast RTs are not to be automatically equated with a large redundancy gain (the RSE) – a point made when evaluating Cappe et al. (2009) in Chapter 1.2.3.

3.3 Chapter discussion

Starting this project, it was noted that looming studies tended to use basic auditory and visual cues to represent looming and receding – auditory intensity change, visual size change (e.g., Cappe et al., 2009). However, looming is very much rooted in the real world, and in the real world, there are multiple cues to looming and receding, in both audition and vision (e.g., Paquier et al., 2016; Zahorik et al., 2005). Hence, the knowledge gap: does the use of more realistic auditory and visual motion-in-depth stimuli produce stronger multisensory effects? To answer the knowledge gap, in this chapter, two behavioural experiments differing only in stimulus realism were conducted: one a replication of Cappe et al. (2009) using basic stimuli, the other experiment using the same experimental paradigm but with additional cues to motion-in-depth. Thus, in comparing the two experiments, the effect of stimulus complexity can be seen, answering the knowledge gap.

Summing up, the replication was successful on several fronts. First, the main behavioural results from Cappe et al. (2009) were reproduced: RTs in audio-visual conditions were faster than in unisensory conditions, and RTs in ALVL (congruent audio-visual looming) were fastest of all the audio-visual conditions. Second, in performing the replication, it was revealed that the auditory conditions were uniquely poor in accuracy and RT, and it is possible that dataset of Cappe et al. (2009) has a similar issue. Third, the replication made slight modifications (fewer trials, quieter auditory stimuli) which appears to have produced accuracy and RTs above that found by Cappe et al. (2009); a good baseline was established for comparison against Experiment 2 with its more realistic stimuli. Recapping Experiment 2 then, the addition of a Doppler-inspired frequency change on top of the auditory intensity change, and using an image background with visual depth cues, all seem to have produced quite positive improvements in accuracy and RT. The auditory conditions in particular improved in response accuracy and speed (a unique weakness back in Experiment 1 with its auditory intensity-only changes), while the non-auditory conditions retained their high level of performance already found in Experiment 1. Also, the Cappe et al. (2009) findings of faster RTs in audio-visual than unisensory conditions, and fastest RTs in ALVL, were also retained here despite the stimulus changes. Altogether, this first look into stimulus realism seems to show that the additional cues may have (small) benefits in detecting auditory signals.

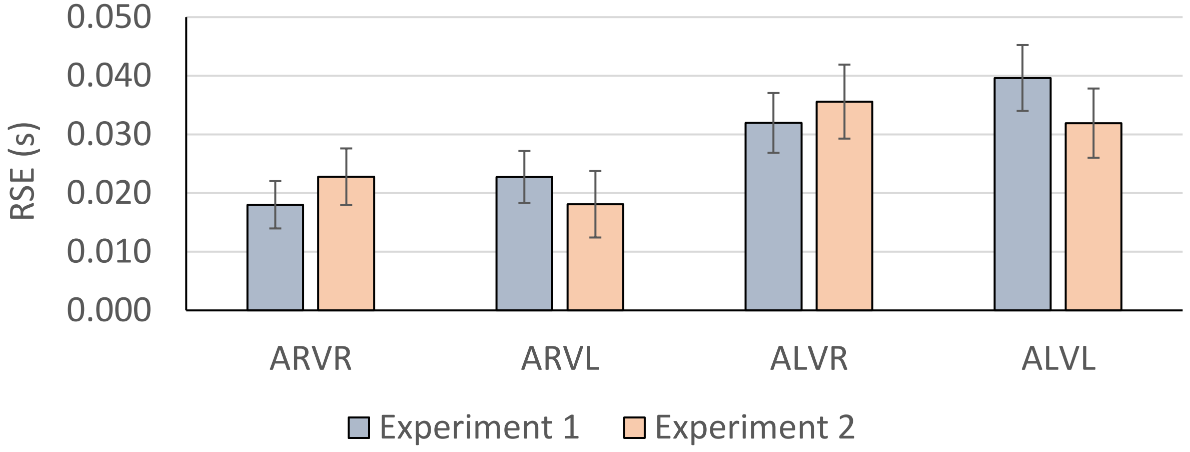


Figure 3.3.1. Comparison of the participant-averaged RSEs (SEM error bars) between the replication experiment using basic stimuli (Experiment 1, grey bars) and the stimulus realism experiment (Experiment 2, orange bars). For each condition, the RSEs were statistically the same between the experiments.

However, a deeper, more critical look into the data seems to paint a different picture about the effectiveness of stimulus realism in this experiment setup. An important measure in this project is the size of the RSE, as a quantity of the multisensory benefit in decision-making, as opposed to receiving only unisensory information. On one level, in both Experiments 1 and 2, the size of the RSEs was only determined by unisensory factors, i.e., large RSEs if the condition included AL not AR. Neither experiment found an audio-visual interaction specific to looming motion: there does not appear to be the special fast processing just for audio-visual looming signals, claimed by Cappe et al. (2009). On another level, the addition of more stimulus cues for realism in Experiment 2 had not increased the size of the RSE, or introduced other RSE-level effects. Experiment 2 RSEs were in fact statistically the same as their counterparts in Experiment 1: an independent-samples t-test comparing the four RSEs between experiments found none of the RSEs were statistically bigger in Experiment 2 than in the Experiment 1 (all four comparisons p>0.05; see Figure 3.3.1, grey bars vs orange bars). Third, with more cues on the stimuli, the RSE was actually larger in the motion-incongruent ALVR condition than in the motion-congruent ALVL condition, albeit the difference did not reach statistical significance (Figure 3.3.1, orange bars). Altogether, no evidence was found for an especially large multisensory benefit just in ALVL conditions, or any audio-visual interactions on the RSE; the addition of more stimulus cues for realism did not find evidence to the contrary.

The apparent insignificance of stimulus realism has been suggested in previous studies (Bach et al., 2009; Rosenblum, Carello, & Pastore, 1987). In a study on auditory looming, participants were biased in judgements and responses towards looming signals: looming signals were judged more unpleasant, change more in loudness, produced a larger skin conductance response, and primed subsequent responses to be faster, compared to the receding counterpart (Bach et al., 2009). However, the above responses were no different between intensity-change looming signals, and complex multi-cue looming signals (involving Doppler effect, and a range of distance-related effects e.g., intensity, filtering, reflection, time delays), apart from a larger skin conductance response with multi-cue looming (Bach et al., 2009). In another study, three auditory looming cues (intensity change, interaural time and level difference, Doppler effect) were compared in a time-to-arrival paradigm (Rosenblum et al., 1987). The basic signal was an ambulance siren, simulating a drive-by (left to right, or vice versa), but on each trial the signal comprised of one to all three auditory looming cues together, and the task was to indicate when the ‘ambulance’ arrived at the participant’s position (Rosenblum et al., 1987). Intensity change was found to be the dominant cue, followed by interaural differences, and lastly the Doppler effect (Rosenblum et al., 1987). As a postulation, intensity change may be the dominant cue to looming because it is most reliable: intensity change is always present in the sound from a looming sound source, whilst interaural differences are irrelevant in head-on looming, and Doppler effect is ineffective if the sound is interrupted (Rosenblum et al., 1987). Altogether, it seems that intensity change is the main cue to auditory motion-in-depth; little is gained from more auditory cues.

Before closing this chapter, an evaluation, in particular the stimulus realism implemented in Experiment 2. Two points of criticism: realism was only implemented through a minimal addition of cues, and the cues were not particularly realistic. Previous studies have explored signal realism with a complex suite of cues (e.g., Bach et al., 2009), not just an addition of auditory frequency change and visual depth cues to the core auditory intensity and visual size changes. There are many auditory and visual cues to depth (e.g., Paquier et al., 2016; Zahorik et al., 2005); Baumgartner et al. (2017) argues in favour of auditory spectral changes instead of intensity change for studying looming, but this cue was not included here. The cues were not particularly real too. The Doppler-inspired frequency change suggests looming at increasing speed, while the linear increase in visual size suggests looming at decreasing speed; the frequency change of 100 Hz was not mapped to the motion suggested by the 20 dB SPL change; and the motion suggested auditorily were not matched to that suggested by the 6° visual size change. Motion is only the depth change, but the absolute distance between the looming object and the observer is also important for the response to looming (see also Bronkhorst & Houtgast, 1999; Zahorik et al., 2005). Past studies suggest that the response to looming is only strong if the looming object comes close to the observer (cf. peripersonal space; e.g., Camponogara, Komeilipoor, & Cesari, 2015; Canzoneri, Magosso, & Serino, 2012; Graziano, Yap, & Gross, 1994; Graziano, Reiss, & Gross, 1999; Serino, Annella, & Avenanti, 2009). To counter, there were methodological limitations to implementing more realistic stimuli: both auditory and visual stimuli were generated live and within Matlab / Psychtoolbox, and some of the easier stimulus changes were auditory intensity/frequency and visual size. More complex stimuli may be taxing for the computer, and could compromise stimulus timing. Importantly, within a laboratory setting, no matter the stimulus complexity, it is not possible to produce perfectly realistic motion-in-depth. The issue is that the auditory and visual apparatus (headphones and computer displays) anchor one's percept of depth: sounds played through headphones seem to originate within the head regardless of the signal, the proprioceptive signal from fixing focus and binocular convergence onto the display plane counters any visual depth signal shown onscreen (Paquier et al., 2016). Also, previous studies have found little difference between single-cue and multi-cue looming (Bach et al., 2009; Rosenblum et al., 1987). Hence a pursuit of absolute stimulus realism may not be most fruitful.

In summary, this exploration into stimulus realism yielded several findings. First, answering the knowledge gap, additional cues on the audio-visual looming stimuli for realism did not produce stronger multisensory looming effects. There was no difference in RSE size using basic or more elaborate audio-visual stimuli. Furthermore, with basic and more elaborate stimuli, the RSE size was only determined by unisensory factors; no looming-specific audio-visual interaction was found, countering the Cappe et al. (2009) suggestion of a special integrative mechanism for processing audio-visual looming signals. Second, additional cues seem to have only aided performance in auditory conditions, which was uniquely poor with basic stimuli. Hence, in the context of a non-effect of stimulus realism, the additional auditory cue (frequency change) may have only served to mark out the different auditory conditions, boosting accuracy and RT, but not via realism per se. Taking the apparent non-effect of stimulus realism in this preliminary exploration, the realism idea is parked for now. The project takes a different direction to address persisting methodological issues, and takes another approach in exploring multisensory looming.

Chapter 4: Answering knowledge gap 2: To what extent can the race architecture be used to explain human decision-making towards redundant audio-visual looming signals?

Cappe et al. (2009) proposed a selective integration mechanism, meaning there was special processing just for audio-visual looming signals (‘selective’), and that this processing is speedy due to a particular type of multisensory processing: integration. However, definitive evidence for selective integration towards audio-visual looming has not been found (see also Chapter 1.2.3). Furthermore, a specific model of selective integration was not provided by Cappe et al. (2009). In Chapter 3, it was shown that an RSE analysis of two behavioural experiments found no special multisensory performance towards audio-visual looming signals. Hence, the question: is multisensory processing ‘selective’ to motion direction, and via an integration process?

Integration is not the only mechanism of multisensory processing, but the alternative, the race mechanism, is often overlooked, arguably on incorrect grounds. In the absence of concrete evidence and specific instantiation of a selective integration mechanism, in this chapter, I explore if race mechanisms can be an alternative framework for understanding the multisensory processing of audio-visual looming signals. First, in Chapter 4.1, an overview of the basics of sensory processing is given, along with the possible mechanisms of multisensory processing. However, the experiments in Chapter 3 had their own problems, and these needed to be solved first before any application of theoretical mechanism. Hence, in Chapters 4.2 and 4.3, two experiments are detailed, as they looked to improve on the data obtained from Experiments 1 and 2 (Chapter 3). Finally, in Chapter 4.4, the foundation of race mechanisms, probability summation, was applied to the best empirical data yet, to explore how well race mechanisms could explain the behavioural performance towards audio-visual motion-in-depth signals.

4.1 A background on mechanisms of multisensory processing

4.1.1 Sensory basics

At the core of perceptual decision-making studies is the question of how sensory information is linked to the motor action in response (Bogacz, 2007; Otto & Mamassian, 2017). On the one hand, sensory information is not complete, because of noise in the external signal, or noise in the internal sensing processes (Bogacz, 2007). On the other hand there is the need to produce fast and accurate responses, to interact well with and survive in the environment (Bogacz, 2007). Hence, in a statistical sense, the optimal linkage between sensing and acting is a three-step process: first, sensory neurons accumulate sensory evidence in favour of a particular motoric outcome, second, cortical neurons combine the accumulated sensory evidence over time to cancel out the noise, and third, the amount of accumulated sensory information is checked against a pre-set decision criterion for producing the motoric action (Bogacz, 2007). The motoric action is produced if the amount of sensory evidence has reached the decision criterion, while the sensory accumulation continues if it is still short of the decision criterion (Bogacz, 2007). In essence, decision-making is characterised as the accumulation of sensory evidence until there is enough for one to confidently produce a particular motoric action.

4.1.2 Multisensory basics

Perceptual decision-making reaches a new level when multiple sensory signals are available (Otto & Mamassian, 2017). The multiplicity and variety of sensory signals makes for a more robust understanding of the environment (Otto & Mamassian, 2017). Recall the approaching car example given at the beginning of this thesis: one can hear the car coming, one can see the car coming, but if one can both hear and see the car coming, then one would most likely make the correct decision of waiting until the car has gone past to cross the road. Notice that this approaching car example is an illustration of redundant signals, because the same motoric outcome is required for the auditory signal, the visual signal, and of course the audio-visual signal; either signal would have been sufficient. The responses towards redundant signals are typically faster than that towards the unisensory signals, in a phenomenon known as the redundant signals effect (RSE; Hershenson, 1962; Kinchla, 1974; Miller, 1982; Todd, 1912), which was also shown in my experiments in Chapter 3. The critical question, then, is how can multisensory effects such as the RSE be explained in the sensory accumulation framework.

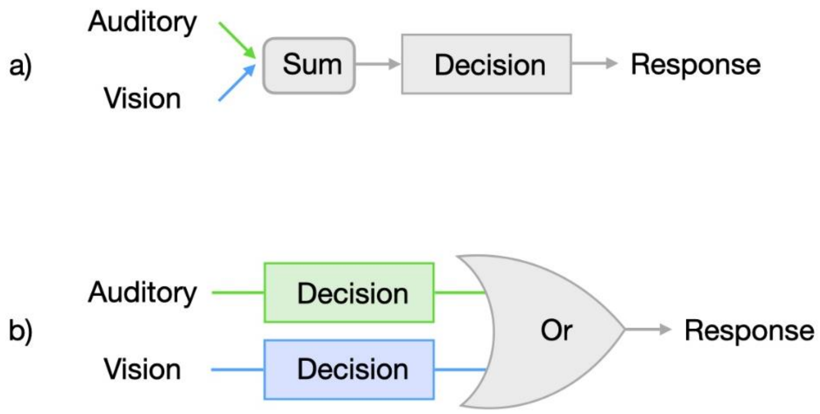


Figure 4.1.1. Schematics of the two possible mechanisms of multisensory processing. a) Integration mechanism, where all the signals from different modalities are pooled together into a single decision unit. Under the integration mechanism, responses in multisensory conditions are faster, because there is more sensory evidence to accumulate to the single decision criterion. b) Race mechanism, where each modality has its own decision unit to accumulate modality-specific sensory evidence. The decision unit that reaches its decision criterion first determines the response, effectively a logical 'or' gate. Figure adapted from Otto and Mamassian (2017; their Figure 2).

Broadly defined, there are two schools of thought regarding a multisensory accumulation framework: integration and race. The integration mechanism proposes that all sensory signals regardless of sensory modality is pooled into one decision unit; the RSE arises because under this integrative framework, the more sensory signals available, the quicker the decision criterion is reached and a response produced, so audio-visual signals naturally lead to faster responses than auditory or visual signals alone (see Figure 4.1.1a; Otto & Mamassian, 2017). On the other hand, the race mechanism proposes that, for example with audio-visual signals, there is an auditory decision unit accumulating auditory signals, and a visual decision unit accumulating visual signals (see Figure 4.1.1b; Otto & Mamassian, 2017). These two decision units are in parallel, and the decision unit that reaches its decision criterion first produces the response (Otto & Mamassian, 2017). Under the race mechanism, the RSE arises because when there are multiple decision units 'racing' each other, there is a quickest decision unit ('winner'), such that overall, with there being a 'winner' decision unit at each decision, responses are faster in multisensory conditions than in unisensory conditions (Otto & Mamassian, 2017). The speedup in the race mechanism is therefore a statistical facilitation, specifically, probability summation (Raab, 1962). Additionally, the race model architecture of two parallel decision units joined by a logic OR gate is desirable, because it perfectly matches the logical ‘or’ between responding to signals in either modality in the redundant signals paradigm (Otto & Mamassian, 2017).

However, in practice, care must be taken in how the race model architecture is used. First and foremost, probability summation underlies all race mechanisms, but probability summation in its simplest form (Raab, 1962) contains two underlying assumptions, neither of which are necessarily correct. The first assumption is that the response times (RTs) measured across trials are statistically independent, in the sense that the RT on a given trial is not affected by the type of stimulation or the response performance in the previous trial. The counterargument for this assumption of statistical independence is that multisensory experiments often randomly interleave trials of different modalities, and it has been empirically shown that switching the modality of stimulation between consecutive trials imparts a cost in RT on the current trial (e.g., Gondan, Lange, Rösler, & Röder, 2004; Innes & Otto, 2019; Otto & Mamassian, 2012; Shaw et al., 2020). For example, the RT on an auditory trial is fast if it was also an auditory trial that preceded it, but the RT on the current auditory trial would be slow if it was a visual trial that preceded it. The second assumption is termed context invariance, and it refers to the notion that there is no interaction between modalities during sensory processing, so for example, a visual decision unit is not affected by the presence or absence of a concurrent auditory signal (e.g., Ashby & Townsend, 1986; Liu & Otto, 2020; Luce, 1986; Otto & Mamassian, 2017; Townsend, Liu, Zhang, & Wenger, 2020; Townsend & Wenger, 2004; Yang, Altieri, & Little, 2018).

4.1.3 Testing the race architecture

The issue of integration versus race mechanisms, and the two assumptions embedded in simple probability summation, all come together in Miller’s (1982) landmark test of multisensory behavioural performance. Miller’s (1982) test starts on the foundation of probability summation. Taking that it is a random variable to describe the time taken for a decision unit to reach the decision criterion, denoted by the letter T, in an audio-visual case, there is an auditory T_A and a visual T_V (Otto & Mamassian, 2017). Therefore, in a race mechanism, the variable for the multisensory decision is the union of T_{AuV}, equal to the minimum of T_A and T_V, as it is the faster of the two unisensory decision units that determines the behavioural response according to the race mechanism (Otto & Mamassian, 2017). The presence of the RSE depends on there being an overlap in the probability densities between T_A and T_V (Otto & Mamassian, 2017). The cumulative probability, P, that audio-visual stimulation AV, triggers a response from the auditory decision unit within time, t, is given by $P_{AV}(T_A \leq t)$. Correspondingly, the cumulative probability (P) that audio-visual stimulation (AV) triggers a response from the visual decision unit within time t is $P_{AV}(T_V \leq t)$. Therefore, the combined probability that either modality has triggered the response, $P_{AV}(T_{A \cup V} \leq t)$, is simply the sum of the unisensory probabilities minus the joint probability of $P_{AV}(T_{A \cap V} \leq t)$:

$P_{AV}(T_{A \cup V} \leq t) = P_{AV}(T_A \leq t) + P_{AV}(T_V \leq t) - P_{AV}(T_{A \cap V} \leq t)$ (adapted from Otto & Mamassian, 2017)

Viewed in a different way, probability summation can be conceptualised in Venn diagram terms (see also Figure 2.4, a Venn diagram used to describe the redundant signals paradigm for audition and vision, also applies perfectly here for probability summation between audition and vision), whereby the union of sets is defined by the addition of the two sets, minus the intersection (Otto & Mamassian, 2017). Note that the intersection of probabilities can only be zero or positive, and in removing the intersection, the left-hand side of the equation can only be equal to or smaller than the right-hand side of the equation, therefore an inequality:

$P_{AV}(T_{A \cup V} \leq t) \leq P_{AV}(T_A \leq t) + P_{AV}(T_V \leq t)$ (adapted from Otto & Mamassian, 2017)

The above inequality is in fact Miller's (1982) test, and it is stating that if probability summation is at work, then responses in multisensory conditions can only be as fast as the fastest unisensory response, in effect an upper bound of redundancy gains under a race architecture (Otto & Mamassian, 2017). Note that in Miller's (1982) test, the assumption of statistical independence has been turned into an assumption of maximal negative correlation between audition and vision, meaning that a fast response in one modality necessarily produces a slow response in the other modality, and this negative correlation is necessary in order to have a redundancy gain at all, and fits with the notion of limited mental capacity (Colonius, 1990).

In practice, Miller's (1982) test has often been violated, meaning multisensory RTs are faster than that predicted by negatively-correlated probability summation. The violation of Miller's (1982) test then commonly leads to the inference that the race mechanism is invalid, and that it must be the alternative mechanism, integration, that is at play (e.g., Cappe et al., 2009), because integration does allow for multisensory RTs to be so fast. However, such an inference in favour of the integration mechanism ignores the fact that Miller's (1982) test, while dropping the assumption of statistical independence, still assumes context invariance. So when there is a violation of Miller's (1982) test, is the problem with probability summation and therefore the race architecture, or is it a problem of the context invariance assumption? It could be that the problem is the assumption of context invariance. For example, in an fMRI study looking into the processing of audio-visual looming signals, concurrent auditory stimulation was found to inhibit visual processing (Gau, Bazin, Trampel, Turner, & Noppeney, 2020), thus showing that the processing in one modality is affected by the presence of signals in the other modality, thereby breaking the assumption of context invariance. Hence, the race architecture could be correct. A comprehensive review of the race model architecture, and the mathematical derivations for probability summation can be found in Otto and Mamassian (2017).

4.1.4 Next steps

If indeed the race architecture is correct, then conveniently, probability summation, which is the foundation of the race architecture, can quite easily be applied to make predictions of multisensory performance. Recall the general equation for probability summation:

$P_{AV}(T_{AuV} \leq t) = P_{AV}(T_A \leq t) + P_{AV}(T_V \leq t) - P_{AV}(T_{A \cap V} \leq t)$ (adapted from Otto & Mamassian, 2017)

If one were to apply probability summation in its simplest form, which assumes both context invariance and statistical independence (independent race model; Raab, 1962), then the intersection term, $P_{AV}(T_{A \cap V} \leq t)$, an unknown, turns into the simple multiplication product of the two unisensory probabilities (Otto & Mamassian, 2017):

$P_{AV}(T_{AuV} \leq t) = P_{AV}(T_A \leq t) + P_{AV}(T_V \leq t) - P_{AV}(T_A \leq t) \text{ x } P_{AV}(T_V \leq t)$ (adapted from Otto & Mamassian, 2017)

Now, with the above equation, which underlies the independent race model (Raab, 1962), there are only unisensory probability terms. If one had a dataset of RTs in auditory and visual conditions, then the RTs can be converted into their respective probabilities, as cumulative distribution functions (CDF) of auditory and visual conditions. Thus, using the above equation, the CDF for the redundant audio-visual condition can be computed. In other words, probability summation, which underlies the race architecture, can be used to make parameter-free predictions of multisensory performance, based only on the empirical unisensory measurements. In an experiment of the redundant signals paradigm (e.g., auditory, visual, audio-visual conditions), one would have the empirically obtained RSE by comparing the redundant condition against the fastest unisensory condition. But crucially, one can also compute the multisensory performance predicted by the independent race model (Raab, 1962), and comparing this prediction against the fastest unisensory condition yields an RSE predicted by the race architecture. Comparing the empirical RSE against the RSE predicted by the race architecture could offer insights into the viability of race mechanisms in general. This is the comparative approach (see also Innes & Otto, 2019 as an example of the comparative approach).

In Chapter 3, I presented two experiments of the redundant signals paradigm, and applying the comparative approach to such experiments would act as a good first test of processing mechanism. However, recall that neither of these two experiments were satisfactory in terms of accuracy and RT performance, and that these issues likely stem from the stimulus design. So first, in two further experiments (Chapters 4.2 and 4.3), I manipulated the stimulus design, in search of better accuracy and RT performance. This culminated in Experiment 4 (Chapter 4.3), which has the best stimulus design thus far in the project. In Chapter 4.4, I apply the comparative approach to the data obtained in Experiment 4.

4.2 Experiment 3: In search of better performance from the audio-visual looming experiment

In Experiment 1 (Chapter 3.1), a replication of Cappe et al. (2009) was performed, and successfully reproduced the finding of faster RTs in redundant than unisensory conditions, and fastest RTs in ALVL. In Experiment 2 (Chapter 3.2), stimulus realism was tested, which improved response accuracy and speed in auditory conditions, but otherwise produced largely similar results as Experiment 1. Both experiments found RSEs to be determined by unisensory factors; there was no looming-specific audio-visual interaction. Hence, no evidence was found supporting the selective integration mechanism proposed by Cappe et al. (2009). However, in performing these two experiments, it was revealed that both response accuracy and speed were uniquely poor in the auditory conditions.

One interpretation of these findings was that there was something in the basic stimulus design that was particularly challenging, manifesting as low response accuracy and slow responses, especially in the auditory modality. Experiment 2 tested additional cues for realism, but improvements seemed minimal and superficial, without changing the multisensory benefits, so perhaps the additional cues were acting as secondary markers to aid signal detection, rather than realism as such. There may be other issues in the stimulus design causing low accuracy and RT performance. Hence, the additional cues for realism was removed, and instead the stimulus design was again examined. On examining the stimulus design of Experiments 1 and 2, and by extension Cappe et al. (2009), a curious feature was identified: concurrent stimulus onset and stimulus motion onset, also known as an onset transient (more on this in Chapter 4.2.1). Thus, this experiment removed the onset transients, to see if this was the fundamental problem of the stimuli.

4.2.1 Methods for Experiment 3

Participants

A new sample of 20 participants (19 female) were recruited for this experiment, using the recruitment techniques (Chapter 2.1). None of these participants participated in the other experiments of this project.

Stimuli

In Experiments 1 and 2, each trial started with a random foreperiod where there was no stimulation. Immediately after the random foreperiod, the stimulus for the trial was presented (stimulus onset). The stimulus motion also started straight away on stimulus presentation (motion onset). This concurrence of the stimulus onset and motion onset is an onset transient (see Figure 4.2.1a), and it could be complicating the experimental task unnecessarily. The experimental task asks for the detection of motion, so only the motion onset is relevant to the task; the stimulus onset is irrelevant to the task, and could be interfering with the detection of motion onset. On auditory trials, where the intensity change may have been difficult, the extra processing demands brought about by this onset transient may have further depressed response accuracy and speed.

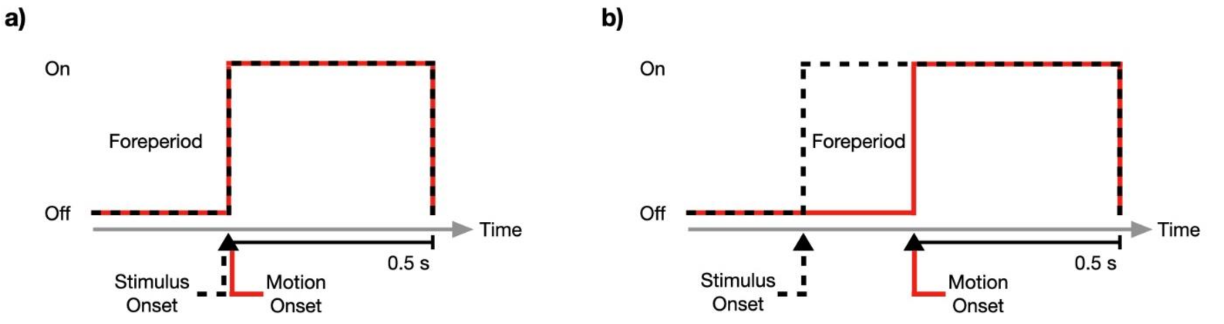


Figure 4.2.1. Stimulus timing on each trial. a) The original stimulus timing from Cappe et al. (2009) and Experiments 1 and 2. After a random foreperiod of no stimulation, the onsets of stimulus (black dashed line) and motion (red line) were concurrent. This sudden presentation of stimulus from no stimulation, with the concurrence of task-relevant motion onset, and task-irrelevant stimulus onset, could have been too demanding. b) The new stimulus timing separates motion onset (red line) from stimulus onset (black dashed line). The stimulus is first presented at the start intensity for a random period, then afterwards, the stimulus undergoes its motion as before. Figure taken from my publication (Chua, Liu, Harris, & Otto, 2022).

Hence, for this experiment, stimulus onset was decoupled from the motion onset (see Figure 4.2.1b). On each trial, the stimulus was first presented for a random duration, at the start value of auditory intensity (40 dB SPL) or visual size (7°). Once the random duration had passed, then the already-present stimulus began its motion as before, i.e., ± 20 dB SPL for simulated auditory motion-in-depth, ± 6° for simulated visual motion-in-depth, over a duration of 0.5 seconds. The set of trials are the same as before, as outlined in the general methods (Chapter 2.3 *Experiment design* and Chapter 2.4 *Stimuli*).

Apparatus, procedures and data analysis are all as described in the corresponding sections of Chapter 2

4.2.2 Results

Accuracy performance

As before, the first aspect of the data to examine is the accuracy performance, which is defined here as the hit rate and false alarm rate. The target is near ceiling levels of performance, because the stimuli are simple and salient, so accuracy should be near perfect, and this is assumed in the modelling that will be performed in Chapter 5. Also of interest is if the removal of the onset transient remedies the low accuracy performance towards the auditory conditions, found in the replication of Cappe et al. (2009). First, an analysis on hit rates, then the false alarm rates.

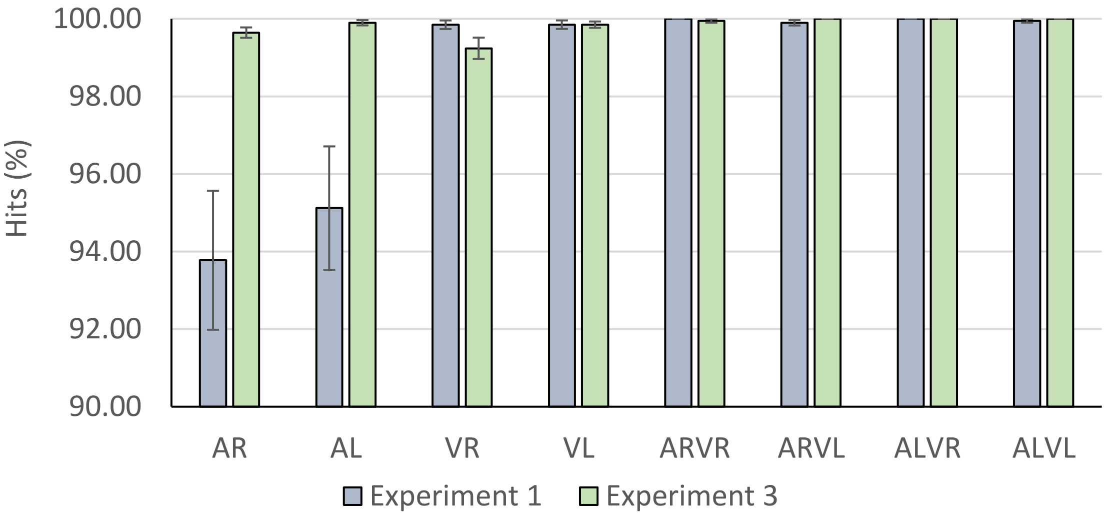


Figure 4.2.2. Participant-averaged hit rates across the conditions (SEM error bars). Experiment 1 is the replication experiment (grey bars, see Chapter 3.1). Experiment 3 is the current experiment with the onset transients removed (green bars). The removal of onset transients has improved the hit rates in both auditory conditions (AR, AL), and are now comparable to the ceiling hit rates in the non-auditory conditions.

To check if the hit rates on the auditory conditions in this experiment (Figure 4.2.2, Experiment 3, green bars) have improved from the replication experiment (Figure 4.2.2, Experiment 1, grey bars), an independent-samples t-test was performed, comparing between experiments the AR and AL hit rates. AR hit rates were significantly higher in Experiment 3 than in Experiment 1 ($t_{(19.213)}$ = 3.265, p = 0.004, mean difference: 5.869% ± 1.798%, equal variances not assumed). AL hit rates were significantly higher in Experiment 3 than in Experiment 1 ($t_{(19.072)}$ = 3.000, p = 0.007, mean difference: 4.779% ± 1.593%, equal variances not assumed). Therefore, the removal of onset transients in this experiment have significantly increased the hit rates in auditory conditions, compared to the replication experiment.

Next, to check if the auditory hit rates are now on par with the non-auditory conditions, which appear to be at ceiling level, multiple paired-samples t-tests were performed, comparing AR or AL against each non-auditory condition. The audio-visual conditions have significantly higher hit rates than the AR hit rate (all $p<0.05$), but the hit rates in unisensory visual conditions were statistically equal to the AR hit rate (AR-VR, $t_{(19)}$ = 1.318, p = 0.203; mean difference: 0.405% ±0.307%; AR-VL, $t_{(19)}$ = -1.472, p = 0.157, mean difference: -0.205% ±1.39%). AL hit rates were statistically on par with all the non-auditory conditions (all $p>0.05$), except for the VR hit rate, where the AL hit rate was significantly higher than the VR hit rate ($t_{(19)}$ = 2.659, p = 0.016, mean difference: 0.659% ±0.248%). Thus, it appears that the hit rates in the auditory conditions are comparable to the hit rates in non-auditory conditions, so hit rates in all conditions are at ceiling level in Experiment 3.

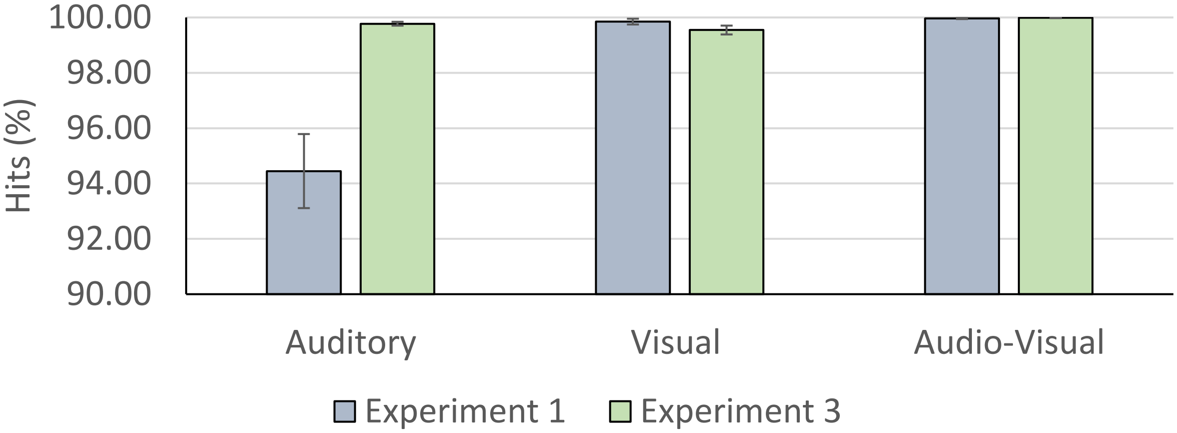


Figure 4.2.3. Hit rates by sensory modality (motion directions averaged, SEM error bars). Experiment 1 (replication; Chapter 3.1) in grey, Experiment 3 (current experiment with onset transients removed) in green bars. The removal of onset transients improved the low auditory hit rates in the replication experiment.

Next, the hit rates were checked for effects on the level of sensory modality. For each sensory modality, the motion directions were averaged, and a one-way ANOVA was performed with sensory modality as the factor. A significant main effect of sensory modality was found ($F_{(1.276, 24.253)}$ = 5.399, p = 0.022, η_p^2 = 0.221, Greenhouse-Geisser corrected). Pairwise comparisons revealed that the audio-visual hit rate was significantly higher than the auditory hit rate (mean difference: 0.216% ±0.068%, p = 0.014) and the visual hit rate (mean difference: 0.0442% ±0.162%, p = 0.039). No difference was found between the auditory hit rate and the visual hit rate (p = 0.399). This result shows that the auditory hit rate is now equal to the visual hit rate, and both are only slightly behind the audio-visual hit rate.

Finally, the sensory modality hit rates were compared against the corresponding values from Experiment 1. A mixed-factors ANOVA was performed on the hit rates, with sensory modality as the within-subjects factor, and experiment as the between-subjects factor. A significant interaction was found between the sensory modality and experiment factors ($F_{(1.032, 39.216)}$ = 16.739, p<0.001, Greenhouse-Geisser corrected). Pairwise comparisons found that it was only the auditory condition which was driving the interaction, with significantly higher auditory hit rates in Experiment 3 than in Experiment 1 (mean difference: 5.324% ±1.340%, p<0.001). The between-experiment differences in other modalities were not significant (p>0.05). This result shows that the removal of the onset transients has exclusively lifted the poor auditory hit rates in the replication experiment.

For false alarm rates in this experiment (Figure 4.2.4, Experiment 3, green bars), a one-way ANOVA was used to check if there was an effect of modality. The ANOVA returned a non-significant result ($F_{(2,38)}$ = 1.601, p = 0.215, η_p^2 = 0.078). To be sure, the pairwise comparisons between the modalities were also checked, but all were non-significant (p>0.05). Hence, it appears that the false alarm rates were statistically the same between all modalities.

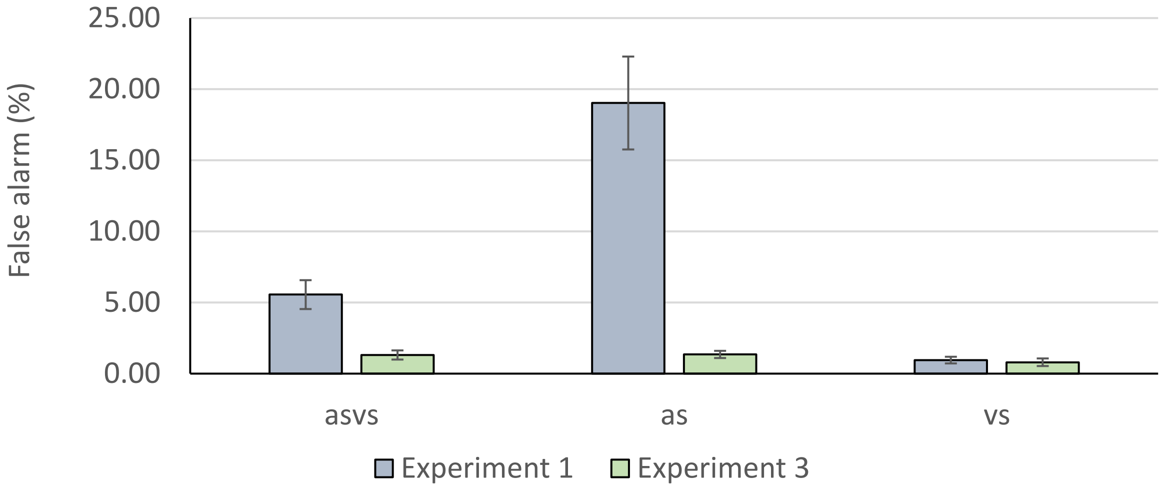


Figure 4.2.4. Participant-averaged false alarm rates (SEM error bars). Experiment 1 (replication; Chapter 3.1) in grey, Experiment 3 (current experiment with onset transients removed) in green. Removing the onset transients reduced the false alarm rate in both auditory and audio-visual conditions, and both are on par with the visual condition. Note that this false alarm rate performance in Experiment 3 is a further improvement over that found with more realistic stimuli (Experiment 2; Chapter 3.2).

To compare the false alarm rates between this experiment (Figure 4.2.4, Experiment 3, green bars) and the replication experiment (Figure 4.2.4, Experiment 1, grey bars), a mixed-factors ANOVA was performed, with modality as the within-subjects factor, and experiment as the between-subjects factor. A significant interaction between modality and experiment was found ($F_{(1.140,\ 43.302)} = 23.927$, $p<0.001$, $\eta_p^2 = 0.386$, Greenhouse-Geisser corrected). Pairwise comparisons on the interaction found that the auditory false alarm rate in this experiment (Figure 4.2.4, Experiment 3, green bar) were significantly lower than in the replication experiment (Figure 4.2.4, Experiment 1, grey bar; mean difference: 17.677% ± 3.275%, p<0.001). The audio-visual false alarm rate in this experiment was also significantly smaller in this experiment than in the replication experiment (mean difference: 4.245% ±1.064%, p<0.001). The visual false alarm rate was statistically the same between this experiment and the replication experiment (p = 0.674). Hence, the removal of the onset transient in this experiment had reduced the auditory and audio-visual false alarm rates, compared to the replication experiment where both these false alarm rates were quite elevated, in absolute terms and in comparison to the visual false alarm rate.

In summary, the removal of onset transients seems highly beneficial to the accuracy performance. Experiment 1 found a significant drop-off in accuracy unique to the auditory conditions, with low hit rates and high false alarm rates. Now, with only the removal of the onset transients in both auditory and visual stimuli, the auditory hit rates were elevated to be on the same level as the visual hit rates, and close to the audio-visual hit rates. Furthermore, the removal of onset transients has reduced the false alarm rates of both the auditory condition and the audio-visual condition, such that the false alarm rates are similarly low in all three sensory conditions. Altogether, the overall hit rate in this experiment was 99.82% ±0.09%, while overall correct (proportion of hits and correct rejections, against all trials) was 99.48% ±0.09%.

Response time performance

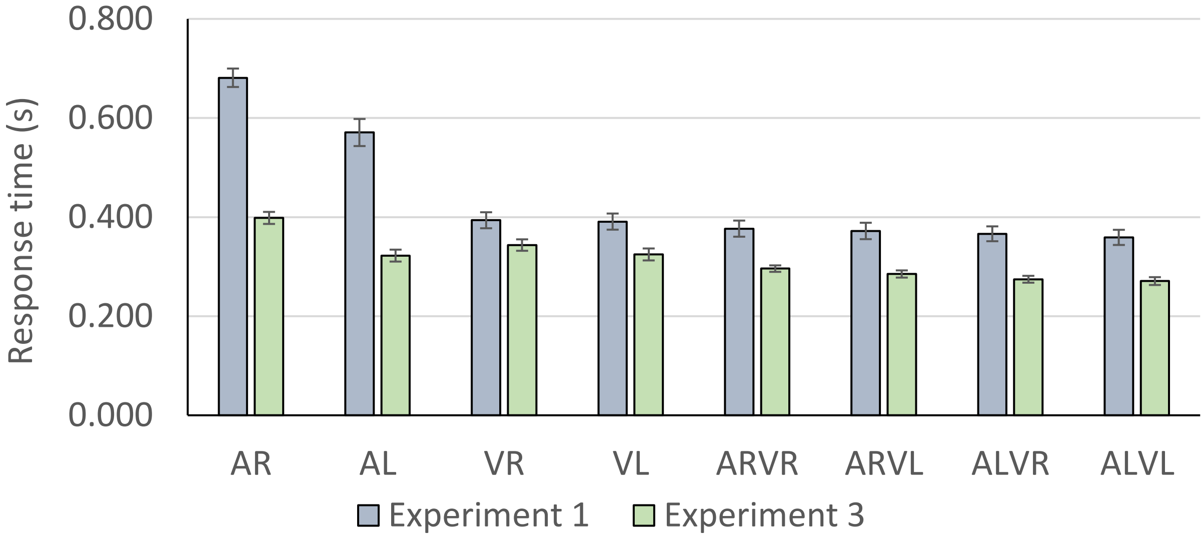


Figure 4.2.5 Participant-averaged RTs across conditions (SEM error bars). The current experiment without onset transients is Experiment 3 (green bars). The replication is Experiment 1 (grey bars). Removing onset transients have sped up the RT in all conditions.

Visually (see Figure 4.2.5), it appears that the removal of the onset transient in this experiment (Experiment 3, green bars in Figure 4.2.5) has reduced the auditory RTs, to the point where the auditory RTs are almost on par with the visual RTs, which represents quite an improvement from the replication experiment (Experiment 1, grey bars in Figure 4.2.5) where the auditory RTs were uniquely slow. The RTs of this experiment also seem to be faster across all conditions, compared to the replication experiment.

To check if the RTs in this experiment (Figure 4.2.5, Experiment 3, green bars) were indeed faster than in Experiment 1 (grey bars), an independent-samples t-test was performed, comparing between experiments all eight conditions. The independent samples t-test revealed that all conditions in this experiment had significantly faster RTs compared to the corresponding RTs in the Experiment 1 (6 out of 8 were $p<0.001$, the other two were the visual conditions, which were $p = 0.016$ for VR, $p = 0.002$ for VL, equal variances not assumed). It appears that the removal of onset transients improved RTs universally across sensory modalities.

To check for the unisensory looming biases, paired-samples t-tests were performed, comparing the RTs towards looming and receding, for each modality. RTs towards AL were significantly faster than RTs towards AR ($t_{(19)} = 14.827$, $p<0.001$, mean difference: 0.076 s ±0.005 s). RTs towards VL were significantly faster than RTs towards VR ($t_{(19)} = 5.064$, $p<0.001$, mean difference: 0.019 s ±0.004 s). Thus, the unisensory auditory and visual looming biases were present in this experiment.

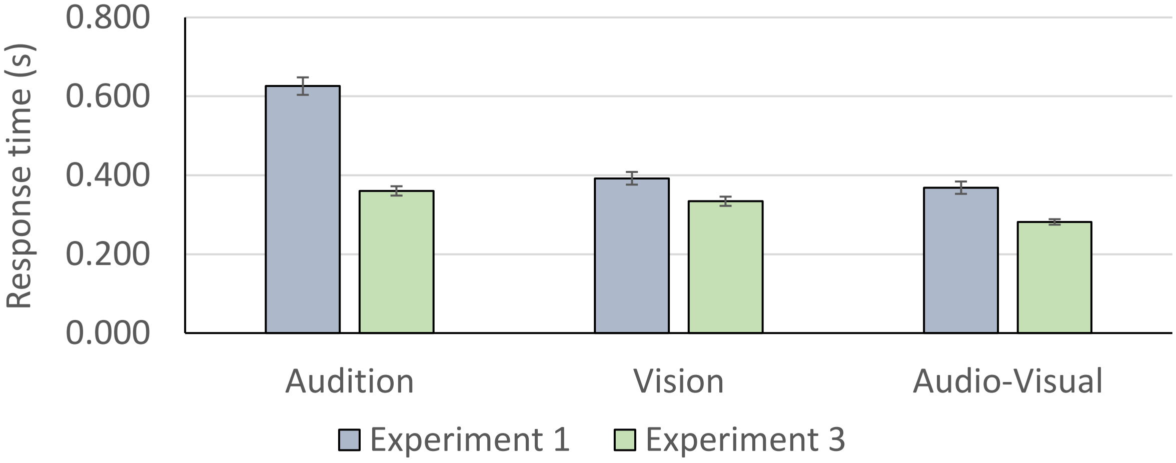


Figure 4.2.6. RTs by sensory modality, motion direction averaged (SEM error bars). Experiment 3 is the current experiment with onset transients removed (green bars). Experiment 1 is the replication (grey bars).

Next, the RTs of this experiment were checked for modality effects. The RTs were grouped by their sensory modality, averaging the motion directions (Figure 4.2.6, Experiment 3, green bars). A significant effect of modality was found ($F_{(2, 38)} = 63.759$, $p<0.001$, $\eta_p^2 = 0.770$). Pairwise comparisons showed that RTs in the audio-visual condition were significantly faster than RTs in the visual condition (mean difference: 0.053 s ±0.007 s, $p<0.001$) and RTs in the auditory condition (mean difference: 0.079 s ±0.006 s, $p<0.001$). RTs in the visual condition were also significantly faster than RTs in the auditory condition (mean difference: 0.026 s ±0.008 s, $p = 0.009$). Hence, Cappe et al.'s (2009) finding that audio-visual RTs are faster than the unisensory RTs were still preserved when the onset transients were removed.

To compare the modality-based RTs between Experiment 1 (Figure 4.2.6, grey bars) and this experiment (Figure 4.2.6, Experiment 3, green bars), a mixed-factors ANOVA was performed, with sensory modality as the within-subjects factor, and experiment as the between-subjects factor. A significant interaction between sensory modality and experiment was found ($F_{(1.380,\ 52.439)} = 149.012$, $p<0.001$, $\eta_p^2 = 0.797$). Pairwise comparisons within this interaction revealed that for all three modalities, the RTs in the current experiment were faster than the RTs in Experiment 1 (mean difference for audition: 0.266 s ±0.025 s, $p<0.001$; mean difference for vision: 0.058 s ± 0.020 s, $p = 0.006$; mean difference for audio-visual: 0.087 s ±0.017 s, $p<0.001$). Thus, the removal of onset transients sped up RTs across all conditions.

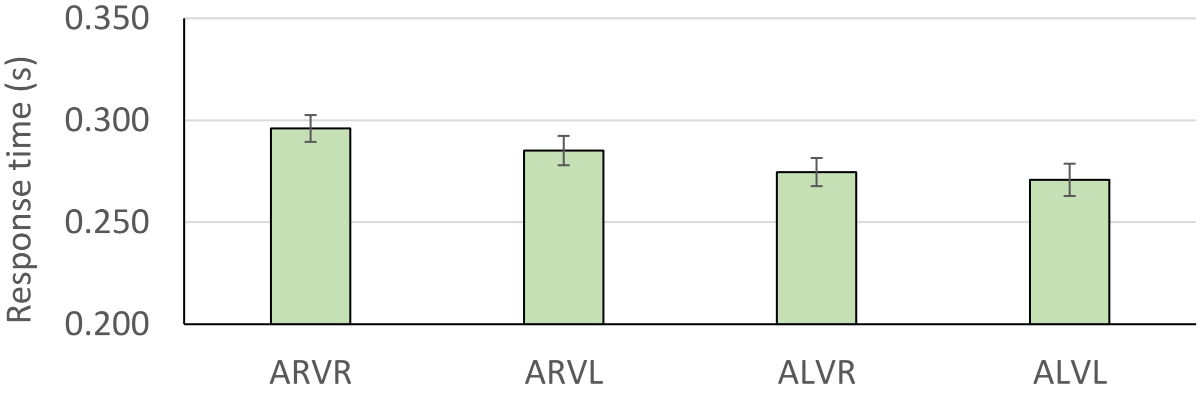


Figure 4.2.7. Participant-averaged RTs in the audio-visual conditions (SEM error bars).

Finally, the RTs in the audio-visual conditions were examined (Figure 4.2.7). A paired-samples t-test was performed to compare the RT of ALVL, to the next fastest RT, that of ALVR. There was no significant difference between the RTs of ALVL and ALVR ($t_{(19)} = 1.840$, $p = 0.081$, mean difference: 0.004 s ± 0.002 s). Cappe et al.'s (2009) finding of fastest RTs in the ALVL conditions among the audio-visual conditions were not replicated in this experiment without onset transients.

To check for unisensory or multisensory effects, a 2x2 ANOVA was performed on the audio-visual RTs (Figure 4.2.7), with factors auditory motion direction (AR, AL) and visual motion direction (VR, VL). A main effect of auditory motion direction was found ($F_{(1,19)} = 62.674$, $p<0.001$, $\eta_p^2 = 0.767$), which pairwise comparisons showed was that the conditions with an AL component had significantly faster RTs than conditions with an AR component (mean difference: 0.018 s ± 0.002 s, $p<0.001$). A main effect of visual motion direction was found ($F_{(1,19)} = 12.174$, $p = 0.002$, $\eta_p^2 = 0.391$), which pairwise comparisons found was that conditions with a VL component had significantly faster RTs than conditions with a VR component (mean difference: 0.007 s ±0.002 s, $p = 0.002$). There was also a significant interaction between auditory motion direction and visual motion direction ($F_{(1,19)} = 7.516$, $p = 0.013$, $\eta_p^2 = 0.283$). On examining the simple effects of this interaction, it was discovered that in AR, VL had faster RTs than VR (mean difference: 0.011 s ± 0.003 s, $p = 0.001$), but in AL, VL and VR had statistically the same RTs ($p = 0.081$). In terms of vision, in VR, AL had significantly faster RTs than AR (mean difference: 0.022 s ±0.003 s, $p<0.001$), while in VL, AL had significantly faster RTs than AR (mean difference: 0.014 s ±0.002 s, $p<0.001$). It appears that there are some dependencies between the sensory modalities and the motion direction, however, one must be clear that the absolute RTs, as shown here, should not be equated with the redundancy gain. The redundancy gain is the RSE.

In summary, the removal of onset transients have sped up RTs across all conditions. Interestingly, Cappe et al.'s (2009) main findings were only partially replicated here. First, RTs in the audio-visual condition were faster than the RTs in the unisensory auditory or visual conditions, which agrees with Cappe et al. (2009). However, RTs in the ALVL condition were not significantly faster than the RTs in the other audio-visual conditions.

Redundant signals effect

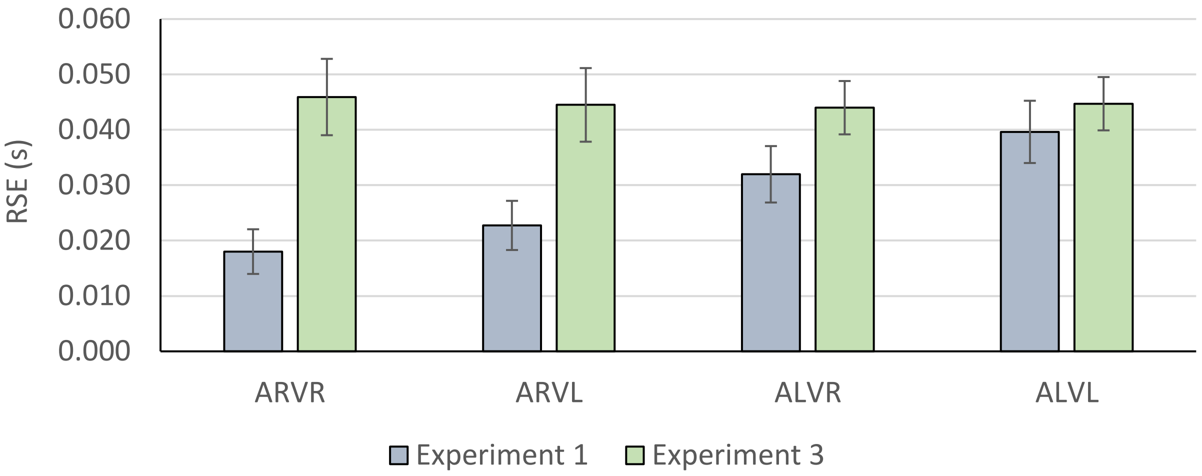


Figure 4.2.8. Participant-averaged RSEs (SEM error bars). Experiment 3 is the current experiment with onset transients removed (green bars). Experiment 1 is the replication (grey bars). With the removal of onset transients, there were increases in RSEs in conditions with an AR component, thus equalising the RSEs between the audio-visual conditions.

The RSE is the correct characterisation of the gains made in redundant conditions (see Chapter 2.6.5 *Redundant signals effect*). To see if the removal of onset transients in this experiment also affected the RSEs, an independent-samples t-test was used to compare the RSEs between this experiment (Figure 4.2.8, Experiment 3, green bars) and Experiment 1 (Figure 4.2.8, grey bars). The independent-samples t-test found that the conditions with an AR component had larger RSEs in this experiment than Experiment 1 (ARVR: $t_{(38)} = 3.484$, $p = 0.001$, mean difference: 0.028 s ±0.008 s; ARVL: $t_{(38)} = 2.725$, $p = 0.010$, mean difference: 0.022 s ±0.008 s). However, the conditions with an AL component showed no difference in RSE between experiments (ALVR: $t_{(38)} = 1.716$, $p = 0.094$, mean difference: 0.013 s ±0.007s; ALVL: $t_{(38)} = 0.687$, $p = 0.496$, mean difference: 0.005 s ±0.007 s). It appears that the removal of onset transients have lifted the RSE in conditions with an AR component.

Curiously, in this experiment, it was the ARVR condition which had the largest RSE, a sensory condition that is consistent with a de-escalation of danger, and should not require quick processing and responses (behavioural urgency hypothesis; Franconeri & Simons, 2003). The second largest RSE belonged to the ALVL condition. To check if this RSE difference between

ARVR and ALVL was statistically significant, a paired-samples t-test was performed. The RSE of ARVR was not significantly larger than the RSE of ALVL ($t_{(19)}$ = 0.244, p = 0.810, mean difference: 0.001 s ±0.005 s).

A 2x2 ANOVA was performed on the RSEs to determine any unisensory or multisensory effects, with factors auditory motion direction (AR, AL) and visual motion direction (VR, VL). Crucially, there was no interaction between auditory motion direction and visual motion direction ($F_{(1,19)}$ = 0.279, p = 0.603, η_p^2 = 0.014). There was also no main effect of auditory motion direction ($F_{(1,19)}$ = 0.037, p = 0.850, η_p^2 = 0.002) nor visual motion direction ($F_{(1,19)}$ = 0.019, p = 0.892, η_p^2 = 0.001). Thus, there was nothing to suggest a motion-specific audio-visual interaction that would be characteristic of a selective integration mechanism towards audio-visual looming signals (Cappe et al., 2009).

Altogether, it appears that the removal of onset transients has increased the RSEs in conditions with an AR component. With onset transients (Experiment 1), the conditions with AR had smaller RSEs than conditions with AL. Thus, without onset transients, the RSEs are of similar size between all four audio-visual conditions.

4.2.3 Discussion

At the beginning of this chapter, it was pointed out that Experiments 1 and 2 (Chapter 3) may have revealed a fundamental design flaw in the stimuli. Experiment 1 (Chapter 3.1), the replication of Cappe et al. (2009), found low accuracy and slow responses, particularly for the auditory conditions. Experiment 2 (Chapter 3.2) explored the use of more motion-in-depth cues on the stimuli, but its improvements were superficial and limited to the auditory conditions, without benefitting the non-auditory conditions. It was postulated that the additional motion-in-depth cues were only acting as a secondary markers to aid signal detection, whilst masking a fundamental stimulus issue. The fundamental issue was suspected to be onset transients on the stimuli: the stimulus onset is concurrent with motion onset. Only the motion onset is relevant to the task, so the concurrent stimulus onset could be an expensive distraction, processing-wise. Hence, in this experiment, the stimuli returned to the basic auditory intensity and visual size changes to represent motion-in-depth, but, the presentation timing was changed (stimulus presented first, motion onset later after random duration) to remove onset transients.

The removal of onset transients produced comprehensive improvements in response accuracy and speed. First, response accuracy, particularly in the auditory conditions, were a weak point in Experiment 1, which had onset transients on its stimuli. Yet, by removing the onset transients in the stimuli, the hit rates in auditory conditions were elevated to a level on par with the ceiling-performance non-auditory conditions. Additionally, the removal of onset transients on the stimuli also greatly reduced the false alarm rate in auditory conditions, which were very high in Experiment 1. As an additional bonus, the removal of onset transients also lowered the false alarm rate in audio-visual conditions, which were also rather high in Experiment 1. Overall, the removal of onset transients has equalised the hit rate to a very high level, while also equalising the false alarm rate to a very low level, across all conditions and sensory modalities. In terms of RTs, the removal of onset transients sped up the RTs across all conditions, and again, the finding of fastest RTs in audio-visual conditions compared to the unisensory conditions was replicated from Cappe et al. (2009). Hence, it appears that by removing the onset transients, I have obtained a seemingly perfect dataset of near perfect accuracy and universally fast RTs. This seemingly perfect dataset might appear to be a good basis for applying multisensory frameworks under the comparative approach – the stated aim of this chapter.

However, within the data, there are a few points to discuss. First, Cappe et al. (2009) found that RTs were fastest in the congruent audio-visual looming condition, ALVL, compared to other audio-visual conditions. The finding of fastest RTs in the ALVL condition (Cappe et al., 2009) was reproduced in Experiments 1 and 2 (Chapter 3). However, this experiment with the onset transients removed did not find RTs to be quickest towards the ALVL condition. Instead, the ALVL condition was equal with the ALVR condition on RT speed, despite the ALVR signal representing an incongruent motion between the sensory modalities – it is not immediately obvious why one should respond quickly to this sensory conflict. Could this result suggest that it is not about looming, the dangerous motion as such, but rather that the multisensory performance is determined from the unisensory performance? Put in another way, ALVL RTs were specially fast in earlier experiments possibly because the unisensory receding conditions were underperforming, perhaps due to stimulus artifacts such as onset transients. Yet, now that all unisensory conditions are within a small range of RTs (Figure 4.2.5), which unisensory conditions are combined together is not so influential on the resulting audio-visual RT (Figure 4.2.7). Cappe et al. (2009) made a claim of selective integration towards ALVL signals, partly on the back of the finding that RTs towards ALVL were particularly fast, and a postulation that fast responses to ALVL are advantageous for survival. If responses towards ALVL were specially fast because of survival, then this advantage should be preserved no matter what the comparison conditions are. Yet, here, in this experiment with faster and more equal RTs across all unisensory conditions, the ALVL RTs were not any faster than the RTs towards the other supposedly non-dangerous audio-visual conditions. It appears that the finding of fastest RTs in the ALVL condition in early experiments was an artifact of difficult stimuli, in particular the auditory stimuli, and not about a special, fast and adaptive process for audio-visual looming signals. Nonetheless, absolute RTs are not quite the correct measure of multisensory processing – the RSE should be examined.

The correct characterisation of the redundancy gain is the RSE. If there was special multisensory processing towards ALVL (Cappe et al., 2009), then this should manifest as a particularly large RSE towards only the ALVL condition. However, numerically, the RSE towards ARVR was actually largest, albeit only by a very small amount and not at all reaching significance (Figure 4.2.8). In fact, the RSEs across all four audio-visual conditions were practically identical (Figure 4.2.8); the ANOVA on these RSEs found no unisensory nor multisensory effects. Such equal RSEs across conditions, derived from unisensory RTs that were all within a narrow range, can be associated with the principle of congruent effectiveness (Otto et al., 2013). The principle of congruent effectiveness states that there can only be a redundancy gain (RSE) if the unisensory components are similar in performance (Otto et al., 2013). This principle (Otto et al., 2013) was exemplified in Experiments 1 and 2 (also Cappe et al., 2009), where RT performance in auditory conditions were poor relative to the visual conditions, in particular with the AR condition. Indeed, the resulting combination of audition and vision produced small RSEs, with particularly small RSEs towards combinations involving an AR component (see also Figure 4.2.8, grey bars). In the current experiment without onset transients, the unisensory auditory and visual RTs were improved and equalised, and the resulting RSEs were also enlarged and equal across the conditions, exactly in accordance with the principle of congruent effectiveness (Otto et al., 2013). Crucially, the principle of congruent effectiveness is based on probability summation, the statistical facilitation that underlies the race architecture. It appears this series of RSE results gives weight to the potential of the race architecture.

However, this experiment is not the complete success that it may appear to be. While the removal of onset transients may have brought about very positive changes (e.g., response accuracy and speed), the specific implementation of this experiment may have introduced another problem. To explain, recall the basic design of the stimuli (see Chapter 2.3 *Experiment design*, also Figure 2.5). In the auditory condition, there is only auditory stimulation throughout the trial. In the visual condition, there is only visual stimulation throughout the trial. In the audio-visual condition, there is both auditory and visual stimulation through the trial. For this experiment, the onset transients were removed by first presenting the stimulus at its start value (40 dB SPL, 7° visual size) for a random duration, after which the motion started (Figure 4.2.1 for a visualisation of this stimulus timing). The problem is that the newly-added period of stimulation at the start value is in effect giving away the modality that will be called on for the motion signal in the trial. It is not clear if this early cueing of the modality ahead of the motion signal is in some way priming the participant, perhaps to focus resources towards the modality that was cued, for example, and could in effect be giving the participant a head start to produce top performance on every trial. This early cueing of the modality was addressed in Experiment 4.

4.3 Experiment 4

Experiment 3 tested the removal of onset transients, whilst retaining the basic stimuli of auditory intensity change and visual size change. The removal of onset transients brought significant improvements to the data, in terms of response accuracy and speed, compared to Experiments 1 and 2 which both had onset transients. Thus, it appears that the onset transients in the stimuli are possibly related in some way to the difficulty of detecting or distinguishing the stimuli, and that the onset transients seems to represent a critical design flaw on the stimulus.

However, in Experiment 3, the way in which onset transients were removed may have opened the door to a new problem. Onset transients were removed by separating stimulus onset from motion onset: the stimulus was first presented, holding at the initial stimulus value (40 dB SPL on the auditory signal, or 7° in the visual signal) for a random duration, after which the motion was presented (see Figure 4.2.1). With the original definitions of the sensory conditions, unisensory signals were presented in the absence of stimulation in the other modality, and multisensory signals were presented with stimulation in both modalities. Hence, the random foreperiod of stimulation in effect gave away which modality the trial would involve, and could perhaps be priming the participant for the response towards the subsequent task-relevant motion onset. It is unknown if there is such priming, or indeed if the early clues to the modality involved is unduly advantageous to response performance. Hence, this experiment re-defined the sensory conditions, thus correcting for this potential problem of modality priming.

4.3.1 Methods for Experiment 4

Participants

For this experiment, a new sample of 20 participants (17 female, 19 right-handed, 10 with corrected-to-normal eyesight, none with declared hearing impairments, average age: 19.9 years ±2.5 years S.D.) were recruited, using the participant recruitment techniques outlined in Chapter 2.1 *Participants*. None of these participants have participated in the other experiments of this project.

Stimuli

The removal of onset transients meant the stimulus was presented first for a random duration (foreperiod), acting as a lead-in to the motion onset. The original definition of the sensory conditions had the stimulation exclusively per the sensory modality. Thus, when both stimulus design features (onset transient removal, modality-exclusive stimulation) came together in Experiment 3, the participant received a cue to the modality during the foreperiod, ahead of the task-relevant motion onset. The remedy was to redefine the sensory conditions, such that in all conditions, there was both auditory and visual stimulation (Figure 4.3.1).

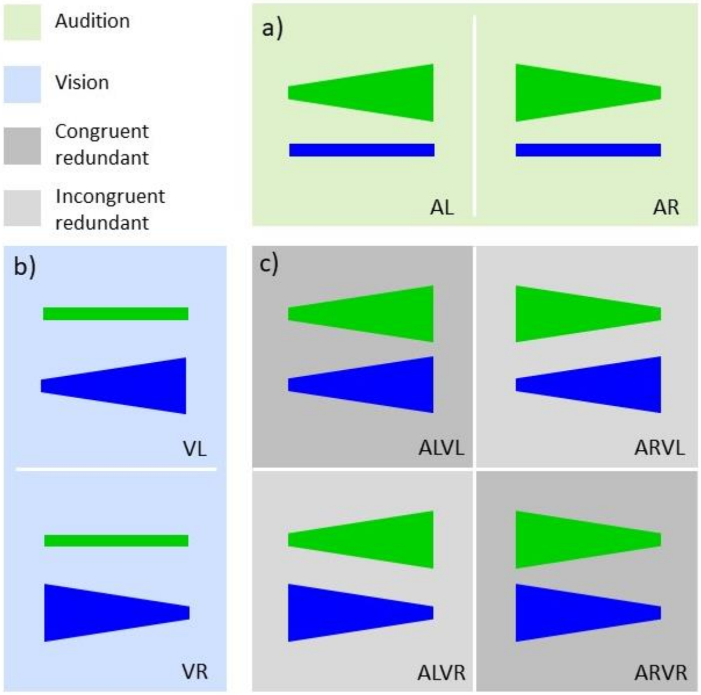


Figure 4.3.1. The newly defined stimulus conditions for Experiment 4. The changes were to the definition of the unisensory conditions, which no longer exclusively feature stimulation of the modality. a) In the auditory modality, there were looming or receding signals, as before, but for this experiment, there was also concurrent visual stimulation of a constant 7° circular object. b) In the visual modality, there were looming or receding signals as before, but in this experiment, there was also concurrent auditory stimulation of a constant 40 dB SPL 1000Hz tone. c) The audio-visual combinations were the same as before. The purpose of this new definition of stimulus condition was so that all modalities have an auditory and visual component, thus avoiding a foreperiod that gives away the modality involved in the trial.

In the unisensory auditory condition, as before, it could either be a looming signal (intensity increase, from 40 dB SPL to 60 dB SPL), a receding signal (intensity decrease, from 40 dB SPL to 20 dB SPL), or an auditory catch (constant 40 dB SPL). However, alongside the auditory stimuli, there would be visual stimulation at the start value (7° visual size) throughout the duration of each auditory trial (see Figure 4.3.1a, green area). This is in contrast to the old definition of the unisensory auditory condition, used in my earlier experiments, which did not have the concurrent visual stimulation alongside. In the unisensory visual condition, as before, it could either be a looming signal (size increase, from 7° to 13°), a receding signal (size decrease, from 7° to 1°), or a visual catch (constant 7°). However, alongside the visual stimuli, there was also auditory stimulation at the start value (40 dB SPL, 1000 Hz) throughout the duration of each auditory trial (see Figure 4.3.1b, blue area). This is in contrast to the old definition of the unisensory visual condition, used in my earlier experiments, which did not have the concurrent auditory stimulation alongside. The audio-visual conditions (Figure 4.3.1c, grey area) were unchanged from Experiment 3, because these conditions already have both auditory and visual stimulation.

The presentation of auditory and visual stimulation were simultaneous, and began right from the stimulus onset, both holding at their start values (40 dB SPL, 7°) for a random duration, after which the motion signal was presented. The advantage of this new definition of sensory conditions is that in the moments before motion onset, all conditions would be the same: 40 dB SPL 1000 Hz tone as auditory stimulation, 7° circular object as visual stimulation. Therefore, the foreperiod stimulation is not informative of the upcoming motion signal, so there should not be cueing or priming effects.

The new operational definitions for the sensory conditions also necessitated changes to the catch trials. Previously, with unisensory conditions that were exclusively of the modality, there was an auditory catch, a visual catch, and an audio-visual catch, each presented in a 2:1 signal-to-catch ratio within the modality (see Figure 2.5 of Chapter 2.3 *Experiment design*). Now, because the unisensory signal conditions feature concurrent non-signal stimulation in the other modality, the catch trials must also be audio-visual, to match (Figure 4.3.2). All catch trials were audio-visual in this experiment, and to preserve the 2:1 signal-to-catch ratio, the one auditory catch of the set was replaced with an audio-visual catch (Figure 4.3.2a), the one visual catch was replaced with an audio-visual catch (Figure 4.3.2b), while the audio-visual condition retained their original two audio-visual catch trials (Figure 4.3.2c). Hence, overall, the set of trials contained eight signal trials, and four catch trials, again the 2:1 signal-to-catch ratio.

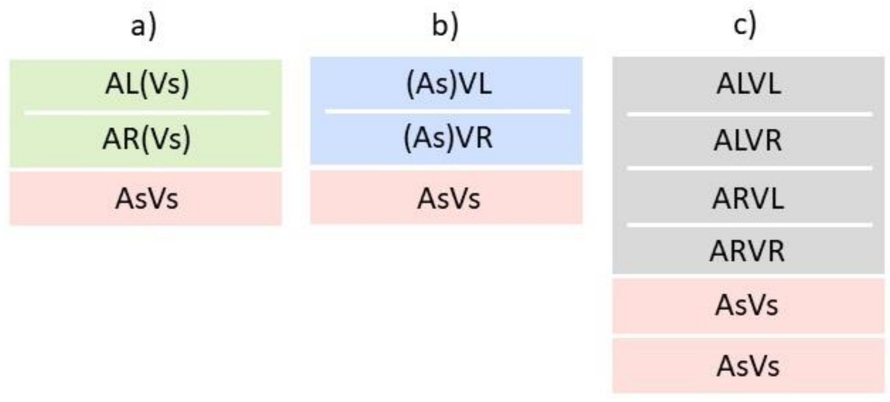


Figure 4.3.2. The new set of trials for this experiment. As there were no unisensory conditions with stimulation exclusively of the modality, the catch trials were all audio-visual, to match. a) The new definition of the auditory condition, with two signal trials now presented with concurrent visual stimulation (AL(Vs), AR(Vs); green box), and an audio-visual catch trial to match. b) The new definition of the visual condition, with two signals trials now presented with concurrent auditory stimulation ((As)VL, (As)VR; blue box), and an audio-visual catch trial to match. c) The audio-visual condition, which is the same as before, with four signal trials (the 2x2 combination of auditory and visual motions), and two audio-visual catch trials. Notice that overall, the 2:1 signal-to-catch ratio was maintained for this experiment.

Apparatus, procedures and data analysis are all as described in the corresponding sections of Chapter 2

4.3.2 Results

Accuracy performance

Ceiling performance accuracy is necessary if more advanced analysis techniques (i.e., modelling, Chapter 5) are to be applied to the data. Thus, the first analysis of this experiment is on the accuracy, defined as hit rates and false alarm rates.

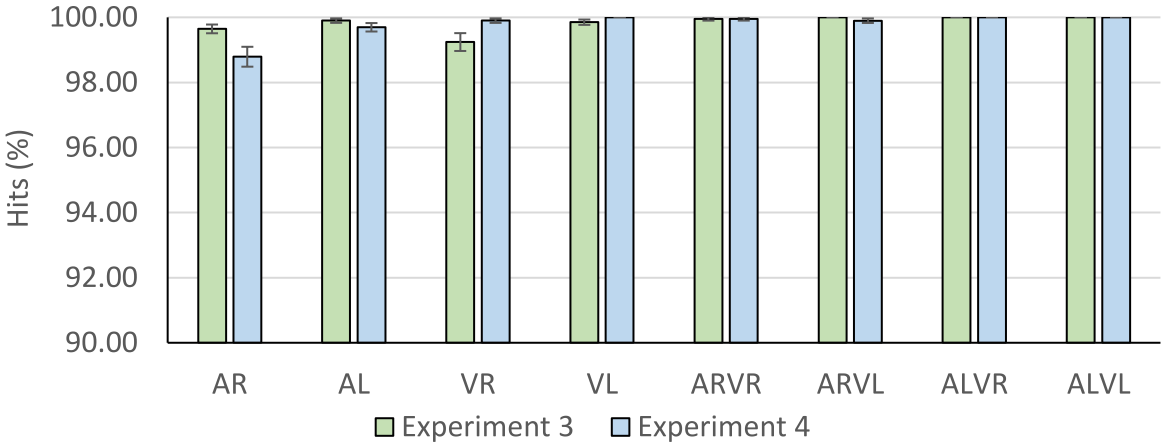


Figure 4.3.3. Participant-averaged hit rates across conditions (SEM error bars). The current experiment is Experiment 4, re-defined stimulus conditions (blue bars). Experiment 3 uses the original stimulus definitions (green bars). The hit rates were similar across conditions, with the only significant difference being the AR and VR conditions.

Taking that Experiment 3 found near-perfect accuracy by removing onset transients (see Figure 4.3.3, green bars; also refer to Chapter 4.2), the first analysis for this experiment (Figure 4.3.3, Experiment 4, blue bars) was to see if the new stimulus definitions changed accuracy performance. An independent-samples t-test was performed, comparing between experiments the hit rates in each condition. The independent-samples t-test found that the conditions did not all respond in the same way to the stimulus re-definition. In AR, the hit rates were significantly lower in this experiment with the new stimulus definitions, than in Experiment 3 ($t_{(26.069)} = 2.563$, $p = 0.016$, mean difference: 0.856% ±0.334%, equal variances not assumed). In VR, the hit rates were significantly *higher* in this experiment with the new stimulus definitions, than in Experiment 3 ($t_{(21.432)} = 2.328$, $p = 0.030$, mean difference: 0.659% ±0.283%, equal variances not assumed). The other conditions were not significantly different in hit rates between the experiments (all $p>0.05$).

Overall, it appears that the hit rates were largely the same between experiments (i.e., stimulus definitions), though perhaps with small differential effects in audition and vision.

The decrease in hit rate only for AR brings the AR condition back in the spotlight though, having been found to be uniquely low on accuracy in my previous experiments. To check if it is currently on par with other non-auditory conditions, multiple paired-samples t-tests were performed, comparing the AR hit rate against each of the other non-auditory conditions. All comparisons against the AR hit rate were significant (p<0.001), meaning the AR hit rate became lower than the other conditions, under this new stimulus definition. As a safety check, multiple paired-samples t-tests were performed to compare the AL hit rate against the hit rates of other non-auditory conditions. The AL hit rate was not significantly different to the VR hit rate or the ARVR hit rate (p>0.05). However, the AL hit rate was significantly lower than the hit rates of VL, ARVL, ALVR, and ALVL (all p<0.05). Hence, it seems like the new stimulus definitions have reduced the hit rates of auditory conditions, with the AR hit below par, and the AL hit rate on par with receding-only conditions.

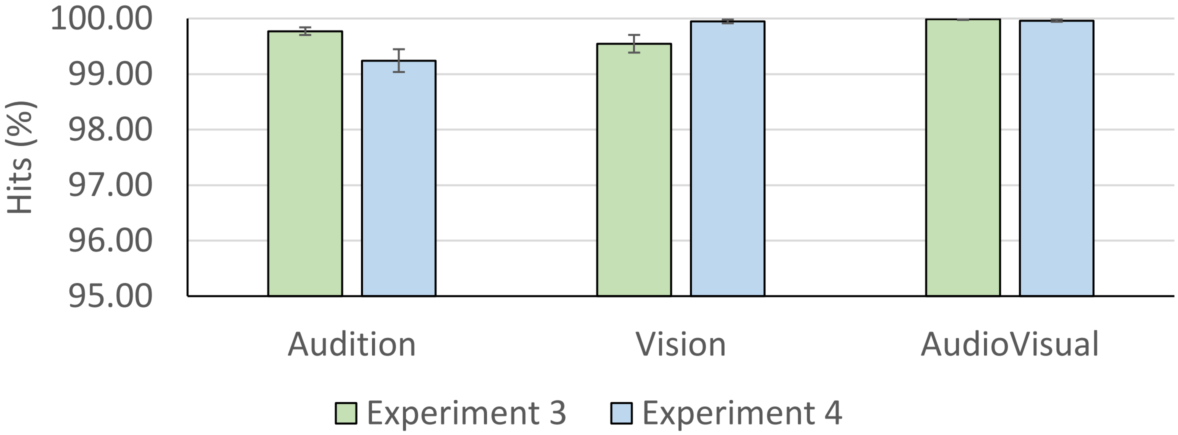


Figure 4.3.4. Hit rates by modality (motion directions averaged within modality, SEM error bars). Experiment 4 (blue bars is the current experiment with re-defined stimuli). Experiment 3 (green bars) uses the original stimulus definitions. The stimulus re-definition produced differential effects: the auditory hit rate was significantly reduced, the visual hit rate was significantly increased.

To check for modality effects on the hit rates, the conditions were collated into sensory modalities, averaging the hit rates across motion directions (Figure 4.3.4, Experiment 4, blue bars).

A one-way ANOVA was performed on the hit rates, with modality as the factor. The one-way ANOVA found a significant main effect of modality ($F_{(1.020, 19.385)} = 14.192$, p = 0.001, $\eta_p^2 = 0.428$, Greenhouse-Geisser corrected). Pairwise comparisons found that the auditory modality had significantly lower hit rates than the visual modality (mean difference: 0.708% ±0.188%, p = 0.004) and the audio-visual modality (mean difference: 0.721% ±0.190%, p = 0.004). The hit rates between the visual and audio-visual modalities were not significantly different to each other (p = 0.924). Thus, it appears that audition is lower than vision and audio-vision, in terms of hit rates.

To compare the modality-hit rates between Experiment 4 (current experiment, stimulus re-defined; Figure 4.3.4, blue bars), and Experiment 3 (original stimulus definitions; Figure 4.3.4, green bars), a mixed-factors ANOVA was performed, with modality as the within-subjects factor, and experiment as the between-subjects factor. A significant interaction between modality and experiment was found ($F_{(1.672, 63.530)} = 10.406$, $p<0.001$, $\eta_p^2 = 0.215$, Greenhouse-Geisser corrected). Pairwise comparisons found that this interaction was driven by the auditory and visual modalities, with significantly lower auditory hit rates in Experiment 4 than Experiment 3 (mean difference: 0.531% ±0.217%, p = 0.019), and significantly *higher* visual hit rates in Experiment 4 than Experiment 3 (mean difference: 0.404% ±0.163%, p = 0.018). The hit rate on the audio-visual modality was not significantly different between experiments. Altogether, on the level of sensory modality, there seems to be differential effects of the new stimulus definition on the hit rate, with reductions in the auditory modality, but improvements in the visual modality.

The re-defined stimulus conditions meant all trials had both auditory and visual stimulation; there was only an audio-visual catch, to examine false alarm rates for. To check if the stimulus re-definition impacted false alarm rates, an independent-samples t-test was performed, comparing the asvs false alarm rate of this experiment (Figure 4.3.5, Experiment 4, blue bar) against its counterpart in Experiment 3 (Figure 4.3.5, green bar). There was no difference in the asvs false alarm rate between experiments ($t_{(31.424)} = 0.116$, p = 0.909, mean difference: 0.435% ±0.376%, equal variances not assumed). Thus, it appears that the re-definition of the stimulus conditions did not impact the good false alarm rate achieved in Experiment 3.

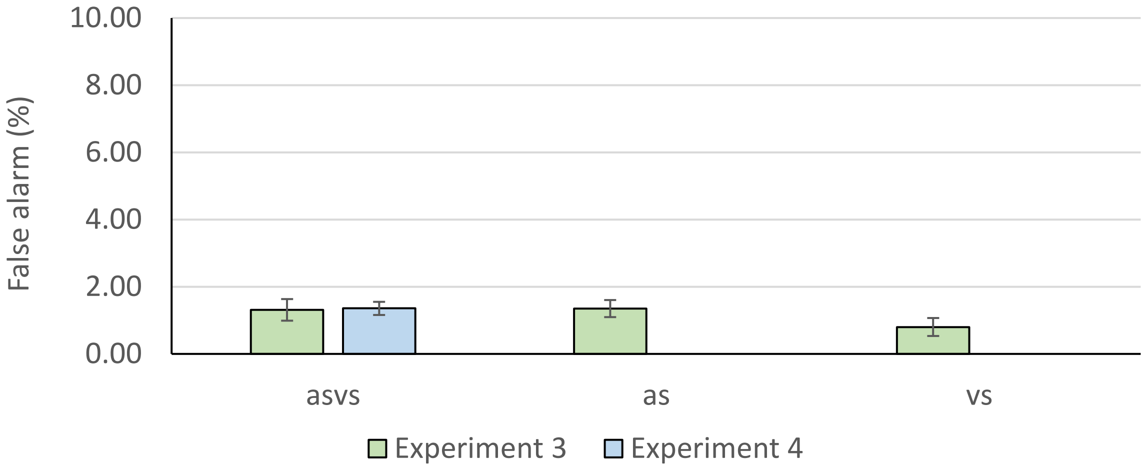


Figure 4.3.5. Participant-averaged false alarm rates (SEM error bars). The current experiment with re-defined stimulus conditions is Experiment 4 (blue bar), while Experiment 3 (green bars) used the original stimulus definitions. Note, Experiment 4 only had audio-visual catch trials (asvs); the false alarm rates for asvs are similar between experiments.

Overall, the stimulus re-definition for this experiment seems to have had quite ambivalent effects on the accuracy performance, compared to Experiment 3 which found near-perfect accuracy using the original stimulus definitions. On a general level, the hit rates in each condition was largely the same between experiments, and the false alarm rates in the same catch condition were statistically identical between experiments. On a smaller level, it was interesting to see that the stimulus re-definition slightly reduced auditory hit rates, yet also slightly improved visual hit rates. Given that Experiment 3 with the original stimulus definitions had near-perfect accuracy, the finding of largely similar hit rates and false alarm rates in Experiment 4 suggests that the stimulus re-definition was not detrimental to the accuracy performance, at least not dramatically. In fact, for the current experiment, the overall hit rate was 99.78% ±0.15%, while the overall correct (proportion of hits and correct rejections) was 99.40% ±0.09%.

Response time performance

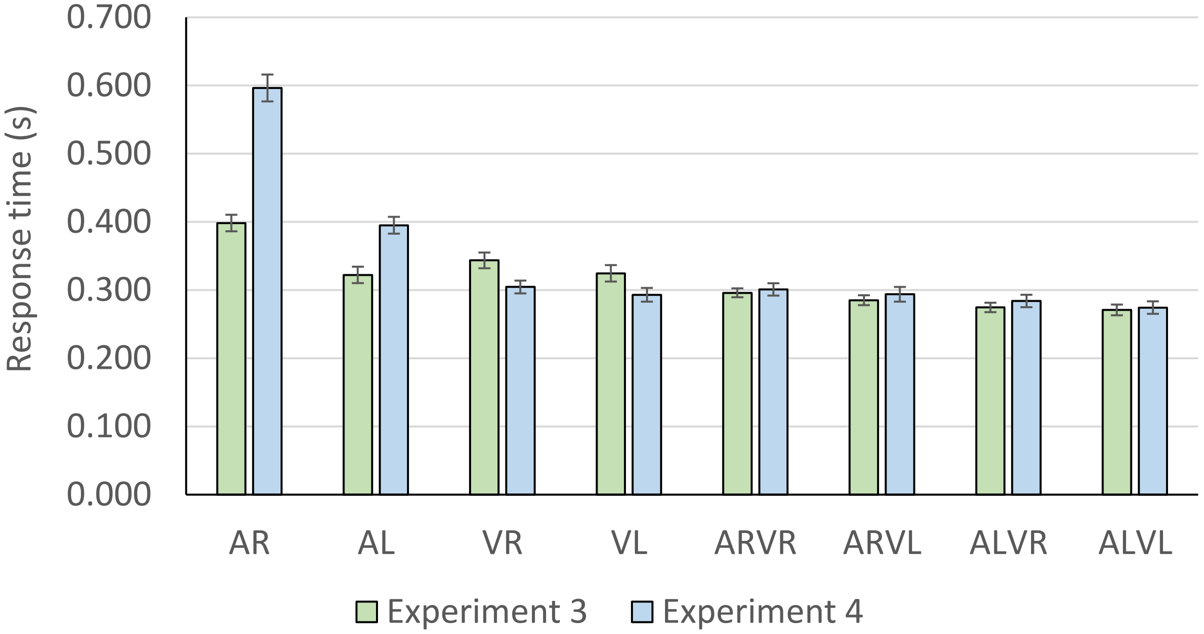


Figure 4.3.6. Participant-averaged RTs across all conditions (SEM error bars). The current experiment with the re-defined stimulus conditions is Experiment 4 (blue bars), Experiment 3 (green bars) used the original stimulus definitions. Comparing Experiment 4 to Experiment 3, the auditory RTs were slower, but the visual RTs were faster.

Visually, the stimulus re-definition seems to have slowed the auditory RTs, especially for the AR condition, although the visual conditions seems to have sped up, all compared to Experiment 3 (Figure 4.3.6; Experiment 4, blue bars, versus Experiment 3, green bars). To check on such apparent differences, an independent-samples t-test was performed, comparing RTs between the two experiments, for each condition. The independent-samples t-test revealed that the stimulus re-definition had differential effects on RTs. The auditory conditions had slower RTs with the re-defined stimulus definition (Figure 4.3.6, Experiment 4, blue bars) compared to the original stimulus definition (Figure 4.3.6, Experiment 3, green bars). The RTs of the re-defined AR condition was significantly slower than to the original AR condition ($t_{(31.631)} = 8.559$, $p<0.001$, mean difference: 0.198 s ±0.023 s, equal variances not assumed). Also, the RTs of the re-defined AL condition was significantly slower than to the original AL condition ($t_{(38)} = 4.218$, $p<0.001$, mean difference: 0.073 s ±0.017 s). However, the visual conditions seemed to have faster RTs with the re-defined stimulus conditions (Figure 4.3.6, Experiment 4, blue bars) than with the original stimulus conditions (Figure 4.3.6, Experiment 3, green bars). The RTs for the re-defined VR condition was significantly faster than for the original VR condition ($t_{(38)} = 2.622$, $p = 0.013$, mean difference: 0.039 s ±0.015 s). The RTs for the re-defined VL condition was close to being significantly faster than the original VL condition, ($t_{(39)} = 2.009$, $p = 0.052$, mean difference: 0.032 s ±0.016 s). Audio-visual RTs were not significantly different between experiments. The stimulus re-definition appears to have had differential effects on the unisensory RTs.

Checking for unisensory looming biases (Figure 4.3.6, Experiment 4, blue bars), paired-samples t-tests were performed, separately for audition and vision, comparing RTs between looming and receding. The RTs towards AL were significantly faster than to AR ($t_{(19)} = 16.096$, $p<0.001$, mean difference: 0.201 s ±0.013 s). The RTs towards VL were significantly faster than to VR ($t_{(19)} = 6.975$, $p<0.001$, mean difference: 0.012 s ± 0.002 s). Hence, with re-defined stimulus conditions, there was still a looming bias: faster responses for looming than receding.

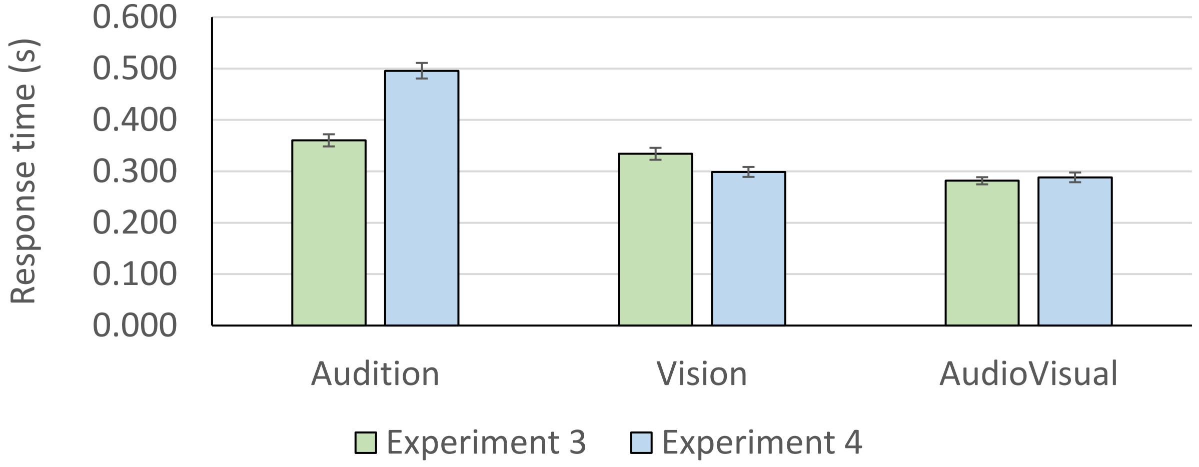


Figure 4.3.7. RTs on a modality level (motion directions averaged within modality, SEM error bars). Experiment 4 (blue bars) uses re-defined stimuli, Experiment 3 (green bars) uses the original stimulus definitions. There was a significant modality effect, with the fastest RTs towards the audio-visual conditions.

To check for modality effects on the RT, the RTs were grouped by sensory modality, averaging the motion-directions within the modality (Figure 4.3.7, Experiment 4, blue bars). A one-way ANOVA was performed on the RTs, with sensory modality as the factor. The ANOVA found a significant effect of modality ($F_{(1.039,\ 19.738)} = 492.553$, $p<0.001$, $\eta_p^2 = 0.963$, Greenhouse-Geisser corrected). Pairwise comparisons revealed that the RTs towards audio-visual conditions were significantly faster than the RTs towards visual conditions (mean difference: 0.011 s ±0.002 s, $p<0.001$) and the RTs towards auditory conditions (mean difference: 0.207 s ±0.09 s, $p<0.001$). RTs towards visual conditions were also significantly faster than RTs towards auditory conditions (mean difference: 0.197 s ±0.009 s, $p<0.001$). Hence, there were clear differences between the three modalities, in terms of RT, with the audio-visual conditions being faster than the unisensory conditions, and the auditory condition being the slowest. The finding of fastest RTs towards audio-visual condition was also found by Cappe et al. (2009).

To compare the modality-based RTs between this experiment with re-defined stimuli (Figure 4.3.7, Experiment 4, blue bars), and Experiment 3 with the original stimulus definitions (Figure 4.3.7, green bars), a mixed-factors ANOVA was performed on the RTs, with modality as the within-subjects factor, and experiment as the between-subjects factor. Crucially, a significant interaction between the factors modality and experiment was found ($F_{(1.577,\ 59.935)} = 149.381$, $p<0.001$, $\eta_p^2 = 0.797$, Greenhouse-Geisser corrected). Pairwise comparisons in this interaction found Experiment 4 had significantly slower auditory RTs (mean difference: 0.135 s ±0.019 s, $p<0.001$), and significantly faster visual RTs (mean difference: 0.035 s ±0.015 s, $p = 0.026$), compared to Experiment 3. This finding on the modality level matches, and reinforces the above findings on the condition level: the stimulus re-definition produced slower auditory RTs, but faster visual RTs.

Next, the RTs in the audio-visual conditions were examined (Figure 4.3.8). Visually, the RTs to ALVL were fastest, and the condition with the next fastest RTs was ALVR. To check if RTs towards the ALVL condition were particularly fast, a paired-samples t-test was performed, comparing RTs of ALVL and ALVR. The paired-samples t-test revealed that RTs towards ALVL were significantly faster than towards ALVR ($t_{(19)} = 4.019$, $p<0.001$, mean difference: 0.010 s ±0.002 s). The finding that RTs towards ALVL were fastest among the audio-visual conditions was also found by Cappe et al. (2009).

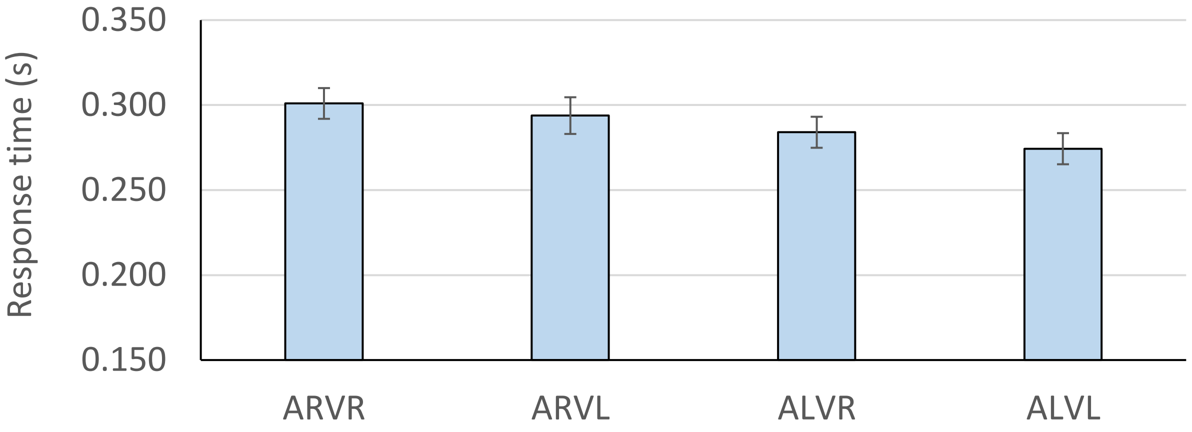


Figure 4.3.8. Participant-averaged RTs on the audio-visual condition (SEM error bars). RTs towards the ALVL condition were significantly faster than RTs towards other conditions.

To explore if there were unisensory or multisensory effects on the audio-visual RTs (Figure 4.3.8), a 2x2 ANOVA was conducted, with the factors auditory motion direction (AR, AL) and visual motion direction (VR, VL). There was no interaction between the auditory and visual factors ($F_{(1, 19)} = 0.534$, $p = 0.474$, $\eta_p^2 = 0.027$). There was a significant main effect of auditory motion direction ($F_{(1,19)} = 94.990$, $p<0.001$, $\eta_p^2 = 0.833$). Pairwise comparisons within the factor of auditory motion direction found RTs towards conditions with AL were faster than to conditions with AR (mean difference: 0.018 s ±0.002 s, $p<0.001$). There was also a significant main effect of visual motion direction ($F_{(1, 19)} = 16.561$, $p<0.001$, $\eta_p^2 = 0.466$). Pairwise comparisons within the factor of visual motion direction found conditions with VL were significantly faster than conditions with VR (mean difference: 0.008 s ±0.002 s, $p<0.001$). Hence, in the RTs, there was no evidence to suggest a looming-specific audio-visual interaction, as would be the case for Cappe et al.'s (2009) claim of selective integration towards audio-visual looming signals.

To summarise the RT findings, stimulus re-definition seems to have produced differential unisensory effects. Auditory RTs were slower, but visual RTs were faster, compared to Experiment 3 which used the original stimulus definitions. Despite the stimulus re-definition, and the removal of onset transients, this experiment reproduced the Cappe et al. (2009) of faster RTs in audio-visual conditions than unisensory conditions, and the fastest RTs were in the ALVL condition. Yet, no evidence was found on the RT level to support Cappe et al.'s (2009) claim of selective integration towards ALVL.

Redundant signals effect

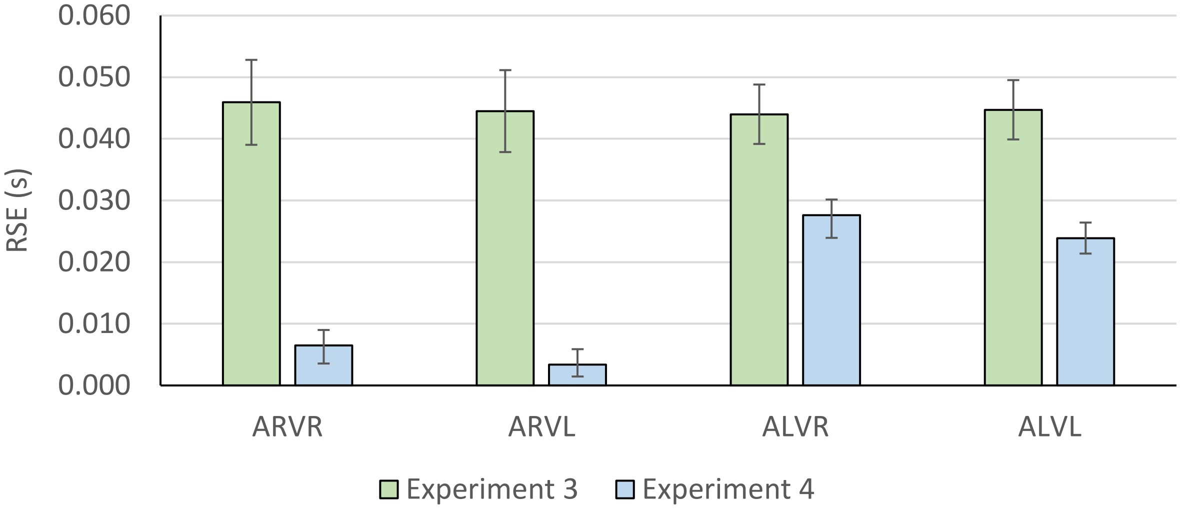


Figure 4.3.9. Participant-averaged RSEs (SEM error bars). The current experiment with re-defined stimuli is Experiment 4, blue bars. Experiment 3 used the original stimulus definition, green bars. With re-defined stimuli, RSEs became smaller across all conditions, and its size was determined by the auditory component: larger RSEs when the auditory component was AL than AR.

The multisensory benefit should be defined by the RSE (see also Chapter 2.6.5 *Redundant signals effect*). Experiment 3 with the original stimulus definitions found large and equal RSEs across the conditions (Figure 4.3.9, green bars). The current experiment re-defined the stimuli (Figure 4.3.9, Experiment 4, blue bars). To check if stimulus re-definition affected the RSEs, an independent-samples t-test was used to compare each condition's RSE between experiments. The independent-samples t-test revealed that all conditions' RSEs were significantly smaller in Experiment 4 with re-defined stimuli, than in Experiment 3 with the original stimulus definition. Specifically, for ARVR ($t_{(25.621)}$ = 5.261, p<0.001, mean difference: 0.039 s ±0.007 s, equal variances not assumed), for ARVL ($t_{(22.162)}$ = 5.955, p<0.001, mean difference: 0.041 s ±0.007 s, equal variances not assumed), for ALVR ($t_{(38)}$ = 2.686, p= 0.011, mean difference: 0.016 s ±0.006 s), for ALVL ($t_{(28.534)}$ = 3.842, p<0.001, mean difference: 0.021 s ±0.005 s, equal variances not assumed). Thus, stimulus re-definition seems to have reduced the RSEs across all conditions.

Next, checking for unisensory or multisensory effects in the RSE, a 2x2 ANOVA was performed on the current experiment's RSEs (Figure 4.3.9, Experiment 4, blue bars). The factors were auditory motion direction (AR, AL) and visual motion direction (VR, VL). Crucially, no interaction between the auditory motion direction and visual motion direction was found ($F_{(1, 19)}$ = 0.031, p = 0.862, η_p^2 = 0.002). No main effect of visual motion direction was found ($F_{(1, 19)}$ = 1.285, p = 0.271. η_p^2 = 0.063). Only a main effect of auditory motion direction was found ($F_{(1,19)}$ = 99.566, p<0.001, η_p^2 = 0.840). Pairwise comparisons on the auditory motion direction found that conditions with an AL component had larger RSEs than conditions with an AR component (mean difference: 0.021 s ±0.002 s, p<0.001). Hence, the size of the RSE was dependent only on the auditory component, with larger RSEs when the auditory component had a looming signal; this auditory main effect can be seen in Figure 4.3.9 (blue bars), where the ARVR and ARVL RSEs are smaller than the ALVR and ALVL RSEs. Finally, the RSE of ALVR was larger than the RSE of ALVL, but a paired-samples t-test revealed that the difference was not significant ($t_{(19)}$ = 1.022, p = 0.320, mean difference: 0.004 s ±0.004 s).

To summarise, stimulus re-definition seems to have reduced the size of the RSEs, and introduced a unisensory effect: the motion direction of the auditory signal. Audio-visual conditions with an AR component have smaller RSEs than audio-visual conditions with an AL component. Also, the RSE towards ALVL was not largest. Note that RTs in the AR condition were particularly slow (see previous section, Figure 4.3.6). The link between cross-modal RTs and the RSE size will be explained later in this chapter (cf. principle of congruent effectiveness; Otto et al., 2013).

4.3.3 Discussion

In Experiments 1 and 2, there were issues with the response accuracy and speed, which might stem from a design flaw in the stimulus: the concurrence of stimulus and motion onsets, or onset transient. Hence, Experiment 3 removed the onset transients, and found near-perfect response accuracy and fast response times. The data of Experiment 3 seemed perfect, however, in removing the onset transients (by introducing a random foreperiod with stimulation), the participant may have been cued to the modality of the motion signal, which may then unduly elevate response accuracy and speed. Hence, Experiment 4 was conducted, which had onset transients removed, but the stimulus conditions were also re-defined such that all conditions were audio-visual, with sensory modality only referring to the motion signal's modality. This re-definition meant that for all conditions, the foreperiod always features the same audio-visual stimulation (40 dB SPL 1000 Hz tone, 7° circular object), thereby removing potential modality priming effects during the foreperiod. Taking that Experiment 3 had near-perfect response accuracy and speed, the first question became: had stimulus re-definition negatively impacted response accuracy and speed?

In short, the answer is mixed. In terms of response accuracy, stimulus re-definition seems not to have produced dramatic reductions in hit rates. In most conditions, the hit rate was largely the same between Experiments 4 and 3; the false alarm rate was also statistically the same between the two experiments. The only conditions which were impacted by stimulus re-definition were the auditory receding (AR), and the visual receding (VR), albeit differentially. Curiously, with stimulus re-definition, the AR hit rate decreased, but the VR hit rate increased, compared to the original stimulus definition. These rather localised changes on AR and VR was however sufficient to change the modality-level dynamics, such that the auditory modality had the worst hit rates, while the visual modality was elevated to hit rates on par with the audio-visual modality.

The same pattern of ‘worsened AR and auditory modality, improved VR and visual modality’ was found with the response times. With the re-defined stimuli, RTs towards AR and AL were significantly slower than with the original stimulus definition. Yet, the RTs towards VR and (marginally) VL were faster with re-defined stimuli than with the original stimuli. Critically, with re-defined stimuli, RTs towards the audio-visual conditions were still fastest, and the RTs were fastest towards the ALVL condition. Both findings were also found by Cappe et al. (2009) and formed the basis of their claim of selective integration towards ALVL signals. However, on neither the level of RT or RSE were there multisensory interactions, such as a looming-specific audio-visual interaction, which would be a trait of selective integration towards ALVL signals. The speed of the RTs and the size of the RSEs were only determined by the auditory component of the audio-visual signal. Specifically, RTs were faster and the RSEs were larger, if the audio-visual signal contained an AL component rather than an AR component.

Ultimately, the question is if this experiment represents the best behavioural data that can be applied to more advanced analysis techniques, such as the comparative approach (e.g., Innes & Otto, 2019), and the computational cognitive modelling of Chapter 5. Experiment 3 with only the onset transients removed was perfect from a data perspective, but was methodologically compromised with its stimuli potentially priming the participant. The critical question is whether in fixing the potential priming by re-defining the stimulus conditions, the response accuracy and speed decreased meaningfully. It was not ideal that in re-defining the stimulus conditions, the auditory conditions decreased in both accuracy and speed, but there was also an increase in response accuracy and speed in the visual conditions. Importantly, the accuracy on an overall level was excellent, with an overall hit rate of 99.78% ±0.15%, and an overall correct value (proportion of hits and correct rejections) of 99.40% ±0.09%. For reference, Experiment 3, with the original stimulus definitions, was actually slightly worse, at an overall hit rate of 98.82% ±0.09%, and hardly better with an overall correct value of 99.48% ±0.09%. Hence, as the experiment which addresses the stimulus design flaws, and achieves high accuracy, this experiment was the one which was used for further analysis, in the comparative approach (Chapter 4.4), and the computational cognitive modelling (Chapter 5).

4.4 The comparative approach

4.4.1 Comparing empirical RSEs against RSEs predicted by a simple probability summation rule

Now with an appropriate dataset (from Experiment 4), this chapter returns to the question posed at the beginning of the chapter: can the race mechanism be a framework that offers insights into the multisensory processing of audio-visual looming signals? Here, an attempt to answer this question is made using the comparative approach, whereby the empirical RSEs are compared to the predictions made by probability summation, the statistical facilitation underlying the race model architecture.

Recall from Chapter 4.1 that probability summation is mathematically described:

$P_{AV}(T_{A∪V} \leq t) = P_{AV}(T_A \leq t) + P_{AV}(T_V \leq t) – P_{AV}(T_{A∩V} \leq t)$ (Otto & Mamassian, 2017)

If two assumptions are made (statistical independence and context invariance, i.e., auditory and visual processes unrelated to each other), then the unknown audio-visual intersection term becomes the product of unisensory probabilities – the independent race model (Raab, 1962):

$P_{AV}(T_{A∪V} \leq t) = P_{AV}(T_A \leq t) + P_{AV}(T_V \leq t) – P_{AV}(T_A \leq t)$ x $P_{AV}(T_V \leq t)$ (Otto & Mamassian, 2017)

Hence, using the independent race model (Raab, 1962), only the auditory and visual performances are needed, from which no further parameters are required to make a probability summation-based prediction of audio-visual performance. Thus, from the empirical RT dataset of Experiment 4, this simple probability summation rule is applied to the unisensory RTs to produce a race model prediction of the RSEs, e.g., taking the RTs in the AL and VL conditions to predict the RSE in the ALVL condition, and so forth for the three other audio-visual conditions. If there is good correspondence between the empirical data and the predictions made by probability summation, then it would be evidence for the potential of the race model architecture in explaining multisensory processing of motion-in-depth.

Technique

In Chapter 2.6.5 *Redundant signals effect*, geometric quantification of the RSE was explained. The same geometric quantification processes are also applied in calculating the predicted RSEs.

Geometric RSE quantification can be visualised (Figure 4.4.1) as the area between the lower bound of race model performance (Grice's bound (Grice et al., 1984), effectively the fastest performance of the two unisensory CDFs at each quantile, red line), and the CDF of the audio-visual performance – empirical (solid line) or predicted (dotted line). Figure 4.4.1, using artificially generated data from known parameters, shows the aforementioned probability summation-predicted RSE in grey shading, while the 'empirical' RSE is faster and therefore is larger than predicted by simple probability summation.

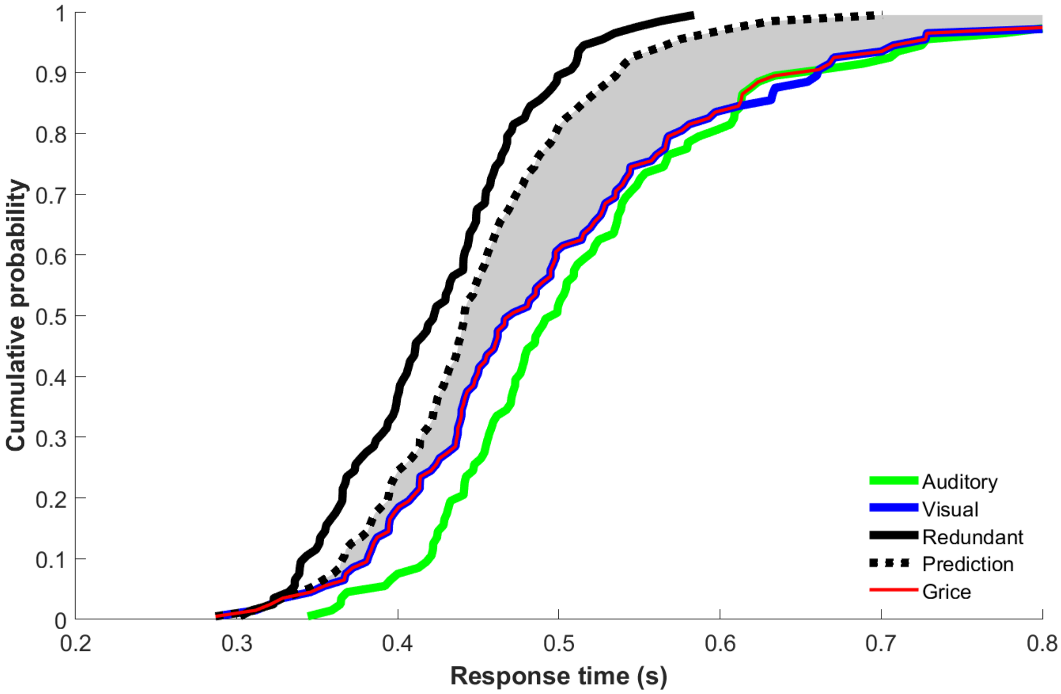


Figure 4.4.1. Graphical representation of the geometric quantification of predicted RSE, using artificially generated data. From the auditory (green line) and visual (blue line) performance, a parameter-free prediction of multisensory performance is made (dotted black line), using a simple probability summation rule which assumes statistical independence and context invariance (independent race model; Raab, 1962). The RSE predicted by the independent race model (Raab, 1962) is the area enclosed (shaded grey) by the predicted multisensory performance and Grice's bound (red line; Grice et al., 1984). In this example, the empirical multisensory performance (solid black line) is faster than predicted by simple probability summation, hence the empirical RSE would also be larger than the predicted RSE.

In practice, as with the geometric quantification of empirical RSEs (Chapter 2.6.5), the above graphical visualisation is not necessary to geometrically quantify the predicted RSEs. Taking the Experiment 4 dataset, first, for each participant, their RTs in the four unisensory conditions (100 trials presented in each condition) were sampled down to 50 RT quantiles using linear interpolation, for each condition. The down-sampling was performed to equalise the number of datapoints between conditions, and is a requirement for the mathematical shortcuts in calculating the geometric RSEs (i.e., operations at each quantile; see also Chapter 2.6.5). Next, the 50 RT quantiles of an auditory condition was paired with the 50 RT quantiles of a visual condition, and fed into the *getRaab* Matlab function of the RSE-box (Otto, 2019), which automatically applies probability summation to the unisensory data inputs, and produces a CDF prediction of multisensory performance. This process was repeated four times in total, as it was the 2x2 combination of auditory conditions (AR, AL) and visual conditions (VR, VL), to produce performance predictions of the audio-visual conditions ARVR, ARVL, ALVR, ALVL. Finally, to obtain the predicted RSE of each audio-visual condition, the difference between the CDF of the predicted audio-visual condition, and its Grice bound, is taken using the *getGain* Matlab function from the RSE-box (Otto, 2019). This difference between the probability summation-predicted audio-visual CDF and the Grice bound is the predicted RSE. In comparison, the empirical RSE was obtained by applying the geometric technique on the empirical audio-visual performances (see also Chapter 2.6.5).

Results

For the first step in analysing the variances in predicted and empirical RSEs (Figure 4.4.2), a 2x2x2 ANOVA was performed, with factors auditory motion direction (AR, AL), visual motion direction (VR, VL) and data type (predicted, empirical). There was a main effect of auditory motion direction ($F_{(1,19)} = 87.1$, $p<0.001$, $\eta_p^2 = 0.821$). Pairwise comparisons within this auditory main effect revealed that the RSE was significantly larger for audio-visual conditions with an AL component than with an AR component (mean difference: 0.016 s ±0.002 s). There was no main effect of visual motion direction, nor any interaction involving vision. Additionally, there was a main effect of data type ($F_{(1,19)} = 29.5$, $p<0.001$,$\eta_p^2 = 0.608$). Pairwise comparisons on this main effect revealed that the empirical RSEs were larger than the RSEs predicted by probability summation (mean difference: 0.008 s ±0.001 s). The finding of larger RSEs than predicted suggested that there was more going on in the processing of audio-visual signals than simple probability summation.

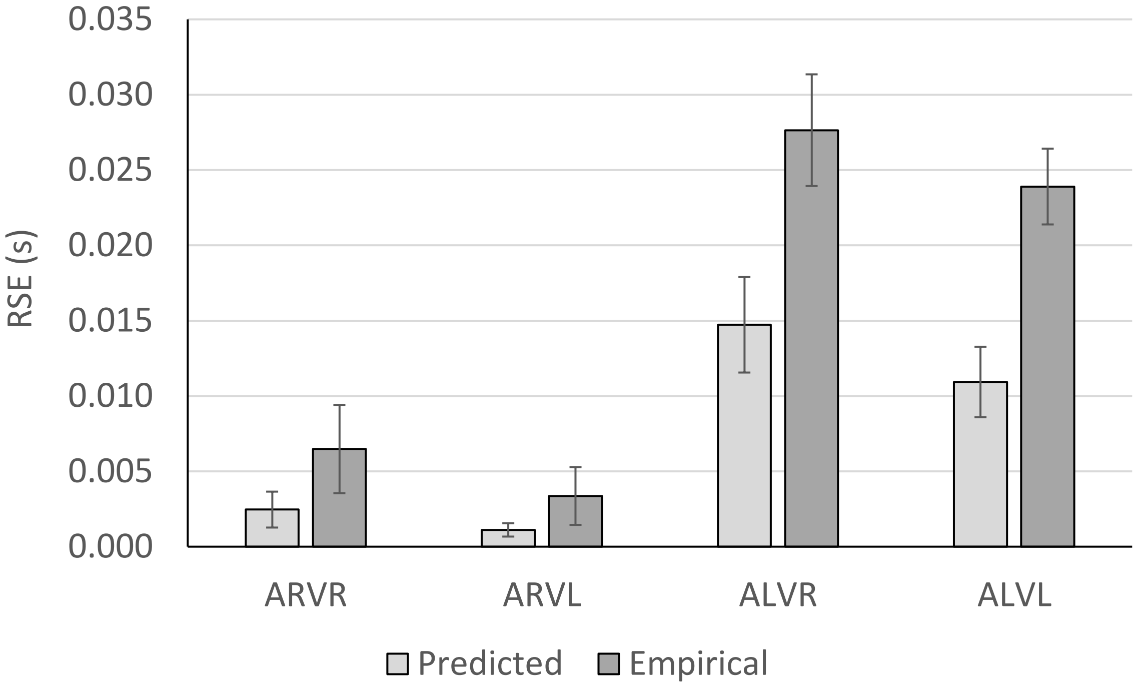


Figure 4.4.2. Predicted (light grey bars) and empirical (dark grey bars) RSEs for each audio-visual condition, averaged across participants (SEM error bars). Notice that although the RSE predictions by probability summation underestimated the size of the empirical RSEs, the predictions closely follow the general pattern of the empirical RSEs, e.g., smaller RSEs predicted when the empirical RSEs were also small.

There was also a significant interaction between the factors auditory motion direction and data type ($F_{(1,19)}$ = 18.0, p<0.001, η_p^2 = 0.486). To understand this interaction, a further analysis on the simple effects was made. Regarding data type, the empirical RSEs were significantly larger than the predicted RSEs, for both AR ($t_{(19)}$ = 2.29, p = 0.034, mean difference: 0.003 s ±0.001 s) and AL conditions ($t_{(19)}$ = 5.76, p<0.001, mean difference: 0.013 s ±0.002 s). Regarding auditory motion direction, the RSEs in conditions with an AL component were larger than RSEs in conditions with an AR component, both in the predicted RSEs ($t_{(19)}$ = 5.45, p<0.001, mean difference: 0.011 s ±0.002 s) and the empirical RSEs ($t_{(19)}$ = 9.98, p<0.001, mean difference: 0.021 s ±0.002 s). Altogether, both main effects (auditory motion direction, data type) were also found on the level of simple effects in the interaction. The interaction effect can be understood in that the difference between empirical and predicted RSE is larger for conditions involving auditory looming compared to auditory receding signals.

Finally, the RSEs towards the ALVR condition appeared to be larger than the RSEs towards the ALVL condition, in both empirical and predicted data. To check if this difference reached significance, paired-samples t-tests were performed. The paired-samples t-test on the empirical data found that ALVR RSE was not significantly larger than the ALVL RSE ($t_{(19)}$ = 1.022, p = 0.320, mean difference: 0.004 s ±0.017 s). The paired-samples t-test on the predicted data found that the ALVR RSE was significantly larger than the ALVL RSE ($t_{(19)}$ = 2.656, p = 0.016, mean difference: 0.004 s ±0.002 s). Although the predicted and empirical data differ on whether the ALVR RSE was significantly larger than the ALVL RSE, note that the mean difference was the same between the predicted and empirical data; it was the larger variability on the empirical data that seems to have prevented a significant difference in favour of the ALVR RSE.

Summary

In examining the RSEs, empirical and predicted by a simple probability summation rule, a major discovery was made: the size of the RSE was modulated only by the unisensory factor of auditory motion direction. The RSE would be large if the audio-visual condition had an auditory looming component, rather than an auditory receding component. Crucially, there was no interaction effect between the auditory and visual factors in the size of the RSEs. Hence there was not at all a looming-specific audio-visual interaction, that would be a trait of a potential multisensory looming bias. These findings do not support the proposal of a selective integration mechanism towards congruent audio-visual looming signals (Cappe et al., 2009).

First, it must be pointed out that in both empirical and predicted RSEs, the condition of congruent audio-visual looming, ALVL, did not have the largest redundancy gain. In fact, numerically, the RSE was larger in the ALVR condition than in the ALVL condition, and this numerical difference was the same in the empirical data and in prediction (although only the predicted data reached significance, due to lower variability in the prediction than in the empirical data). If there is indeed multisensory processing that is special towards congruent audio-visual looming, then the RSEs should be particularly large in the ALVL condition. My data did not find such an absolute advantage in the RSEs towards the ALVL condition, so there was no support of there being a process ‘selective’ towards processing audio-visual looming signals.

Second, there were two indications from the RSE patterns that a race mechanism could be a suitable explanatory framework for the processing of redundant audio-visual motion-in-depth signals. As the first indication, it must be noted just how closely the predictions by simple probability summation follows the empirical data. This can be seen visually from comparison of predicted and empirical RSEs across all four audio-visual conditions (Figure 4.4.2). Moreover, in the statistical analyses, both empirical and predicted RSEs had the auditory main effect of larger RSEs with an auditory looming component than with an auditory receding component. The closeness of the empirical RSEs and predicted RSEs was also shown in their significant correlation with each other (r = 0.662, p = 0.001; considering individual data points of all four audio-visual combinations). This correspondence between empirical and predicted RSEs, although not perfect, shows a strong explanatory power of probability summation and race models.

As the second indication that the race mechanism could be a suitable explanatory framework, the auditory main effect of larger RSEs with an AL component than with an AR component is in fact predicted by probability summation, based on the unisensory RTs of AL and AR. From the perspective of the race architecture, the RSE is a race between unisensory processes. To produce a redundancy gain in the race architecture, the unisensory units need to be competitive, that is, they both have a chance of ‘winning’ the race to decision. If there is a unisensory unit that always ‘wins’ the race, by being quickest to reach its decision criterion, then decisions are based on that modality alone, and no redundancy gain is predicted by probability summation and the race architecture. Hence, the implication of probability summation is the principle of congruent effectiveness (Otto et al., 2013), which states that a redundancy gain (RSE) is only possible if both sensory units are similar in performance. Now, recall that in the results section of this experiment (see Chapter 4.3.2, RT performance, and also Figure 4.3.6), the RTs towards the unisensory AR condition (0.596 s ±0.020 s) were notably worse than the RTs towards the unisensory AL condition (0.395 s ±0.012 s), a difference of about 0.2 seconds. Meanwhile, the visual conditions had still faster RTs, with little difference between VR (0.305 s ±0.009s) and VL (0.293 s ± 0.010 s). Practically, then, the RT difference between AR and visual conditions was about 0.3 seconds, while the RT difference between AL and visual conditions was only about 0.1 seconds. From the perspective of probability summation and the race architecture, it is not at all surprising then that the audio-visual conditions with AR had smaller RSEs. The auditory processing of the AR signal is simply too slow in relation to the visual processing, that there cannot be much of an RSE. Whereas the auditory processing of the AL signal is closer to the visual performance, so there is a larger RSE. The prediction from probability summation and the race architecture matches with what was empirically obtained, so there is strong evidence to suggest that probability summation and the race architecture has strong explanatory power in audio-visual processing.

However, it must be noted that the size of the empirical RSEs were consistently and significantly larger than the RSEs predicted by a simple probability summation rule (independent race model; Raab, 1962). Is this therefore evidence against the probability summation and the race architecture? As a reminder, there are two assumptions of the independent race model (Raab, 1962), neither of which may be correct (Otto & Mamassian, 2017). First, the RT on a given trial is not affected by the type of stimulation or the performance on the previous trial (assumption of statistical independence; Ashby & Townsend, 1986; Colonius, 1990; Luce, 1986; Miller, 1982; Otto & Mamassian, 2017). Second, processing of a signal is the same whether the signal is presented alone or with concurrent stimulation in another modality, because the modalities do not interfere with each other (assumption of context invariance; Ashby & Townsend, 1986; Liu & Otto, 2020; Luce, 1986; Otto & Mamassian, 2017; Townsend et al., 2020; Townsend & Wenger, 2004). Thus, the underestimates of the probability summation predictions of RSE may not be an issue of the race architecture; the problem may lie in the two assumptions that were used to make the predictions. Hence, the next section tests for the first assumption, the assumption of statistical independence.

4.4.2 Testing the assumption of statistical independence

To test the assumption of statistical independence (the performance on the current trial is unaffected by the previous trial), history effects were examined. Typically, in sensory experiments, trials of different modalities are randomly interleaved, such that for example an auditory trial can sometimes be preceded by a visual trial, and other times an auditory trial can be preceded by an auditory trial (Shaw et al., 2020). The experiments of my project also randomly interleaved trials of different modalities. Hence, a prime example of history effects would be if the RT after a modality switch is different to the RT after the same modality repeated. Such a difference in RT is the modality switch cost, and if there is a significant switch cost, then it means the assumption of statistical independence is incorrect, and there is scope for using the race architecture in understanding audio-visual processing of motion-in-depth signals.

Technique

The modality switch cost is the RT slowdown on the current trial when the previous trial was of a different modality (switch), compared to if the previous trial was the same modality as the current trial (repeat). To calculate the modality switch cost, first, for each participant, the mean RT on auditory trials preceded by a visual trial was calculated (V → A; RT_{VA}), then the mean RT on auditory trials preceded by an auditory trial (A → A; RT_{AA}). Thus, the auditory switch cost is the difference between RT_{VA} and RT_{AA}. Next, for each participant, the mean RT on visual trials preceded by an auditory trial was calculated (A → V; RT_{AV}), then the mean RT on visual trials preceded by a visual trial (V → V; RT_{VV}). Thus, the visual switch cost is the difference between RT_{AV} and RT_{VV}. The average between the auditory switch cost and visual switch cost was taken per participant, to arrive at a combined switch cost, which was then finally averaged across participants into a single value to summarise the history effects of this experiment.

Results

A modality switch cost was found (0.010 s ±0.004 s). A one-sample t-test showed that this modality switch cost was significantly different from zero ($t_{(19)} = 2.191$, $p = 0.041$).

Summary

While the costs found in this experiment was small compared to previous RSE studies (Gondan et al., 2004; Innes & Otto, 2019; Miller, 1982; Otto & Mamassian, 2012; Shaw et al., 2020), the finding of a significant modality switch cost here nonetheless illustrates that the modality of the previous trial does affect performance on the current trial. More specifically, if there is a modality switch (the previous trial's modality is not the same as the current trial's modality), then RTs on the current trial are slower than if the modality was repeated across the two trials. Crucially the finding of significant modality switch costs shows that the RTs in the unisensory conditions are not statistically independent. Hence, the assumption of statistical independence is incorrect. The underprediction of the RSE sizes in Chapter 4.4.1 may therefore be attributable to the usage of the statistical independence assumption in the predictions, rather than any fundamental issues of the probability summation or the race architecture.

4.4.3 Applying Miller’s (1982) test to the data

The assumption of statistical independence is incorrect (see previous section), and this was a realisation that was embodied in Miller’s (1982) landmark test of race architecture performance. Miller’s (1982) test replaces the assumption of statistical independence with a maximal negative correlation (Colonius, 1990), thereby setting the largest possible RSEs predicted by the race architecture with negatively correlated RTs, in effect an upper bound of performance predicted by the race architecture. If the RTs in redundant conditions exceed any part of this bound (termed ‘violation’), then the implication is that the multisensory processing is outside of a race process with correlated RTs. Often, violations of Miller’s (1982) bound are taken as an indication of integrative processing (e.g., Cappe et al., 2009; see also Gondan & Minakata, 2016). Hence, the results of my dataset in Miller’s (1982) test are critical for the determination of whether a race or integrative process best characterises the processing of audio-visual looming signals.

Technique

Miller’s (1982) test is defined by an inequality, stating that the probability for a multisensory decision is less than or equal to the sum of each unisensory decision’s probability (see also Chapter 4.1.3 *Testing the race architecture*). In effect, Miller’s (1982) upper bound of race architecture performance with correlated RTs can be calculated from the unisensory performances. To calculate Miller’s (1982) bound, the RTs of the auditory and visual components (100 trials presented per condition per participant) were taken and down-sampled to 50 RT quantiles each, and the inequality was applied to calculate the bound for a given auditory and visual combination. The violation is simply the difference between the CDF of the corresponding audio-visual condition’s RTs, and its Miller’s (1982) bound, i.e., if the performance in the audio-visual condition is faster than the bound, then there would be a violation. This quantification of the violation is again the geometric technique, which was also used for RSE quantification (see Chapter 2.6.5 *Redundant signals effect* for the analogous geometric quantification of the RSE; see also Figure 4.4.3 for a graphical representation). The quantification of the violations were automated using the Matlab function *getViolation* from the RSE-box (Otto, 2019).

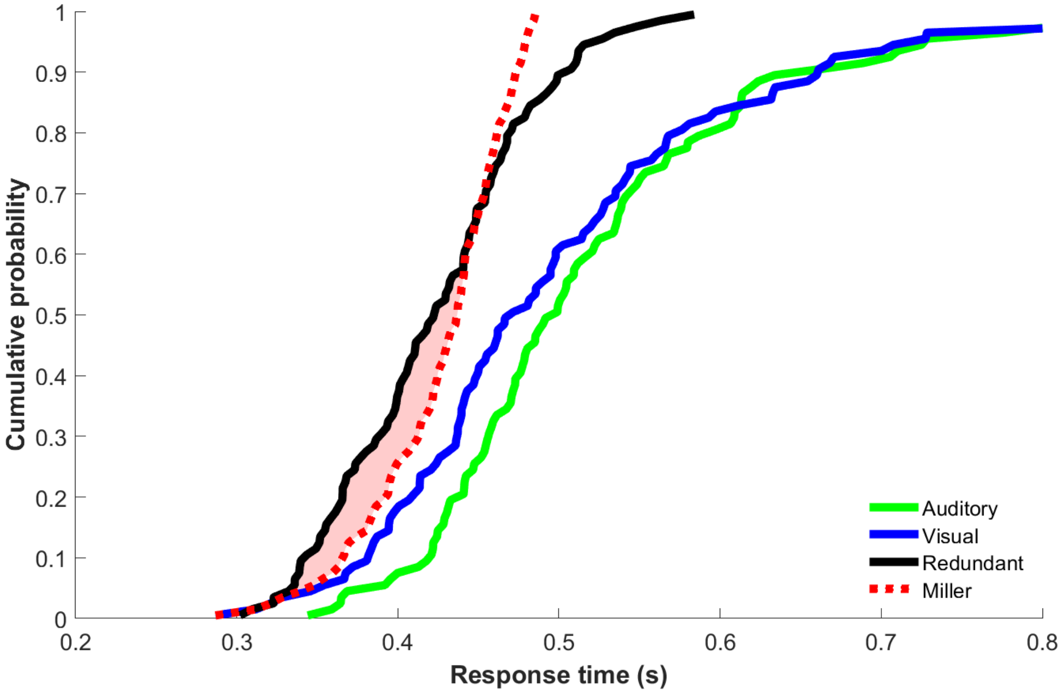


Figure 4.4.3. The graphical representation of the geometric quantification of Miller's (1982) bound violation. The bound (dotted red line) is calculated from the auditory (solid green line) and visual (solid blue line) components. If the empirical audio-visual performance (solid black line) is in excess of the bound, then there is a violation. The size of the violation is measured by the area in excess of the bound (shaded pink).

Results

Analysing the violations (Figure 4.4.4), a 2x2 ANOVA was performed, with the factors auditory motion direction (AR, AL) and visual motion direction (VR, VL). First, there was a significant intercept, $F_{(1,19)} = 119$, $p<0.001$, $\eta_p^2 = 0.862$, meaning that violations of Miller's (1982) bound was significantly different from zero. Additional bootstrap simulations showed that violations were significant on the group level in each of the four conditions. On the level of individual participants, violations were significant for all participants in conditions with an AL component, while the violations were significant for 12/20 participants in the ARVR condition, and 14/20 participants in the ARVL condition. The ANOVA also revealed a main effect of auditory motion direction ($F_{(1,19)} = 22.3$, $p<0.001$, $\eta_p^2 = 0.539$). Pairwise comparisons on the auditory main effect found that violations were larger in conditions with an AL component than in conditions with an AR component (mean difference: 0.005 s ±0.001 s). There was no main effect of visual motion direction, nor an interaction between auditory motion direction and visual motion direction. Additionally, a paired-samples t-test was performed to check if the violation on the ALVL condition was largest. ALVL was compared against the condition with the next largest violation, ALVR. The t-test returned a non-significant result ($t_{(19)} = 0.767$, $p = 0.453$, mean difference: 0.001 s ±0.002 s).

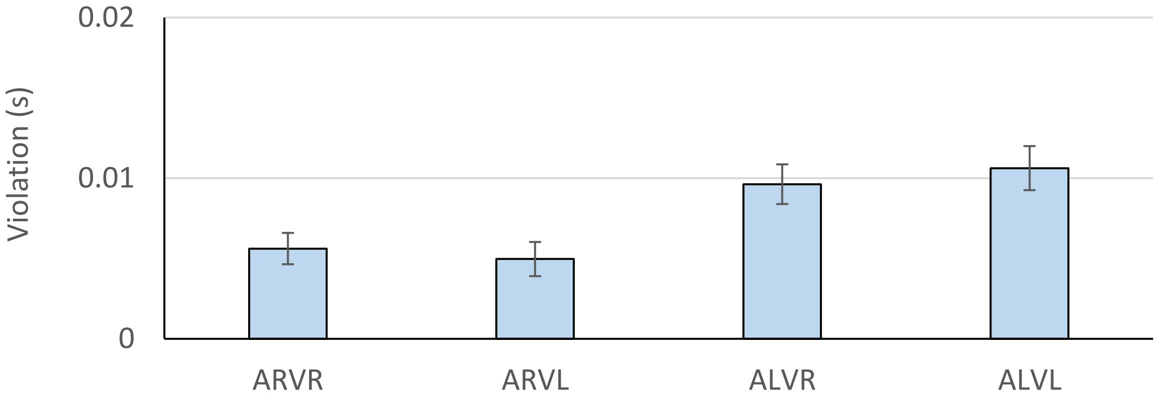


Figure 4.4.4. Participant-averaged violations of Miller’s (1982) bound (SEM error bars). The size of the violations was only determined by the auditory component: larger violations if the audio-visual condition featured an AL component rather than an AR component.

Summary

Significant violations of Miller’s (1982) bound were found, meaning that the responses were faster than possible under a probability summation rule with negatively dependent RTs. Often, such violations are taken as direct evidence against the race architecture. Should the finding of violations in this experiment be taken as evidence for integrative processing, and perhaps selective integration towards audio-visual looming signals (Cappe et al., 2009)? The answer is not so clear.

First, it is important to refer back to the claim of selective integration towards audio-visual looming signals (Cappe et al., 2009). If there was indeed selective integration towards audio-visual looming signals (ALVL), then its violations of Miller’s (1982) bound, as an indication of extra processing speed, should be particularly large, while the violations in other audio-visual conditions should be small or non-existent. Yet, it was found that the violations in ALVL were no larger than the incongruent ALVR. In fact, all audio-visual conditions tested exhibited significant violations of Miller’s (1982) bound. Moreover, the size of the violations were only determined by a unisensory auditory factor, with larger violations for conditions with an auditory looming component, than for conditions with an auditory receding component. There were no indications that a multisensory effect was at play. Hence, there does not appear to be a multisensory looming bias, much less special processing unique to audio-visual looming signals.

Interestingly, this unisensory auditory effect on the violations is identical to that found on the RSEs themselves, where there were larger RSEs / violations for conditions with an auditory looming component than with an auditory receding component. As with the RSEs, this unisensory auditory effect is easily explainable with probability summation, as the principle of congruent effectiveness (Otto et al., 2013), because the RTs towards auditory looming was faster than towards auditory receding, and therefore closer to the visual RTs. Hence a larger redundancy gain in conditions with an auditory looming component (ALVL, ALVR), than conditions with an auditory receding component (ARVL, ARVR).

To summarise, there were significant violations. Yet, the size of the violations follows a pattern that can be explained by probability summation. The violation on ALVL was neither unique nor particularly large among the audio-visual conditions. It appears, therefore, that the ‘selective’ aspect towards ALVL cannot be supported. However, it also seems ambiguous whether the best explanations for the processing of audio-visual signals are in the race or integration architectures. Here, it is important to remember that Miller’s (1982) bound still assumes context invariance. So, if there are violations of Miller’s (1982) bound, then the problem could be the race architecture, or the assumption of context invariance (Otto & Mamassian, 2017). Taking that it is the assumption of context invariance that is incorrect, a computational modelling approach was taken in Chapter 5 to answer whether the race architecture which accounts for context *variance* would be a powerful explanatory framework for understanding the audio-visual processing of looming signals.

4.5 Chapter discussion

The impetus for the present knowledge gap was the old conundrum of multisensory research: are multisensory decisions made using a race or integration architecture? The redundant signals paradigm of the experiments, where either the auditory or visual signal is sufficient for a response, perfectly matches the logical ‘or’ of the race architecture, where parallel unisensory units ‘race’ each other to accumulate enough sensory evidence to their decision criteria, the ‘winning’ unit then determines the behavioural output (Otto & Mamassian, 2017). Yet, in the landmark Miller’s (1982) test, which is an upper bound of race architecture performance using negatively correlated RTs, RTs in multisensory conditions often exceed the bound, and such violations are often interpreted as evidence for the integration architecture. Furthermore, a redundant signals experiment on audio-visual looming signals found fastest RTs towards congruent audio-visual looming signals, and made the claim of a selective integration mechanism towards congruent audio-visual looming (Cappe et al., 2009).

However, there are in fact two incorrect assumptions underlying basic probability summation, the statistical facilitation enabling redundancy gains in the race architecture (Otto & Mamassian, 2017). These two assumptions of basic probability summation, which are statistical independence and context invariance, are often used but not always acknowledged (Gondan & Minakata, 2016; Otto & Mamassian, 2017). Hence, when multisensory data apparently does not fit with the race architecture, the problem could be the inclusion of the two incorrect assumptions, rather than the race architecture itself. The aim of this chapter was therefore to test the race architecture, and its applicability in explaining the responses towards audio-visual looming signals.

4.5.1 A quest for the best possible redundant signals experiment on audio-visual looming

Before being able to test the race architecture, there needed to be good behavioural data from a redundant signals experiment on audio-visual looming. Yet, before this chapter, there were Experiments 1 and 2, neither of which had satisfactory response accuracy and speed. Experiments 1 and 2 (and by extension, Cappe et al. (2009)) had an odd stimulus feature: onset transients. On each trial, stimulus onset and motion onset were concurrent, but it is only the motion onset which is relevant to the experimental task. The concurrence of stimulus and motion onset could be a distraction that impacted response performance.

To test the idea that onset transients were the issue, Experiment 3 was performed (Chapter 4.2). Experiment 3 removed onset transients by separating out stimulus onset from motion onset, and this resulted in significant improvements in both response accuracy and speed. However, the removal of onset transients, while keeping to the original stimulus definition, meant each trial had a foreperiod potentially cueing the modality of the motion signal, and could be priming participants into unduly excellent performance. Hence, Experiment 4 (Chapter 4.3) was conducted, removing onset transients while also re-defining the stimulus conditions such that all conditions had audio-visual stimulation, and no foreperiod cueing was made. Experiment 4 was the last in the series for this project, having good performance likely from the removal of onset transients, but without the shortcomings of Experiment 3. However, the question may still arise about how one can be sure that the onset transients are indeed the fundamental problem of the original stimuli. In Experiment 2, testing stimulus realism, it was concluded that the additional cues to motion-in-depth seemed only to bring superficial and marginal performance improvements, and were also masking a fundamental design flaw in the stimulus. Is there an indication that the onset transients are indeed the fundamental design flaw, i.e., the removal of onset transients is not of the same superficial nature as adding more motion-in-depth cues?

In the latter part of this chapter, the comparative approach called for a test of the assumption of statistical independence, and the key metric was the modality switch costs (Chapter 4.4.2). The modality switch cost is the slower performance on the current trial if a modality switch was involved, versus if the modality was repeated onto the current trial. The existence of a significant modality switch cost is evidence against the assumption of statistical independence. However, in calculating the modality switch costs for all four experiments, an interesting trend became apparent (see Figure 4.5.1). Specifically, Experiment 1 (replication) and Experiment 2 (replication with more cues to motion-in-depth) both had onset transients in their stimuli, and both have large modality switch costs. Experiment 3 (basic stimuli, no onset transients) and Experiment 4 (re-defined basic stimuli, no onset transients) both have small modality switch costs. It appears as though that onset transients in the stimuli are related to large modality switch costs.

To be sure about such a difference across the experiments, a one-way ANOVA was performed on the modality switch costs, comparing all four experiments. There was a significant main effect of experiment ($F_{(3,57)} = 8.469$, $p<0.001$, $\eta_p^2 = 0.308$). The modality switch costs of

Experiment 3 (basic stimuli, no onset transients) were significantly smaller than that of the experiments with onset transients, which is Experiment 1 (mean difference: 0.025 s ±0.007 s, p = 0.021), and Experiment 2 (mean difference: 0.030 s ±0.008 s, p = 0.006). The modality switch costs of Experiment 4 (re-defined basic stimuli, no onset transients) were also smaller than that of the experiments with onset transients, approaching significance against Experiment 1 (mean difference: 0.026 s ±0.009 s, p = 0.055), and significant against Experiment 2 (mean difference: 0.031 s ± 0.010 s, p = 0.036). Between the two experiments with no onset transients (Experiment 3 and Experiment 4), there was no significant difference in the modality switch costs (mean difference: 0.001 s ±0.006 s, p = 1.000). In short, experiments without onset transients have similarly low modality switch costs, lower than the experiments with onset transients.

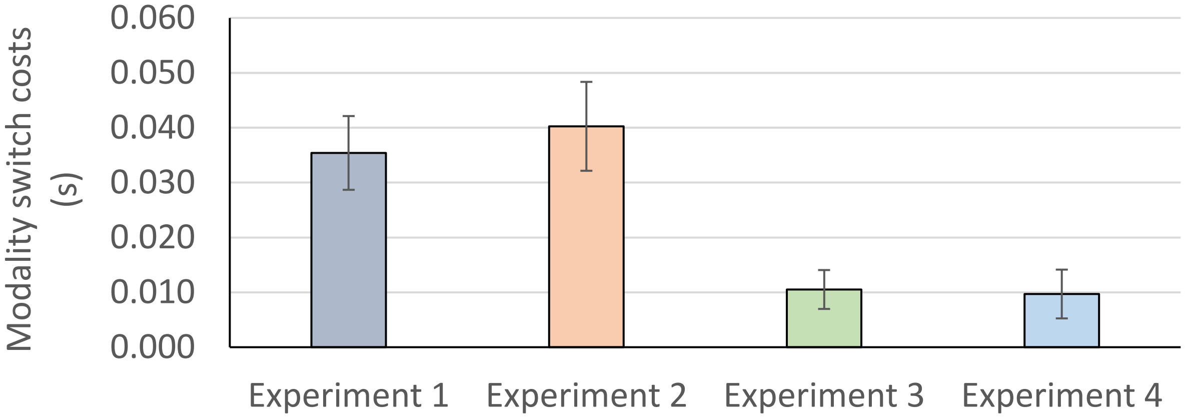


Figure 4.5.1. Participant-averaged modality switch costs (SEM error bars), for all four experiments. Experiment 1 is the replication of Cappe et al. (2009), with basic stimuli and onset transients (Chapter 3.1). Experiment 2 modifies on Experiment 1, with additional cues on the stimuli, and onset transients (Chapter 3.2). Experiment 3 builds returned to basic stimuli, but removed onset transients (Chapter 4.2). Experiment 4 builds on Experiment 3, with re-defined basic stimuli, without onset transients (Chapter 4.3). Notice that the two experiments with onset transients (Experiments 1 and 2) have higher modality switch costs, than the two experiments without onset transients (Experiments 3 and 4).

The intrigue around the modality switch costs comes together when one also considers the response performance in these experiments. Regarding experiments with onset transients, Experiment 1, the replication of Cappe et al. (2009) was notable for having low accuracy and slow RTs, particularly in the auditory conditions, while Experiment 2, the addition of more motion-in-depth cues, only marginally improved on Experiment 1. Both of these experiments have high modality switch costs (Figure 4.5.1). Regarding experiments without onset transients, Experiment 3, despite using the same basic stimuli as the replication, produced near-perfect accuracy and fast RTs, while Experiment 4 with re-defined basic stimuli was not far behind Experiment 3. Both of these experiments have low modality switch costs (Figure 4.5.1). It seems as though there is some link between onset transients on the stimuli, the response accuracy and speed, and the modality switch costs, e.g., with onset transients, response accuracy and speed are poor, and the modality switch costs are high. Causality cannot be ascertained from these results alone, but is nonetheless an intriguing finding.

If one were to speculate, then it could be that the onset transients represent a processing cost, because it means having to selectively attend to the task-relevant motion onset while filtering out the task-irrelevant stimulus onset at the same time. Assuming that the auditory motion signal – intensity change – was also not easy to detect or distinguish, then the auditory signal with onset transients could be particularly difficult to process and decide on. The visual condition did not suffer as much with the added processing cost of onset transients, because the visual signal of size change is an obvious signal. The difficulty in processing the auditory conditions would reflect in the poor response accuracy and speed, but it might also negatively impact the ability to swiftly change processing between modalities, i.e., into or out of audition, and is thus reflected in high modality switch costs as well. By removing the onset transients, the auditory stimuli was no longer as difficult to processing, so the auditory response performance improved, and the modality switch costs shrank, as was seen in Experiments 3 and 4 (Figure 4.5.1) which had the onset transients removed.

The above is pure speculation for what might be the underlying link between onset transients, response performance and modality switch costs. If the speculation is true, then modality switch costs seems to be an indirect indicator of the hidden factor, stimulus difficulty. As for the original question of why onset transients were considered to be the fundamental problem of the stimulus, and that the addition of more motion-in-depth cues was only superficial, the answer is in the modality switch costs (Figure 4.5.1). Taking that the modality switch costs are indeed an indirect indicator of stimulus difficulty, then notice that the addition of more motion-in-depth cues (Figure 4.5.1, Experiment 2) has not reduced the modality switch costs compared to the replication (Figure 4.5.1, Experiment 1); neither of these two experiments had particularly good response accuracy and speed. However, taking the basic stimuli but removing its onset transients (Figure 4.5.1, Experiment 3), the modality switch costs were significantly lower than in Experiment 1, with much better response accuracy and speed. The low modality switch costs and good performance persisted in Experiment 4 (Figure 4.5.1), which also had onset transients removed, but with re-defined stimulus conditions. Thus, taking that the modality switch costs indicate stimulus difficulty, only onset transients were a critical factor in stimulus difficulty, while the addition of more motion-in-depth signals seems not to make the stimuli easier.

4.5.2 Testing the race architecture: The comparative approach

With the best possible data from a redundant signals experiment, the project is in the position to answer the knowledge gap: to what extent can the race architecture explain multisensory decision-making to audio-visual motion-in-depth signals?

The approach taken to answer the knowledge gap was to use the race architecture in its simplest form to predict multisensory performance using empirical unisensory data, then compare the prediction to the empirical multisensory data: the comparative approach. The race architecture in its simplest form is the independent race model (Raab, 1962), which uses a simple probability summation rule that assumes both statistical independence and context invariance. Probability summation is core to the race architecture, so if the predictions of probability summation in some way matches the real multisensory data, then there are prospects to the race architecture. Crucially, it turned out that the RSEs predicted by simple probability summation captures the pattern of the empirical RSEs. Effects which were significant in the empirical RSEs were also significant in the predicted RSEs. Namely, there was only an effect of auditory motion direction on the RSEs; there were no multisensory effects. Moreover, there was a significant correlation between the predicted and empirical RSEs. Hence, it appears that probability summation, which underlies the race architecture, is capable of capturing the variances in the RSE data, across the four combinations of audio-visual motions. It seems that there is potential in the race architecture for explaining multisensory decision-making.

However, it cannot be ignored that although probability summation captured the patterns in the RSEs, the predictions by probability summation were nonetheless underestimates of the empirical RSEs. Therefore, the question is whether this is a failure of the race architecture.

In reality, it would be an over-simplification to declare the failure of the race architecture on the back of its underestimation of RSEs. The problem is that the particular form of probability summation used, although simple to implement, assumes both statistical independence and context invariance. Both assumptions reduce the redundancy gain possible in the race architecture, and neither are likely to be correct (e.g., Otto & Mamassian, 2017). Hence, the problem could be the assumptions, rather than the race architecture itself. Indeed, significant modality switch costs were found, which is evidence against the assumption of statistical independence.

The crucial test comes in the form of violations towards Miller's (1982) bound, which drops the assumption of statistical independence by assuming a maximal negative correlation (Colonius, 1990). Violations of Miller's (1982) bound are in effect extra speed in multisensory processing that cannot be explained by a race architecture with negative dependencies; the violations are often taken as direct evidence for the integration architecture. Critically, in this experiment, significant violations were found in the empirical RSEs of every audio-visual condition tested. Care must be taken with the interpretations at this point. On the first level, no support was found for the claim of selective integration (e.g., Cappe et al., 2009), because the violations (as a measure of extra processing speed) were not unique, nor especially large, towards the audio-visual looming signals. Instead, there was only an effect of auditory motion on the violations, namely, larger violations when the condition contained an auditory looming component, than an auditory receding component. In fact, this auditory effect is identical to that found with the RSEs, and is predicted by probability summation, because the RT towards auditory looming is more similar to visual RT, compared with RT towards auditory receding (cf. principle of congruent effectiveness; Otto et al., 2013). On the second level, the finding of violations towards Miller's (1982) bound should not be taken as automatic evidence for integration, because Miller's (1982) bound in fact still assumes context invariance. So when violations of Miller's (1982) bound are found, the issue could be with the assumption of context invariance, rather than the race architecture as such. Nonetheless, a proper test of the context invariance assumption is necessary. To definitively answer the question of whether the race architecture explains audio-visual processing of looming signals, the next chapter applies a novel race model that accounts for statistical dependencies and context variances (Otto & Mamassian, 2012), under a computational cognitive modelling technique.

Chapter 5: Answering knowledge gap 3: What can a computational modelling analysis show about the multisensory processing of audio-visual looming signals?

5.1 Computational modelling analysis on the processing of audio-visual looming signals

5.1.1 Modelling: the basics

Up to this point in the project, the analyses have been on the level of hit rates, overall accuracy, RTs, RT-based quantities and so on; essentially, basic behavioural metrics. However, the reality is that ostensibly simple behaviours are in fact the product of many underlying processes (Heathcote, Brown, & Wagenmakers, 2015). Therefore, these basic behavioural metrics offer only basic insights into the behaviour at hand, and it would be inaccurate to speculate mechanism based only on basic behavioural metrics (Heathcote et al., 2015). Similarly, it is haphazard to theorise only on a verbal level, because the spoken language cannot contain enough details to fully and accurately instantiate the theory, not to mention that it is susceptible to fallacies and inconsistencies (Farrell & Lewandowsky, 2015; Lewandowsky & Farrell, 2011).

Instead, if one had a theory of how a cognitive process works, then it must be instantiated in a model which formalises the architecture of the mechanism, and specifies all the assumptions (Farrell & Lewandowsky, 2015; Lewandowsky & Farrell, 2011), such that there are testable hypotheses from the theory. A model contains parameters, which are values, each of which exemplify a process underlying the behaviour (Farrell & Lewandowsky, 2015; Lewandowsky & Farrell, 2011). Depending on the architecture of the model, some or all of the parameters can be adjusted (adjustable parameters are also known as free parameters), and the adjustments affect the behaviour predicted by the model, given the bounds specified by the model's architecture (Lewandowsky & Farrell, 2011). The essence of computational modelling is to adjust the values of each free parameter, until the behaviour predicted by the model fits the observed behavioural data (Lewandowsky & Farrell, 2011).

Yet specifically, how does one adjust the values of a free parameter, to achieve good fit between model and data? First, the fit between model and data is quantified by their discrepancy, and a good fit means small discrepancies (Lewandowsky & Farrell, 2011). With different values on a free parameter, or different combinations of values on multiple free parameters, the model should vary in its predicted behaviour, and hence would become less or more discrepant with the real behaviour (Lewandowsky & Farrell, 2011). On a basic level, there can be a trial and error process of adjusting the parameter values until minimum discrepancy is found, however, this would be a computationally expensive and time-consuming technique of estimating the parameter values (Lewandowsky & Farrell, 2011). Instead, several start values of the parameters are tried, and are iterated from there to seek for minimal model-data discrepancy, using functions that are built-in to Matlab (Lewandowsky & Farrell, 2011). Over the field of all possible parameter value combinations, there may be local minima of discrepancy, but the goal is to find the global minimum of discrepancy; having multiple start values helps minimise the risk of producing parameter estimates for a local minimum (Lewandowsky & Farrell, 2011). If there is a set of parameter estimates which produces small discrepancies between the model and data, then such a model has shown itself to be a possible explanation for the observed behavioural phenomenon at hand (Lewandowsky & Farrell, 2011). This brief description of cognitive modelling is the basis of the modelling performed in this chapter.

5.1.2 Race models of interactive audio-visual processing: a range for model comparison

In Chapter 4, under the comparative approach, a simple probability summation rule could predict the general pattern of RSEs found from a redundant signals experiment on audio-visual motion-in-depth signals. Probability summation is the statistical facilitation that allows for redundancy gains in the race architecture. The similarity between the empirical and predicted RSEs, namely good empirical-prediction correlation (r = 0.662, p = 0.001), and both empirical and prediction finding a unisensory effect of the auditory component, suggests that there is good explanatory potential in probability summation and hence the race architecture.

However, although the predicted RSEs followed the patterns in empirical RSEs, the predicted RSEs were underestimates. The issue is that the prediction used a simple probability summation rule which assumes both statistical independence and context invariance (independent race model; Raab, 1962), both of which are likely to be incorrect, and both reduce the size of the RSE prediction (e.g., Colonius, 1990; Miller, 1982; Otto & Mamassian, 2017; Townsend et al., 2020). For one, the significant modality switch costs found in Chapter 4.4.2 is evidence against the assumption of statistical independence. If the assumptions of statistical independence and context invariance were to be accounted for, could the race model architecture be a good explanatory framework for understanding audio-visual processing of motion-in-depth signals?

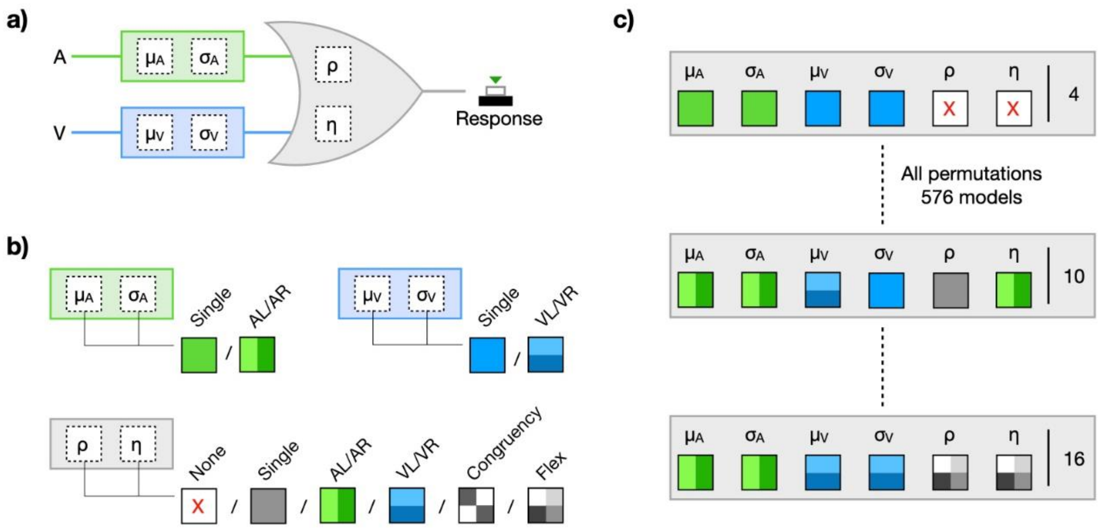


Figure 5.1.1. Schematic of a race model with interactions between unisensory units (Otto & Mamassian, 2012). a) The basic architecture of the race model with interactions: the unisensory units (auditory and visual) are in parallel 'race' configuration, and described by a rate μ and its standard deviation σ. The interaction parameter ρ allow for dependencies in RTs between the two modalities (thus allowing for statistical *dependence*). The interaction parameter η is an additional noise term to allow for additional noise in audio-visual conditions than in unisensory conditions, and allows for context *variance*. b) To extend the basic interaction race model (Otto & Mamassian, 2012), the parameterisation could vary to accommodate looming unisensory, multisensory, looming factors. The unisensory parameters of μ and σ can be singular, or vary with motion direction to account for looming biases. The multisensory parameters of ρ and η can be absent, singular, vary with auditory or visual motion direction, vary with audio-visual motion congruency, or be fully flexible. c) Permutations on these parameterisations produces a total of 576 nested model variants, ranging from a simple four-parameter model that has neither interactions nor looming biases, to a 16-parameter model. Figure taken from my publication (Chua et al., 2022).

Such an audio-visual-interaction race model exists (Otto & Mamassian, 2012). In the Otto and Mamassian (2012) model, for each modality, there is a LATER unit of sensory evidence accumulation (Carpenter & Williams, 1995; Noorani & Carpenter, 2016). The LATER unit assumes a reci-normal distribution of unisensory RT, meaning that the reciprocal of RTs, 1/RT rate, is normally distributed with a mean rate, μ, which has a standard deviation, σ (Carpenter & Williams, 1995; Noorani & Carpenter, 2016). The exact distribution is described using two random Gaussian numbers (Nadarajah & Kotz, 2008). In the case of audio-visual processing, there would be two LATER units in parallel, one unit for audition with parameters μ_A and σ_A, the other unit for vision with parameters μ_V and σ_V. As a race model, the unit with the higher rate 'wins the race', thus determining the behavioural output (Otto & Mamassian, 2012). To allow for interactions between the parallel units, a correlation ρ parameter was added to allow for statistical dependencies between unisensory RTs, and an additional noise η parameter to account for additional noise in redundant conditions versus unisensory conditions, in effect allowing for context variance (see Figure 5.1.1a for an illustration of the model; Otto & Mamassian, 2012).

For the purposes of this project, the model of Otto and Mamassian (2012) needs to be extended to account for the eight auditory, visual, and audio-visual conditions tested (see also Figure 5.1.1b). For the unisensory units, the parameters can be extended to account for the factor unisensory motion direction. For example, the auditory unit can have a rate parameter of singular μ_A, or vary with auditory motion direction, μ_{AR} and μ_{AL}. The variance parameter can also be of singular σ_A, or vary with auditory motion direction, σ_{AR} and σ_{AL}. The unisensory visual parameters can also be similarly extended to account for visual motion direction. For the interaction, each parameter was extended to account for six possibilities. The correlation ρ parameter can be (1) not used, (2) a singular ρ, (3) varying with auditory motion direction, ρ_{AR} and ρ_{AL}, (4) varying with visual motion direction, ρ_{VR} and ρ_{VL}, (5) varying with audio-visual motion congruency, $\rho_{congruent}$ and $\rho_{incongruent}$, or (6) fully flexible, having a ρ value for each audio-visual condition, ρ_{ARVR}, ρ_{ARVL}, ρ_{ALVR}, ρ_{ALVL}. Similarly, the η parameter can be (1) not used, (2) a singular η, (3) varying with auditory motion direction, η_{AR} and η_{AL}, (4) varying with visual motion direction, η_{VR} and η_{VL}, (5) varying with audio-visual motion congruency, $\eta_{congruent}$ and $\eta_{incongruent}$, or (6) fully flexible, having an η value for each audio-visual condition, η_{ARVR}, η_{ARVL}, η_{ALVR}, η_{ALVL}.

Thus, taking all possible combinations of parameter values, 2^2 x 2^2 x 6^2 (auditory x visual x interaction), there were 576 nested variants of this extended, interactive race model derived from Otto and Mamassian (2012). Each model variant is the combination of different parameters, and suggests slightly different factors at play (see Figure 5.1.1c for example parameterisations, from a 4-, 10-, and 16-parameter model). For example, the simplest model variant has only four parameters (μ_A, σ_A, μ_V, σ_V), and suggests only unisensory factors needed, no interactions between modalities, and no looming bias. The most complex variant has 16 parameters (μ_{AR}, μ_{AL}, σ_{AR}, σ_{AL}, μ_{VR}, μ_{VL}, σ_{VR}, σ_{VL}, ρ_{ARVR}, ρ_{ARVL}, ρ_{ALVR}, ρ_{ALVL}, η_{ARVR}, η_{ARVL}, η_{ALVR}, η_{ALVL}). Altogether, the set of 576 models encompasses a wide range of factors (unisensory, multisensory, interaction, non-interaction, et cetera). Model selection was performed to choose the model which 'best' explains the data.

5.1.3 Model comparison: fitting and selection

The model fitting technique was quantile maximum probability estimation, which uses RT quantiles for robustness against outliers and higher computing efficiency, whilst maintaining comparable accuracy to continuous maximum likelihood estimation which uses the full RT distribution (Heathcote, Brown, & Cousineau, 2004; Heathcote, Brown, & Mewhort, 2002). Maximum likelihood estimation, in general terms, is a measure of discrepancy between model and data, but is statistical in the sense that it finds parameter values that are most likely given the data (Lewandowsky & Farrell, 2011). The first step of model fitting was to have the valid RTs of each condition in quintiles, and a count of RTs in the corresponding quintile bins. Then, with Matlab's fmincon function, fitting was performed by adjusting the parameter values until the discrepancy between model and data was minimised, quantified by twice the negative log quantile likelihood summed across all eight conditions. In doing so, the quantile probability was also maximised.

As described in the general description of modelling in Chapter 5.1.1, a potential pitfall in model fitting is that the search for minimal model-data discrepancy leads to, and concludes at a local minimum, rather than the global minimum; sub-optimal parameter estimates would be taken. Hence, for this study, to avoid local minima, there were a range of start values for each parameter. The μ parameters had start values from separately-obtained best-fitting estimates for the unisensory conditions, with ±2% either side. The σ parameters had start values from separately-obtained unisensory best-fitting estimates, with ±2.5% either side. The ρ parameters had 10 start values, evenly spaced from -0.9 to 0.9. The η parameters had four start values evenly spaced between 0% and 30% of the best-fitting σ estimates. Thus, there were up to 600 start values for each model fitting, the exact number depending on the parameters involved, as dictated by the model's parameter structure.

The fitting of one model to an individual participant's data produces a set of parameter estimates, and a measure of the model-data discrepancy, quantified by twice the negative log quantile likelihood. There was a set of 576 models, and each model was fitted to the data, thus producing a set of 576 parameter estimates and model discrepancies. From the 576 model fits, the objective was to select the model that best fits the data: a model with minimal model-data discrepancy, achieved using as few parameters as possible (e.g., principle of parsimony; Lewandowsky & Farrell, 2011). The fundamental issue in model selection is that the model must have enough parameters to accurately account for important effects in the data, but not an excess of parameters which would fit real effects and noise in the data (Lewandowsky & Farrell, 2011).

The objective of selecting the best model is achieved using an information criterion, which balances model-data discrepancy with a penalty for increasing number of parameters (Lewandowsky & Farrell, 2011). The two most well-known information criteria are the Akaike Information Criterion (AIC; Akaike, 1974) and Bayesian Information Criterion (BIC; Schwarz, 1978). However, views are divided over which criterion is best to use (Chakrabarti & Ghosh, 2011). On the one hand, the BIC finds favour owing to its Bayesian underpinnings, and its preference for simpler models due to the BIC featuring a heavy penalty term for model complexity, compared to the AIC (Chakrabarti & Ghosh, 2011; Lewandowsky & Farrell, 2011; Raftery, 1999). On the other hand, the BIC has been questioned over its theoretical integrity, sensitivity to priors, and its preference for simplicity which may neglect important parameters and lead to inaccurate model selection; classical information criteria such as the AIC may be better (Weakliem, 1999). A balanced approach in choosing the AIC or BIC is to apply both to select their respective 'best' model, and of the two, choose the information criterion with the model having the least model-data discrepancy (Chakrabarti & Ghosh, 2011). In my study, without a clear prior theoretical justification for using one information criterion over the other, both AIC and BIC were used for model selection, and the information criterion with the least discrepant 'best' model was taken. Additionally, owing to the lengthy computation time required for model selection (in excess of one week, using several dozen computers interlinked together working non-stop), both AIC and BIC model selections were performed and provided here, for theoretical completeness and practical convenience.

In terms of implementation, the AIC (balancing model fit with model complexity, i.e., number of parameters), was applied on an individual basis to determine the best model for each of the 20 participants (Appendix B). For the group-level best model, the individual AIC scores were summed across participants, for each of the 576 models, then converted into AIC weights and the model with the highest weight was the best group model (Lewandowsky & Farrell, 2011; Rae, Heathcote, Donkin, Averell, & Brown, 2014; Wagenmakers & Farrell, 2004). Similarly, the BIC (balancing model fit with a heavier penalty term for model complexity), was applied on an individual basis to determine the best model for each of the 20 participants (Appendix C), from which the group-level BIC selection was also calculated.

5.1.4 Modelling results

AIC Results

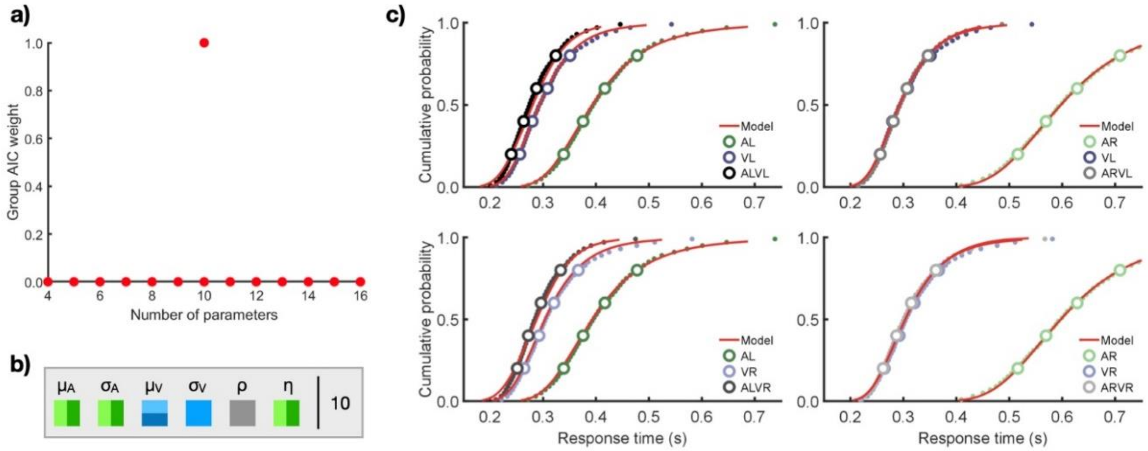


Figure 5.1.2. In the model comparison, AIC selected a 10-parameter model as best fitting for the group. a) The best group-level model among the 576 was determined by selecting the one with the highest group AIC weight, which was a 10-parameter model. b) The winning 10-parameter model had auditory parameters which varied with motion direction, a visual rate μ that varied with motion direction but its σ was singular. Critically, the interaction terms were either singular or depended only on the auditory motion direction. c) The model (lines) fits nearly perfectly to the data (dots, quantiles marked with large circles), in all conditions. Figure taken from my publication (Chua et al., 2022).

Out of the 576 models in the model comparison, the AIC selected a 10-parameter model as best-fitting to the empirical data, on a group level (see Figure 5.1.2a showing this 10-parameter model having the highest group AIC weight among the 576 models, Figure 5.1.2b for the winning model's parameters). Crucially, this winning 10-parameter model had near-perfect fits to the data (Figure 5.1.2c). Interestingly, seven out of the 10 parameters were for describing the unisensory differences in motion direction, while the other three were interaction parameters which were not related to any audio-visual congruency effects (Figure 5.1.2b). Following is a more detailed analysis of the parameters.

Selection criterion	Parameters	μ_{AL} (s^{-1})	μ_{AR} (s^{-1})	μ_{VL} (s^{-1})	μ_{VR} (s^{-1})	σ_{AL} (s^{-1})	σ_{AR} (s^{-1})	σ_V (s^{-1})	ρ	η_{AL} (s^{-1})	η_{AR} (s^{-1})
AIC	10	2.594 ±0.084	1.709 ±0.058	3.492 ±0.115	3.350 ±0.096	0.527 ±0.029	0.316 ±0.015	0.619 ±0.021	-0.394 ±0.109	0.192 ±0.027	-0.007 ±0.013

Table 5.1. The parameter estimates of the best-fitting model on a group level, as selected by the AIC. The model uses 10 parameters.

For the auditory parameters, both the rate μ and standard deviation σ varied with auditory motion direction. Thus, the model had the auditory parameters μ_{AL}, μ_{AR}, σ_{AL} and σ_{AR} (Table 5.1). A paired-samples t-test found μ_{AL} had a significantly larger rate than μ_{AR} ($t_{(19)} = 17.319$, $p<0.001$, mean difference: 0.885 s^{-1} ±0.226 s^{-1}). Thus, sensory evidence was accumulated faster in receiving the AL signal than the AR signal, matching with the empirical finding that RTs towards AL were faster than towards AR, by about 0.2 seconds. A paired-samples t-test also found σ_{AL} was significantly larger than σ_{AR} ($t_{(19)} = 8.094$, $p<0.001$, mean difference: 0.211 s^{-1} ±0.026 s^{-1}), meaning that sensory accumulation was noisier with AL than with AR. Altogether, the modelling results show a faster but noisier rate for AL signals than AR signals, and is evidence for a unisensory auditory looming bias.

For the visual parameters, the rate μ varied with visual motion direction, but the standard deviation σ was singular. Thus, the model had the visual parameters μ_{VL}, μ_{VR}, and σ_V (Table 5.1). A paired-samples t-test found a significantly faster rate on μ_{VL} than μ_{VR} ($t_{(19)} = 5.540$, $p<0.001$, mean difference: 0.141 s^{-1} ±0.026 s^{-1}). Thus, sensory evidence was accumulated faster with VL than with VR. Altogether, the model captured a unisensory visual looming bias by finding a significantly faster rate with VL than with VR signals. However, only a singular σ_V parameter was used, which could be a reflection of similar RTs in the two visual conditions.

The interaction parameters are, however, most interesting in terms of understanding the multisensory processing in this experimental task. First, a negative correlation ρ parameter was found (-0.394 ±0.109, Table 5.1), and a one-sample t-test found it to be significantly different to zero ($t_{(19)} = -3.62$, $p=0.002$). The finding of a significant negative correlation is important, because it means that the content of previous trials do affect current performance, thereby refuting the assumption of statistical independence in basic probability summation (i.e., independent race model; Raab, 1962). The negative dependency is in line with Colonius (1990), which pointed out that negative dependencies are necessary for there to be statistical facilitation at all, and fits with the notion of limited mental processing capacity. The negative dependency also links with the notion that a modality switch has an RT cost compared to a modality repeat (Shaw et al., 2020), a RT cost which was indeed present in my data (see Chapter 4.4.2 for the modality switch cost analysis). Crucially, the ρ parameter was only found to be singular, so there does not appear to be a multisensory effect in the correlation term.

The other interaction parameter, additional noise η, was found to vary with auditory motion direction. Thus there were the parameters η_{AL} and η_{AR} in the model (see Table 5.1). A paired-samples t-test found η_{AL} to be significantly larger than η_{AR} ($t_{(19)} = 7.72$, $p<0.001$, mean difference: 0.198 s^{-1} ±0.0026 s^{-1}), meaning that there was more additional noise in audio-visual conditions with an AL component than with an AR component. The η parameterisation with auditory motion direction could be due to faster responses in AL than AR, hence AL processing temporally overlaps with visual processing, and produces more interaction noise, more than with an AR component. Nonetheless, the fact that the best-fitting model included η terms is evidence against the assumption of context invariance, because the η term is to allow for additional noise in multisensory conditions compared to unisensory conditions, thus showing that processing is not the same with or without stimulation in another modality. Therefore, the η term allows race architecture performance to violate Miller's (1982) bound. Crucially, the η terms found here only varied with auditory motion direction; there does not appear to be a multisensory effect in the additional noise term.

BIC results

From the pool of 576 nested models, the BIC selected a 6-parameter model as best-fitting to the data, on a group level (see Figure 5.1.3a showing this 6-parameter model having the highest group BIC weight among the 576 candidate models, Figure 5.1.3b for this model's parameterisation). Curiously, all six parameters are to do with the unisensory processes – there are no interaction parameters. This BIC-selected 6-parameter model does not appear to fit closely with the data (Figure 5.1.3c, deviation between model (line) and data (dots)), an indication that perhaps this model is too simple and does not capture all the effects in the data.

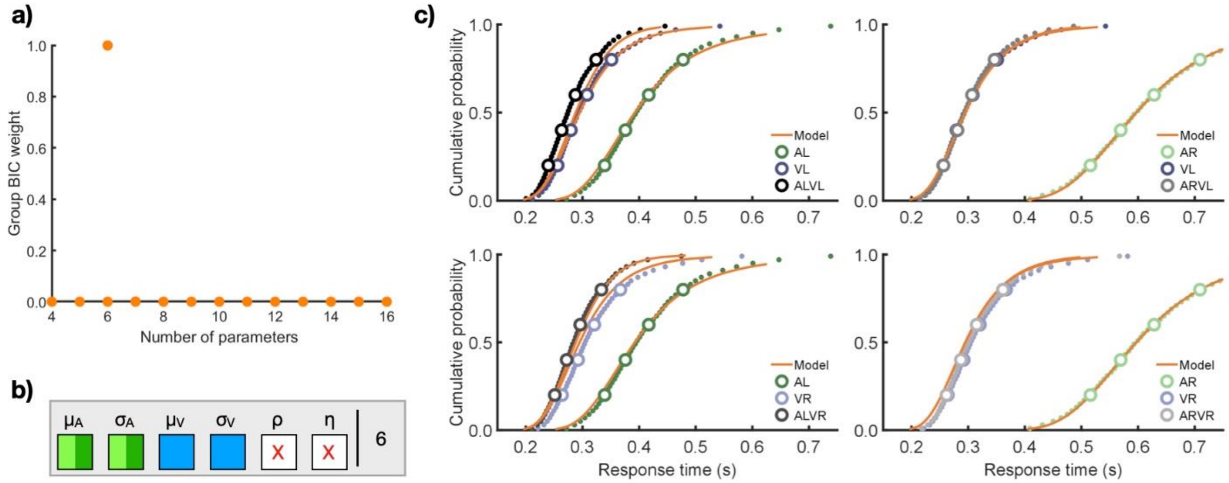


Figure 5.1.3. In the model comparison, BIC selected a 6-parameter model as best-fitting. a) The best group-level model among the 576 candidate models was determined by selecting the model with the highest group BIC weight, which was a 6-parameter model. b) The winning 6-parameter model had unisensory auditory parameters which varied with motion direction, but the unisensory visual parameters were singular, and there were no interaction parameters. c) With only six parameters, this BIC-selected model (lines) showed some discrepancy from the data (dots, quantiles marked with large circles). Figure taken from my publication (Chua et al., 2022).

Selection criterion	Parameters	μ_{AL} (s^{-1})	μ_{AR} (s^{-1})	μ_V (s^{-1})	σ_{AL} (s^{-1})	σ_{AR} (s^{-1})	σ_V (s^{-1})	ρ	η (s^{-1})
BIC	6	2.603 ±0.089	1.710 ±0.058	3.456 ±0.107	0.567 ±0.040	0.318 ±0.015	0.658 ±0.021	x	x

Table 5.2. The parameter estimates of the model selected as best-fitting on a group level, according to the BIC. This model uses 6 parameters, none of which interaction parameters. This 6-parameter model selected by the BIC is less complex than the 10-parameter model selected by the AIC.

For the auditory parameters, both the rate μ and standard deviation σ varied with motion direction. Hence, the model had the parameters μ_{AL}, μ_{AR}, σ_{AL} and σ_{AR}. A paired-samples t-test found μ_{AL} to be significantly larger than μ_{AR} ($t_{(19)} = 16.108$, $p<0.001$, mean difference: 0.893 s^{-1} ± 0.055 s^{-1}). Hence, sensory evidence accumulation was faster in conditions that had an AL component than those that had an AR component, which matches with the empirical finding that RTs for AL- conditions were about 0.2 seconds faster than in AR- conditions. A paired-samples t-test found σ_{AL} to be larger than σ_{AR} ($t_{(19)} = 6.873$, $p<0.001$, mean difference: 0.249 s^{-1} ±0.036 s^{-1}), meaning that the sensory accumulation was noisier in AL- conditions than in AR- conditions. Altogether, the model found faster but noisier sensory accumulation for conditions with an AL signal than with an AR signal, which is evidence for an auditory looming bias. These findings of motion dependency on the auditory parameters, and an auditory looming bias, were also found in the AIC-selected model.

For the visual parameters, both the rate μ and standard deviation σ were singular, possibly as a reflection of the similarity in RTs between conditions with VL or VR signals. Note that the selection of singular visual parameters by the BIC is a further simplification of the AIC selected model, which still had separate μ_{VL} and μ_{VR} rates.

Lastly, the model does not include interaction parameters, meaning it does not allow for statistical dependencies nor additional noise in multisensory conditions. This BIC-selected 6-parameter model is in effect the independent race model (Raab, 1962), which is in line with BIC's favouring of simpler models (e.g., Heathcote et al., 2015). However, seeing that the model seems not to closely fit the data (see deviations between model lines and data dots, Figure 5.1.3c), this BIC-selected model of six parameters may have been too simple to fully capture all effects in the data (see also Weakliem, 1999 for a critique of BIC).

5.1.5 Summary

In this model comparison, 576 nested models were tested against the data. The 'best' model, which is a balance of its closeness to the data and model complexity, has parameters which would offer insights to the multisensory processing of audio-visual motion-in-depth signals. The AIC selected a 10-parameter model with a near-perfect fit to the data, and its parameterisations suggested unisensory looming biases. Crucially, interaction terms were used by this AIC-selected 10-parameter model, the ρ parameter being singular and suggesting a negative correlation, while the η parameter varied with auditory motion direction and suggested more additional noise in conditions with an AL component than in conditions with an AR component. On speculation, the parameterisation of η with auditory motion direction might be explained by faster RTs towards AL than AR (i.e., AL RTs closer to visual RTs). Hence, the distribution of AL RTs should overlap more with the visual RTs, more than with AR RTs, so more noise would be present in AL-conditions than AR- conditions. Altogether, the utilisation of interaction parameters suggested that both assumptions of statistical independence and context invariance are incorrect, and should be dropped when using the race architecture. However, the BIC favoured simplicity, selecting a 6-parameter model with no interactions, in effect the independent race model (Raab, 1962). The BIC is known to be more punitive towards increasing numbers of model parameters than the AIC (Heathcote et al., 2015; Lewandowsky & Farrell, 2011), so it was not a surprise that the BIC selected a simpler model than the AIC, but the absence of interaction parameters was surprising. Perhaps the ρ and η interaction parameters in the Otto and Mamassian (2012) model do not quite capture the processing of audio-visual motion-in-depth signals, so the BIC omitted both interaction parameters, even if this meant selecting a too simplistic model that does not closely fit the data.

Altogether, in this analysis which compares AIC and BIC for model selection, with the basis on having the least model-data discrepancy, it seems that the 10-parameter model selected by the AIC is more appropriate than the 6-parameter model selected by the BIC. Crucially, neither AIC- nor BIC-selected models were parameterised with audio-visual motion congruency, or with different parameters for each audio-visual combination (i.e., a different parameter set only for audio-visual looming). Thus, this model comparison found no evidence of multisensory processing special to congruent audio-visual looming motion. Instead, a relatively simple model, based on probability summation (the foundation of the race architecture) with a small number of interaction parameters, is sufficient to explain the processing of audio-visual motion-in-depth signals.

5.2 Parameter and model recoveries

Having selected the 'best' model, a good practice in computational modelling is to perform parameter and model recoveries, as a sanity check for any biases or incorrect premises in the modelling methodology (Heathcote et al., 2015). With the empirical experiments, behavioural data was generated, and modelling was applied to this data to deduce the model and parameter estimates within the model. With recovery, it is about using the known parameter estimates (i.e., after the model fitting / selection process) to generate synthetic data, and seeing if the same model fitting and selection process as used with the empirical data recovers the data-generating parameters or model from the synthetic data (Heathcote et al., 2015). If the data-generating parameters are recovered unbiased, and if the same model is chosen again amongst competing models, then the modelling procedures are robust. The recovery procedure is typically repeated many times e.g., 100 repetitions, to see the spread in recovery parameter estimates, or the proportion of times the generating model is recovered. First, parameter recoveries using the winning AIC-selected model were performed (Chapter 5.2.1), then a series of model recoveries (Chapter 5.2.2), to check for the robustness of this project's experimental data and techniques. Although likely too simple (see Chapter 5.1), out of interest, the BIC-selected 6-parameter model was also subjected to the parameter and model recoveries.

5.2.1 Parameter recovery

Technique

First, each participant's (20 participants) empirical parameter estimates to the AIC-selected 10-parameter model were used to generate synthetic data in the same size as the empirical dataset (8 conditions, 100 trials per condition, 20 participants), 1000 times over (iteration). In effect, the synthetic dataset simulated 1000 experiments (1000 x 16,000 trials). To prepare for model fitting using quantile maximum probability estimation (Heathcote et al., 2004; Heathcote et al., 2002; as used for the empirical data), the synthetic RTs were converted to RT quintiles and a count of RTs falling into the quintile bins, per condition, participant and iteration. Second, the same model fitting function as with the empirical data was used on this synthetic dataset. The model was fixed as the AIC-selected 10-parameter model. The model fitting was repeated 1000 times, owing to the dataset being in effect 1000 experiments. At the conclusion of this model fitting, the 20 participants each had 1000 estimates on their 10 parameters. Finally, the mean, median and 95% confidence interval were taken on the 1000 estimates for each of the participant's 10 parameters. The results were then plotted in graphs (Figure 5.2.1). For completeness, the same technique as above was used for the parameter recovery of the BIC-selected 6-parameter model. The only difference was that the BIC used six parameters.

Results

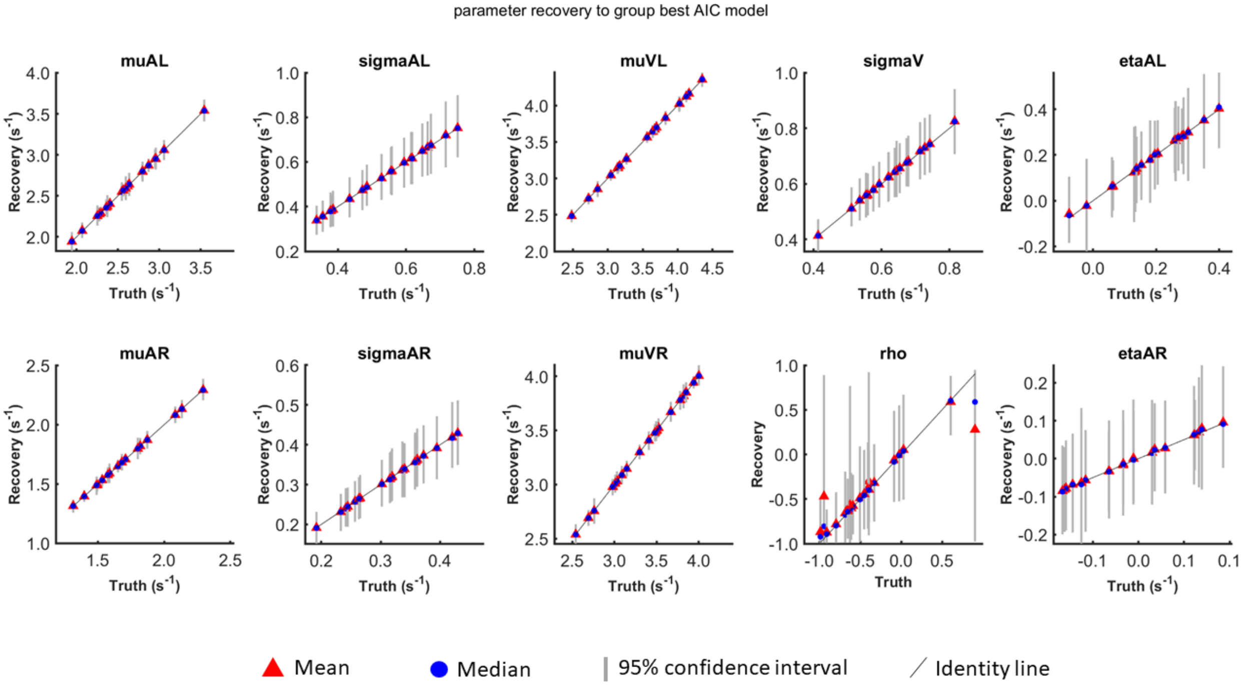


Figure 5.2.1. Parameter recoveries on each of the 10 parameters in the AIC selected model. Each participant's 10 parameters each had 1000 estimates. The red triangle and blue circle represents each participant's mean and median of their parameter estimates, respectively. The vertical lines represent the 95% confidence interval of each participant's 1000 parameter estimates. The diagonal line is the identity line, where the generating parameter value ('truth') equals the recovered parameter estimate.

Looking at each AIC-model parameter, it appears that the recovery performance was mixed (Figure 5.2.1). First, for all parameters, the mean and median of each participant's parameter estimates were largely on the identity line (generating value = recovered value), meaning that on average, the generating parameter values were recovered unbiased. However, the amount of variation on the parameter estimate is not homogenous across all 10 parameters. The unisensory rate parameters (μ) all have tight 95% confidence intervals. The unisensory standard deviation parameters (σ), the interaction η parameter, and especially the ρ parameter have large 95% confidence intervals.

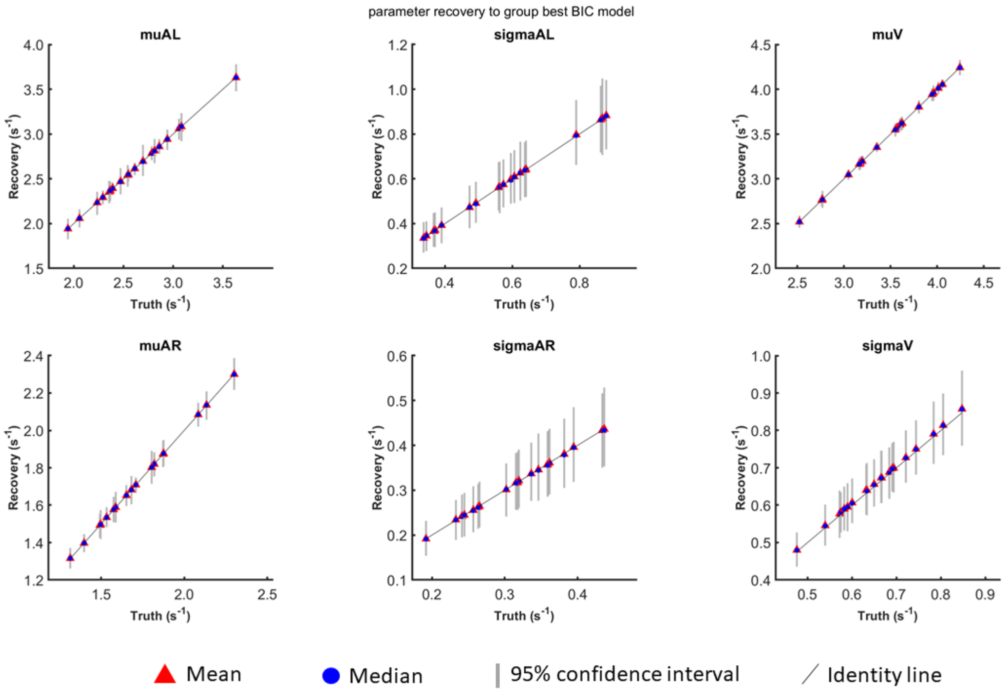


Figure 5.2.2. Parameter recoveries on each of the six parameters of the BIC selected model. Each participant's six parameters each had 1000 estimates. The red triangle and blue circle represent the mean and median of each participant's 1000 parameter estimates, respectively. The vertical line represents the 95% confidence interval of the 1000 parameter estimates. The diagonal line is the identity line, where the generating value ('truth') is equal to the recovered parameter estimate.

Looking at each BIC-model parameter, the recovery performance again looks mixed (see Figure 5.2.2). On the positive, the mean and median of each participant's parameter estimates are on the identity line (generating value = recovered value), so the parameters have been recovered unbiased. However, the spread of recovery parameter estimates is wide for the standard deviation (σ) parameters.

Summary

To summarise, on average the recovered parameters were true to the generating parameter values, without biases, hence suggesting that the model fitting and selection procedures were robust and unbiased. However, on parameter types to do with unisensory noise (σ) and especially the interaction effects (ρ and η), the spread of the parameter estimates over 1000 iterations was wide. The wide spread of parameter estimates is a sign that the parameter is unstable, and could relate to the notion that the ρ and η interaction were not the optimal constructs for characterising the audio-visual processing of motion-in-depth.

5.2.2 Model recoveries as a function of model complexity

Basic model recovery

The basic model recovery used the empirical individual parameter estimates under the AIC-selected 10-parameter model to generate a synthetic dataset, in 100 iterations, thus simulating 100 experiments (100 x 16,000 trials). Then, the same model fitting and selection technique as with the empirical data was applied to the dataset, 100 times over, using a competitive set of 26 models. The 26 models were chosen based on them having been selected in the original model fitting and selection procedure on the empirical data (e.g., AIC best-, AIC 2nd best-, BIC best-, each participant's best-fitting model et cetera). A smaller model pool instead of the full 576 models was essential to keep computing time reasonable. Using AIC model selection, it was found that the generating 10-parameter model was recovered in 98% of the 100 iterations. The other 2% of recoveries went to a closely-related 11-parameter model, which used two visual σ parameters (σ_{VL}, σ_{VR}) instead of the singular σ_V of the generating model. Using the same technique as above, the recovery was 100% for the BIC-selected 6-parameter model. Altogether, the generating model was reliably recovered, thus showing that the selected 10- or 6-parameter models were stable, and the employed modelling technique was robust and unbiased.

Model recovery as a function of model complexity

As a further step, model recoveries were performed on various levels of model complexity (as defined by the number of parameters). This step was to explore if the model recovery would still be robust for less or more complex models.

Technique

The levels of model complexity were 6, 7, 8, 9, 10, 11, 12, 13, and 14 parameters. At each level of model complexity, the best-fitting model of that complexity was chosen and the individual participant fits to that model were used to generate synthetic datasets in 100 iterations. Then, the datasets (representing the various model complexities) underwent model fitting and selection (as used with the empirical dataset). A wide and competitive pool of 33 models was used for the model selection. The pool comprised of the three best-fitting models (from the fitting to empirical data) at each of the following model complexities: 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15. Once model recovery was complete, it was examined if the generating models were recovered.

Results

When the group AIC was used as the model selection criterion, recovery performance was at or near 100% for models that had 10 or fewer parameters (Figure 5.2.3a). However, with 11 parameters in the model, recovery performance dipped to 51%, and deteriorated further towards 1% when the model had 14 parameters. On closer inspection, with 11 parameters generating, the model recovery was split between recovering the generating 11-parameter model (51%), and the 'winning' 10-parameter model (49%). With 12 parameters generating, seven models were recovered, though none of the recovered models had more than 12 parameters. Instead, the majority of recoveries were towards 11-parameter models and the 'winning' 10-parameter model. With 13 and 14 parameters generating, again, there was a spread of models recovered, and the majority were towards models with fewer parameters than that generating.

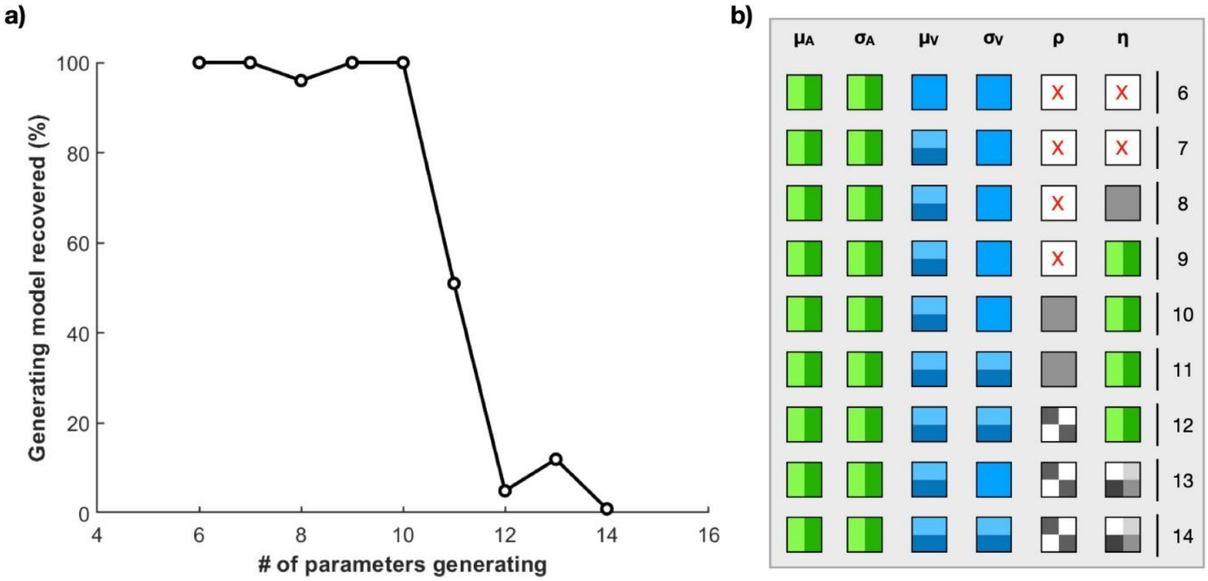


Figure 5.2.3. Model recovery using group AIC for model selection, over a range of model complexities (number of parameters). a) Recovery performance was at or near 100% for data generated with up to 10 parameters. However, there was a rapid drop in recovery performance for data generated with 11 or more parameters. The falloff in recovery performance suggests that past 10 parameters, the additional parameters are superfluous to the data. b) The parameterisation of the best-fitting models (obtained from the original fitting of empirical data) at each model complexity. Parameter symbols as used Figure 5.1.1; unisensory parameters can be single or vary with motion direction (bi-colour); interaction parameters can be not used, singular, motion-dependent in either modality, congruency-dependent, different to each audio-visual condition. Figure adapted from my publication (Chua et al., 2022).

Hence, it appears that with up to 10 parameters, the generated data matched the complexity in the empirical data. However, with 11 or more parameters generating, the additional parameters were unnecessary to the data, so the recovery tended towards simpler models. For example, the difference between the 10- and 11-parameter models is the σ_V parameter (Figure 5.2.3b). The 10-parameter model has a singular σ_V, while the 11-parameter model has σ_{VL} and σ_{VR}. Yet, as σ_{VL} and σ_{VR} can be near identical for some participants, the additional parameterisation is superfluous to the data, and the 10-parameter model is selected.

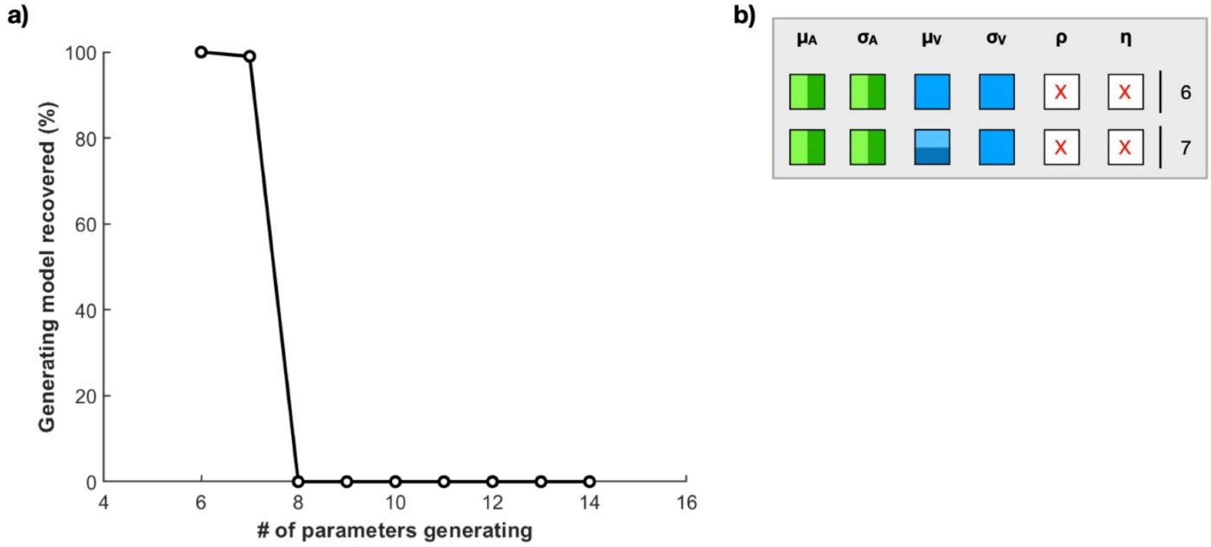


Figure 5.2.4. Model recovery using group BIC for model selection, over a range of model complexities (number of parameters). a) Recovery performance was 100% or near 100% when the data was generated using seven or less parameters. More complex models were not recovered at all, as the group BIC universally recovered a 7-parameter model at such higher levels of generating model complexity. b) The parameterisation of the 6- and 7-parameter model, showing that apart from μ_V, they are identical: both require motion-dependent parameterisations for audition, a singular σ_V, and no interaction parameters.

For completeness, model recoveries using group BIC selection were also considered (Figure 5.2.4a). Apart from recovering the 6-parameter model at 100%, and the 7-parameter model at 99%, the group BIC-selection did not recover any of the more complex generating models. Instead, group BIC universally selected a 7-parameter model when a model with more than seven parameters was generating. The 6- and 7-parameter models are nearly identical, both not featuring any interaction parameters, and are only differentiated by the unisensory visual μ parameterisation (Figure 5.2.4b). It appears that the BIC selection has a strong bias for selecting simple models, perhaps overly so given the BIC-selected 6-parameter model fits poorly to the empirical data (see Figure 5.1.3c; Chapter 5.1).

Summary

Altogether, the model recoveries here show that the 'winning' 10-parameter model selected by the AIC is indeed the best model to explain the audio-visual processing of motion-in-depth. First, this 10-parameter model was recovered at or near 100% in two separate model recoveries with different competitive model pools. Then, in performing model recoveries with data generated from different levels of model complexity, this 10-parameter model was the most complex to still achieve 100% recovery. Recovery performance fell rapidly with data generated from 11 parameters or more. Hence, this 10-parameter model has a sufficient amount of parameters to fully capture the effects in the data; more parameters would be superfluous. Adding to the earlier suspicion that the BIC is choosing overly simple models for this study, model recovery using the BIC steadfastly chose 6- or 7-parameter models, which in Chapter 5.1, the 6-parameter model was shown to be discrepant to the data.

5.2.3 Model recovery as a check of non-congruency bias, sample size

The modelling so far has found no congruency effects that would suggest a multisensory looming bias. However, instead of concluding that there are indeed no multisensory looming biases, another possibility can be entertained: the modelling technique here may in some way be biased against congruency effects. Hence, in this model recovery, the winning 10-parameter model was adapted to include rho parameters that vary with audio-visual motion congruency. This exercise was to see if a congruency model would be recovered using the modelling technique here. Additionally, a sample size analysis was conducted as part of the recovery, to check if the combination of 100 trials per condition and 20 participants was sufficient for robust modelling.

Technique

As the basis for generating synthetic data, the participant-averaged parameter estimates of the winning 10-parameter model was used, retaining the unisensory and auditory motion-dependent η parameters (η_{AL} = 0.2, η_{AR} = 0), but with modifications to ρ. Namely, instead of a singular ρ parameter, the ρ parameter varied with audio-visual motion congruency ($\rho_{congruent}$ = -0.5, $\rho_{incongruent}$ = 0.5), making for a 11-parameter model. To vary the sample size, the dataset was generated with three levels of trials per condition (50, 100 and 200 trials per condition), with participant size variable between 1 and 40. Hence, the overall sample size could vary between 400 trials (50 trials per condition, 1 participant) to 64000 trials (200 trials per condition, 40 participants). The synthetic data was generated in 100 iterations for model recovery. Then, the same model fitting and selection technique as with empirical data was used on this synthetic dataset. Only AIC selection was used, as it was shown earlier that the BIC preferred overly simple models. To keep computing times reasonable, a shortlisted model pool was used instead of the full set of 576 models. This shortlisted model pool consisted of all models which shared the same seven unisensory parameters (μ, σ) as the generating model (and also the 10-parameter winning model). Thus, with six possibilities each on ρ and η, there were 36 models in the model pool.

Results

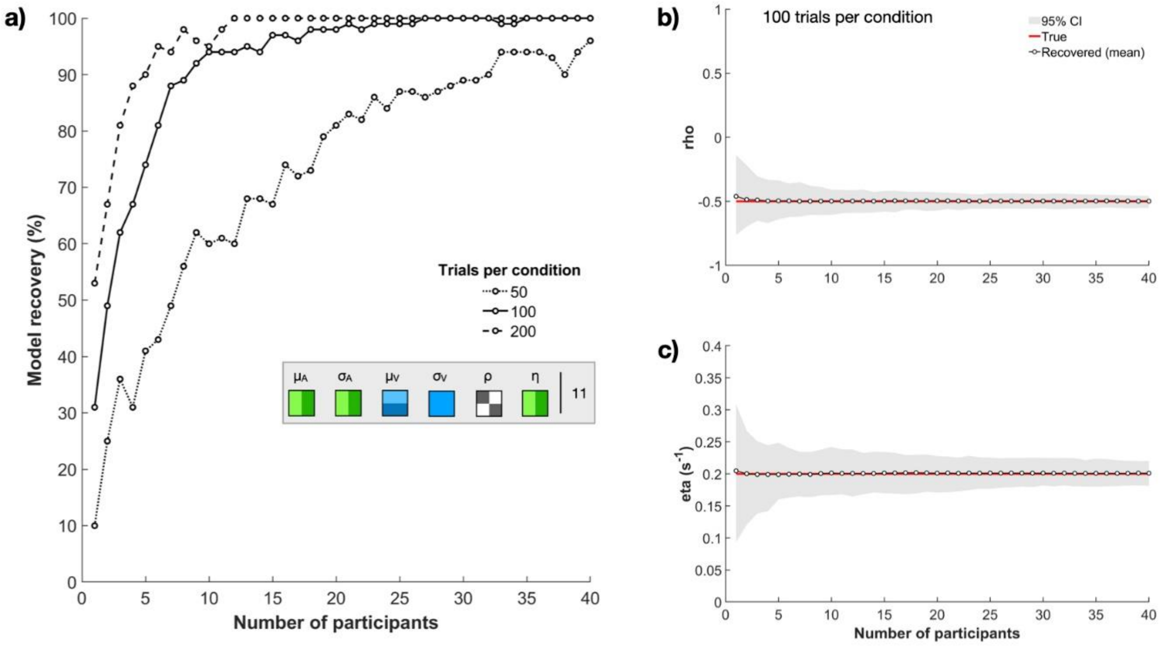


Figure 5.2.5. Recovery performance for a 11-parameter model with induced ρ parameters varying with audio-visual motion congruency. A sample size analysis was also performed as part of the analysis. a) Model recovery performance (%, vertical axis) improved with sample size, with higher recovery performance for more trials per condition (separate lines), and more participants (horizontal axis). Crucially, the congruency-ρ model was reliably recovered with the sample size used in the experiments of this project (100 trials per condition, 20 participants). The parameterisation of the congruency-ρ model is illustrated in the grey box. b) With 100 trials per condition, the mean of recovered ρ estimates (circles) was unbiased to the generating ρ value (red line, 'true'), with decreasing uncertainty for more participants (grey area, 95% confidence interval). c) With 100 trials per condition, the mean of recovered η estimates (circles) was unbiased to the generating η value (red line, 'true'), with decreasing uncertainty for more participants (grey area, 95% confidence interval). Figure adapted from my publication (Chua et al., 2022).

The recovery performance of the congruency-ρ model is in Figure 5.2.5a. Generally, an increase in sample size led to increased recovery performance. Specifically, with only 50 trials per condition (dotted line), recovery performance was low, and reached the 90% recovery level only if there were more than 30 participants. In contrast, with 100 trials per condition (solid line) and 200 trials per condition (dashed line), recovery performance approached 90% with 10 or fewer participants. In terms of participant size, more participants lead to higher recovery performance. Taking 100 trials per condition, the recovery of ρ and η estimates were analysed as a function of participant size (Figures 5.2.5b, c, respectively): the uncertainty in the parameter estimates were large with few participants, but rapidly decreased with more participants, and had stabilised by 20 participants. Crucially, over the range of participant sizes tested (up to 40), the average of the recovered parameter estimates were accurate to the generating parameter value. Altogether, the experiments in this project had 100 trials per condition and 20 participants, which this analysis shows to have sufficient power for robust modelling.

Summary

Overall, this analysis showed that the employed modelling technique does not have a non-congruency bias. First, there seems to be sufficient power in the sample size used (100 trials per condition, 20 participants). This model recovery analysis also showed that a model with audio-visual motion congruency parameters could be reliably recovered with unbiased parameter estimates, using the employed modelling techniques. Hence, if indeed the best model to explain the behavioural data needed a parameter based on audio-visual motion congruency, then there are no biases against the selection of such a model.

5.3 Chapter discussion

Prior to this chapter on modelling, basic behavioural metrics (accuracy, RT) towards audio-visual motion signals were analysed, and no evidence was found in support of a multisensory bias towards audio-visual looming signals. However, ostensibly simple behaviours can be the product of complex underlying processes, so it would be inappropriate to claim mechanism from basic behavioural metrics (Heathcote et al., 2015). Rather, a theory and its assumptions should be instantiated as a model, from which there are hypotheses testable in modelling (Farrell & Lewandowsky, 2015; Lewandowsky & Farrell, 2011).

The race architecture seems to have potential in explaining multisensory processing. From the comparative approach in Chapter 4, it was seen that a basic probability summation rule (the statistical facilitation core to the race architecture) correctly predicted the patterns in the empirically obtained RSEs. However, the probability summation predictions underestimated the full size of the empirical RSEs, which may have been caused by the assumptions of statistical independence and context invariance. Hence, it could be that the race architecture is fundamentally sound, but the incorrect assumptions need to be relaxed. Otto and Mamassian (2012) proposed a model based on the race architecture, but with two interaction parameters: a correlation ρ to account for statistical dependencies in unisensory RTs, and an additional noise term η to allow for additional noise in multisensory conditions than in unisensory conditions. To determine the best model, i.e., the parameterisations that best describe the empirical data, a model comparison and selection was performed, using 576 nested permutations on the unisensory parameters and novel interaction parameters.

From the model comparison and selection using the AIC (Akaike, 1974), a 10-parameter model was found to be best-fitting to the data. Interestingly, seven out of the 10 parameters were to do with unisensory processes, with a higher but noisier sensory accumulation rate for auditory looming than receding signals, and just a higher rate for visual looming than receding signals. These unisensory parameters reflect unisensory looming biases. More important to understanding the multisensory process are the interaction parameters. In this model, there was a singular negative ρ parameter, and two η parameters varying with auditory motion direction. First, the negative ρ was expected, because negative dependencies in RTs between modalities are necessary for there to be redundancy gains in the race architecture (see Colonius (1990); the negative dependency is typified in modality switch costs, see also Chapter 4.4.2, and Shaw et al. (2020)). Critically, the finding of negative dependencies refutes the assumption of statistical independence. Next, two η parameters were found. The η parameter is to allow for additional noise in multisensory conditions than unisensory conditions, thus, the usage of η shows that the processing in multisensory conditions is not the same as in unisensory conditions – the assumption of context invariance is incorrect. The two η parameters varied with auditory motion direction, such that there was more additional noise in multisensory conditions with an auditory looming component than with an auditory receding component. This η parameterisation based on auditory motion direction could stem from the faster responses in AL than with AR – the processing of AL temporally overlapped with visual processing, more than with AR, hence more audio-visual interaction and a larger additional noise η term required for conditions with AL, than conditions with AR. Crucially, none of the parameters in this best-fitting model suggested a special process only towards audio-visual looming signals: there were no parameters varying with audio-visual motion congruency, or fully flexible such that processing audio-visual looming signals has its own parameter values. Hence, this modelling analysis found no evidence of selective processing towards audio-visual looming signals. There were only unisensory effects, and a race architecture was sufficient to explain the audio-visual processing of motion.

To verify these model findings, and as a sanity check of the modelling techniques, parameter recovery and a series of model recoveries were performed. In short, the 10 parameters were recovered unbiased, and the best-fitting 10-parameter model was reliably recovered. Further recoveries showed that additional parameters were superfluous to the data. Yet, if the data indeed was so complex (i.e., generated from 11 parameters), including an interaction parameter that varied with audio-visual motion congruency, then this modelling technique, and the employed sample size (100 trials per condition, 20 participants), would have reliably selected such a model. Thus, the recoveries added weight to the modelling results, demonstrating that the modelling technique is unbiased, and it selected a non-congruency 10-parameter model as best-fitting. Altogether, this modelling chapter shows that a race architecture, with interactions, is a powerful explanatory framework for understanding audio-visual processing of motion-in-depth.

On a final note, the BIC was also used for model selection but its preference for model simplicity (defined by the number of parameters) seems too strong, consistently selecting models that are too simple to fully account for the effects in the data, as shown by its choice of a 6-parameter model which only loosely fitted the data. The BIC consistently avoided models with interaction parameters in the model recoveries, so perhaps the ρ and η constructs are not the most apt to the audio-visual processing of motion. Nevertheless, the BIC-selected models also did not feature parameters that would suggest special processing towards congruent audio-visual looming signals.

On a critical note, one modelling practice is to analyse on the level of individual participants, e.g., perform model fitting and selection on each participant's data (Lewandowsky & Farrell, 2011). However, modelling on the level of individual participants produced heterogenous results. It could be that each participant processes the audio-visual motion signals differently, perhaps even changing moment to moment, so a homogenous model selection across the participants did not emerge. Another explanation is that there may not have been enough data on an individual level to model robustly. With one participant, 100 trials per condition, there would be 800 valid trials at maximum, which referring to the sample size analysis in Chapter 5.2.3, is at the low end of sample sizes. Taking that sample size analysis for reference (see Figure 5.2.5a, solid line for 100 trials per condition, one participant), 800 valid trials would have produced poor model recovery performance of around 30%. A much more robust and consistent picture emerged with group-level data. Hence, the analysis here focused on group-level data. For reference, the results of modelling on an individual level are shown in Appendix B for AIC selection, Appendix C for BIC selection.

Chapter 6: Section discussion

This project on audio-visual looming reaches its conclusion, with the three knowledge gaps answered. The impetus for this project was to investigate how audition and vision could work together in processing looming motion. Looming is motion towards oneself, which in the real environment can be associated with an impending collision or attack, so there is a need to quickly detect and evade this dangerous motion (cf. behavioural urgency hypothesis; Franconeri & Simons, 2003). With earlier and quicker evasive responses, there can be a larger gap between oneself and the looming object, thus increasing the survival odds (Neuhoff, 2001). Aligning with such theoretical postulations, there is a wealth of research to show preferential responses towards looming signals, but not towards receding signals, in the perceptual, behavioural, physiological and neurological domains, separately for audition (e.g., Bach et al., 2008; Freiberg et al., 2001; Neuhoff, 1998, 2001; Rosenblum et al., 1993; Seifritz et al., 2002) and vision (e.g., Ball & Tronick, 1971; Franconeri & Simons, 2003; Lin et al., 2008; Moher et al., 2015).

However, the real world is multisensory, and there are advantages to multisensory perception (Ernst & Bülthoff, 2004), one of which is the speedup of responses (redundant signals effect, RSE; e.g., Giard & Peronnet, 1999; Hershenson, 1962; Kinchla, 1974; Miller, 1982; Todd, 1912). Hence, the question is whether there is a multisensory looming bias as well, such that there can be especially fast responses towards audio-visual looming signals. One suggestion was for a selective integration mechanism, to the effect of producing especially fast responses only towards audio-visual looming signals (Cappe et al., 2009). In this project, audio-visual processing of looming motion was examined in a series of four experiments (see Table 6.1 for a summary), using a variety of experimental techniques and levels of analysis, and the findings are discussed below.

Experiment 1	Replication of Cappe et al. (2009)	**Chapter 3.1**
Experiment 2	Additional stimulus cues on Experiment 1 setup	**Chapter 3.2**
Experiment 3	Onset transients removed, basic stimuli	**Chapter 4.2**
Experiment 4	Onset transients removed, basic stimuli, stimulus conditions re-defined	**Chapter 4.3**

Table 6.1. Summary of the four experiments in this project

6.1 Onset transients are the immediate problem here, not stimulus realism

Cappe et al. (2009) was successfully replicated in Experiment 1 (Chapter 3.1), but in doing so, problems with accuracy and RT performance were revealed. Despite achieving higher overall accuracy (and apparently faster RTs too) than Cappe et al. (2009), the auditory conditions were uniquely poor, with low accuracy and slow RTs. The stimuli in Experiment 1 were basic though (auditory intensity change, visual size change). Hence, the first research direction examined if more motion-in-depth cues on the stimuli would improve response accuracy and speed, thereby increasing any multisensory effect. In Experiment 2, more motion-in-depth cues were added to the same experiment paradigm: the auditory signal had a frequency change on top of the basic intensity change, while the visual stimuli received perspective cues in the background. However, the additional cues produced superficial improvements limited to auditory conditions (which were poor in Experiment 1); there was no impact on other modalities, nor the redundancy gains (RSE).

Instead, examining the stimuli more closely, a curious feature was found: the task-irrelevant stimulus onset was concurrent to the task-relevant motion onset. Hence, Experiments 3 and 4 tested the removal of onset transients, which greatly improved response accuracy and speed. Incidentally, a later analysis on modality switch costs seemed to reinforce the notion that onset transients were detrimental to performance: Experiments 3 and 4 (without onset transients) had similarly small modality switch costs, while Experiments 1 and 2 (with onset transients) had similarly large modality switch costs. So, without onset transients, response accuracy and speed improved, and seemingly indicated by low modality switch costs. Perhaps modality switch costs act as an indicator of stimulus difficulty, because difficult stimuli (particularly in the auditory condition) could manifest as slowness switching in and out of modalities across trials. If indeed modality switch costs indicate stimulus difficulty, then stimulus realism seems not to reduce stimulus difficulty (modality switch costs were equal between Experiments 1 and 2); only the removal of onset transients significantly reduced modality switch costs, hence stimulus difficulty.

Altogether, the additional motion-in-depth cues seemed to serve as an alternative, superficial marker that aided performance, rather than tapping into realism and generating enhanced or even new multisensory effects. This apparent non-effect of additional motion-in-depth cues corroborates with past research, which found that in audition, intensity cues are essential, whilst additional cues contribute minimally (e.g., Bach et al., 2009; Rosenblum et al., 1987). The topic of realism will be discussed again (Chapter 6.3), but at this stage, the more immediate problem was the onset transients on the stimuli, not such much the stimuli's realism. For the purposes of this project, the analysis used the data from Experiment 4, which featured basic stimuli with no onset transients.

6.2 Is there selective integration?

Cappe et al. (2009) proposed a selective integration mechanism for processing audio-visual looming signals, in effect a multisensory looming bias. This project has four experiments that cover several variations on the stimuli. From four datasets, is there anything in the behavioural metrics to suggest selective integration? Moreover, the data from Experiment 4 was also subjected to the comparative and modelling approaches, to determine a possible mechanism for processing audio-visual looming signals.

6.2.1 Behavioural metrics: RTs versus RSEs

Cappe et al. (2009) claimed a selective integration mechanism for processing audio-visual looming signals, on the back of finding faster RTs in audio-visual conditions than unisensory conditions, and fastest RTs in the congruent audio-visual looming condition, ALVL. However, the finding of absolute fastest RTs is not necessarily an indication of fast multisensory processing, much less a mechanism. Rather, the benefit of signal redundancy is the *speedup* of multisensory RTs compared to its unisensory RT components (RSE, redundant signals effect; Giard & Peronnet, 1999; Hershenson, 1962; Kinchla, 1974; Miller, 1982; Todd, 1912). If there was a process selective towards audio-visual looming, then the RSE of ALVL should be largest, while the RSEs in the other conditions should be small or non-existent.

From my four experiments, no evidence for selective integration was found in the RSEs. First, contrary to the implicit suggestions that fast RTs are the same as large RSEs, Experiments 2 and 4 both found ALVL RTs to be fastest of all, but their respective ALVL RSEs were statistically on par with ALVR RSEs. Furthermore, all audio-visual conditions produced RSEs, and Experiments 2, 3 and 4 even showed that the ALVL RSE was not uniquely large. Moreover, if there was selective processing towards audio-visual looming, then the RSE sizes should be modulated by a motion-specific audio-visual interaction. Yet, none of my experiments found such an interaction. Instead, the RSE sizes were only modulated by unisensory effects, with larger RSEs for conditions with a looming component, than with a receding component. Altogether, there was nothing special about the processing towards ALVL conditions, and the redundancy gains were simply determined from the unisensory components. Hence, there does not appear to be any selective integrative processing towards audio-visual looming signals.

Instead of integration, the finding of only unisensory effects on the RSE can be explained by the race architecture. It is important to recall that audition and vision each have looming biases (e.g., Bach et al., 2008; Moher et al., 2015), which was typically expressed as faster RTs towards AL than AR, and faster RTs in VL than VR, for Cappe et al. (2009) and for my experiments. The race architecture samples from the faster of two parallel unisensory decision units. So if the unisensory RTs were fast, then the sampled multisensory RTs would also be fast. As ALVL is the combination of the fastest auditory and visual conditions, then according to the race architecture, it is not surprising to find fastest RTs towards ALVL.

6.2.2 The comparative approach: a simple probability summation rule predicts RSE patterns

To directly test the potential of the race architecture, unisensory data from Experiment 4 was put through a simple probability summation rule (independent race model; Raab, 1962), thus a race model prediction of multisensory performance and RSE was produced, for each of the four audio-visual conditions. If there is potential in the race architecture, the RSEs predicted by probability summation must match with the empirical RSEs. Crucially, the predicted RSEs matched the unisensory effect on the empirical RSEs: an effect of auditory motion direction, such that conditions with an AL component have larger RSEs than conditions with an AR component. Furthermore, there was strong correlation between the empirical and predicted RSEs (r=0.662, p=0.001).

The effect of auditory motion direction can be explained by the race architecture principle of congruent effectiveness (Otto et al., 2013). AL had faster RTs and was therefore closer to the visual RTs, compared to AR, so the similarity in AL and visual performance produced stronger audio-visual interactions that would have allowed for a greater redundancy gain. The presence of only a unisensory effect on the RSE again shows that there is no selective processing towards ALVL. The correct prediction of empirical RSE patterns using probability summation shows the potential of the race architecture.

Nonetheless, despite correctly predicting RSE patterns, the predictions were underestimates. The problem is that the employed probability summation rule assumes both statistical independence and context invariance (independent race model; Raab, 1962). Statistical independence was shown to be incorrect by the finding of significant modality switch costs. Still, the audio-visual performance was in excess of a race architecture with negative dependencies (violation; Miller, 1982). If statistical dependencies and context variance could be accounted for, would the race architecture correctly predict audio-visual processing performance?

6.2.3 The model approach: race architecture with simple interactions explain audio-visual processing of looming motion

Computational modelling represented the third approach in deducing the mechanism in processing audio-visual looming signals. If one has a theory for how something works, then it must be instantiated as a model, from which hypotheses can be tested (Farrell & Lewandowsky, 2015; Lewandowsky & Farrell, 2011). Otto and Mamassian (2012) devised a race model with two interactions: a ρ correlation parameter to account for audio-visual RT dependencies (addressing the assumption of statistical independence), and an η additional noise parameter which allows for more noise in multisensory than unisensory conditions (addressing the assumption of context invariance). This race model with interactions (Otto & Mamassian, 2012) formed the theoretical basis of this project's computational modelling approach.

This project's model comparison used 576 model permutations, which encompassed a wide range of effects, including audio-visual congruency, and unique processing for ALVL. Crucially, neither model selection criteria (AIC, BIC; Akaike, 1974; Schwarz, 1978) chose a model which featured audio-visual congruency or unique processing for ALVL. Instead, the AIC chose a 10-parameter model which allowed for unisensory looming biases, a singular negative ρ parameter for negative audio-visual RT dependencies, and η additional-noise parameters varying with auditory motion direction. Corresponding to the earlier postulation that the similarity between AL and visual RTs would have produced stronger audio-visual interactions (thus producing a larger RSE) than with the slower AR (Otto et al., 2013), the modelling analysis found that the η_{AL} parameter was indeed significantly larger than the η_{AR} parameter. Crucially, this 10-parameter model produced near-perfect fits to the data. Subsequent recovery analyses showed that the modelling methodology here was robust and unbiased. Hence, the modelling analysis showed that a race architecture, with simple interactions, could explain audio-visual processing of looming signals.

Owing to the extensive processing time needed for model selection, and the theoretical disagreement over the use of AIC versus BIC for model selection (e.g., Chakrabarti & Ghosh, 2011; Lewandowsky & Farrell, 2011; Raftery, 1999; Weakliem, 1999), both information criteria were used and the results were presented here. Unlike the AIC selection, BIC selected a 6-parameter model with no ρ or η parameters, in effect the independent race model (Raab, 1962) that assumes both statistical independence and context invariance. Such a non-interaction race model is likely too simple, in theory and in practice, as this 6-parameter model did not fit the empirical data closely. Hence, the AIC-selected 10-parameter model was chosen here. However, it is also possible that the BIC underperformed here because neither the ρ nor η parameters were optimal constructs for audio-visual processing of motion. Thus, the BIC, which puts more weight on model simplicity than the AIC, opted out of the ρ and η parameters, even if this meant choosing too simple a model. As a future step, the interaction parameters could be optimised, perhaps as a single interaction parameter that accounts for both RT dependencies and context variance.

6.3 Two discussion points, closing remarks

1. Looming outside of a collision/avoidance context

In this project, looming was presented as a signal of collision or predatory attack, in which a timely avoidance response is needed. The collision narrative of looming seems typical of the literature in this field (e.g., Tyll et al., 2013). Recall from the literature (Chapter 1) that there is a perceptual exaggeration of auditory looming signals (Neuhoff, 1998, 2001), attentional capture by visual looming signals (Franconeri & Simons, 2003; Lin et al., 2008), and underestimation of the time-to-arrival in looming signals (Rosenblum et al., 1993; Schiff & Oldak, 1990); all seemingly to facilitate a quick evasive response, which increases the margin of safety, increasing survival odds. Furthermore, looming signals brought about avoidance and withdrawal responses in humans and primates (Ball & Tronick, 1971; Freiberg et al., 2001; Schiff et al., 1962), reinforcing the notion that looming means collision, and an avoidant response is needed. However, humans (and other animals) evolved as predators. In a hunting context, an object's approach is positive (e.g., successful chase), whilst an object's furthering is a negative (e.g., failed chase). The reverse is true in a collision context (looming is negative; receding is positive). With this context-dependent reversal in the meaning of looming, how might the findings of this project be applicable across both contexts?

The experiments in this project did not specifically discern between collision or hunting looming; participants were not asked to interpret the looming/receding stimuli with a particular context in mind. As a thought, one possibility is that while looming may have different meanings in collision or hunting contexts, the performance requirements are the same across both contexts. In collision looming, quick responses are needed for timely evasion. Intuitively, when hunting, if the prey is looming, then the hunter is closing in, the hunt is likely successful, so there is an impetus for quick responses to secure the hunt's success. Similarly, in both collision and hunting contexts, receding motion signals a de-escalation of the situation: collision unlikely, or prey getting further signifying a failed hunt, so fast responses are no longer necessary. If this thought is true, then the meaning of looming affects the motor output (evade or chase), but not the underlying processes – there would still be the biases for looming (but not receding) signals, in both auditory and visual modalities. Hence, taking that processing may not vary between the collision and hunting contexts, then the multisensory processing, which this project proposes could be explained by an interactive race architecture, would be valid in both collision and hunting contexts.

2. Another look at stimulus realism, impact on modelling

Experiment 2 tested stimulus realism, with seemingly limited effect. Concurrently, there was the realisation that the stimuli inherited from Cappe et al. (2009) may be problematic in featuring onset transients. Hence, the project focused on the removal of onset transients, on basic stimuli only, to avoid making two manipulations (stimulus realism, onset transient removal) at once. At this stage, the basic stimuli seem sufficient for the purposes of this project; the removal of onset transients improved performance, while past studies have used basic looming signals to good effect (e.g., Franconeri & Simons, 2003; Neuhoff, 2001), and there were solid findings here, including a robust model to explain multisensory processing of looming signals. However, now that the problem of onset transients has been resolved by their removal, arguably the question of stimulus realism reopens. On one hand, some past studies have not found greater effects for realistic looming stimuli than basic stimuli (e.g., Bach et al., 2009). On the other hand, the apparent failure of stimulus realism in Experiment 2 could have been self-fulfilling, having implemented 'realism' via minimal cue additions which themselves were not particularly realistic, and still featuring onset transients. There are a few knowledge gaps surrounding stimulus realism; some possibilities for studying stimulus realism further are presented below.

First, a practical matter. True realism may not be achievable in a laboratory setting: visual imagery is fixed at the distance of the display, auditory signals presented through headphones are heard within the head (Paquier et al., 2016). One solution to this practical problem may be a physical setup, where a loudspeaker is physically moved to loom and recede to the participant, the moving loudspeaker acting as both visual stimulation and the source of auditory stimulation (see setup in Neuhoff, 2001). Nonetheless, if a physical setup is impractical, then with standard laboratory equipment, there are still a few study directions to explore more about realism, and possibly learn how realism could be better implemented in future looming studies. To knowledge, in studies examining multiple looming cues (e.g., Bach et al., 2009; Rosenblum et al., 1987), there is not much explanation for their selection of looming cues to test, which cue combinations would be realistic, or whether multiple cues are actually combined and beneficial to sensory processing. For example Baumgartner et al. (2017) found looming biases using only auditory spectral cues, and even argued against the typical auditory intensity change cue. If the auditory spectral cue is so strong, presumably more so than the pseudo-Doppler effect used here in Experiment 2, would the combination of intensity and spectral change have produced more robust auditory processing, that would then feed into audio-visual processing? Yet, is the implication – some cue combinations (e.g., intensity and spectral changes) are stronger than others (e.g., intensity change and Doppler) – plausible? Or instead, if there is a dominant cue, then is it only the dominant cue that drives sensory processing, regardless of other concurrent cues? A future study could look into the processing of multiple concurrent cues, and examine perhaps the behavioural and neural responses (via neuroimaging) to various combinations. If it can be shown that multiple concurrent cues can be combined in sensory processing, in which there is a particularly powerful cue combination, then this could be implemented as a follow-up to Experiment 4, from which the comparative and modelling approaches could again be applied.

On the topic of modelling, if stimulus realism was important to the multisensory processing of audio-visual looming signals, then what are the implications for the current modelling results? Should there be an Experiment 5 using realistic stimuli (using a physical setup, or the optimal cue combinations found empirically), would the same computational modelling technique select a different model? Without significant findings from the existing stimulus realism experiment (Experiment 2), the response at this stage is mostly speculative, and hinges on the currently unknown relevance of stimulus realism.

Known is that the selection of a 10-parameter model (parameterised to account for auditory and visual looming biases, a single negative correlation parameter, and additional noise depending on auditory motion direction) was based on Experiment 4, which used basic stimuli (auditory intensity change, visual size change) to simulate looming and receding. The selection of a model that parameterises the unisensory parameters (on both modalities) to account for processing differences between looming and receding signals suggests that the basic stimuli were at least somewhat valid representations of motion-in-depth, which bodes well for the validity of the 10-parameter model. Furthermore, at least in the auditory modality, more realistic multi-cue looming signals elicit similar responses as basic looming signals do (Bach et al., 2009); if this holds true also in an audio-visual context, then the basic stimuli used here is inconsequential to the modelling. Moreover, even if stimulus realism as such is important, the experiments here test for decision-making towards highly salient and brief motion signals, and the modelling embodies these assumptions; stimulus realism may not have a material impact on the decision-making here. Furthermore, the pattern of faster RT towards looming than receding (the looming bias), faster RT in visual than auditory (giving more weight to the more reliable sensory modality, see also ventriloquism; Ernst and Bülthoff (2004)), seem quite plausible and likely would not change just from adding more cues for realism; without major changes in the RT patterns, the model selected is also unlikely to change. Hence, the computational modelling technique, which was shown to be robust and unbiased (Chapter 5), may still choose the same 10-parameter model even for a hypothetical Experiment 5 using realistic stimuli.

Completing this discussion, perhaps in another experiment paradigm where stimulus realism has a larger potential role (e.g., harder to detect signals obscured by noise, slower detection due to a different sensory information accumulation method), the impact of stimulus realism might be on faster RTs, and a smaller gap between auditory and visual performances. The audio-visual interactions would be key, albeit in a model different to that used here, owing to the different nature of the detection task, which is beyond the scope of the current project.

Closing Remarks

This project has shown that selective integration is not necessary to explain the processing towards audio-visual looming signals. From the analysis on RSEs, to a comparative approach against a simple race model, to a comprehensive computational modelling approach, the evidence suggests that the race architecture is a potential framework in explaining multisensory processing of motion-in-depth signals. Computational modelling is a novel and powerful analytical approach (e.g., Lewandowsky & Farrell, 2011), and here it was shown that a race model with simple interaction parameters (none of which suggesting audio-visual congruency or special processing towards ALVL) explains audio-visual processing of looming signals. In previous studies of other experimental paradigms, the selective integration mechanism for processing audio-visual looming signals was also rejected (Huygelier, van Ee, Lanssens, Wagemans, & Gillebert, 2021). Looking at the bigger picture, this project is further evidence that the race architecture with interactions fits with the RSEs obtained in sensory decision-making experiments (see also Mercier & Cappe, 2020 as another recent example). This project, among other research (e.g., Townsend et al., 2020), adds to the evidence that the race architecture with interactions allowed could be a powerful explanatory framework in multisensory decision-making.

Section 2: Short-term monocular patching pilot study

Chapter 7: General methods of the short-term monocular patching pilot study: test phase

In the short-term monocular patching pilot study, every experiment consists of two phases: an eye patching phase, and a test phase. The eye patching phase differs according to the needs of the experiment, and will be detailed in the relevant chapters. Participants were also different for each experiment, so the participant particulars are also detailed in the relevant chapters.

The test phase is constant for all experiments. In this chapter, the apparatus for the test phase will first be described (Chapter 7.1), then an overview of the general procedures (Chapter 7.2), and finally an in-depth explanation of the three visual tests: binocular rivalry (Chapter 7.3), contrast letters (Chapter 7.4) and stereopsis (Chapter 7.5).

7.1 Apparatus

The apparatus for this project centres around the VIEWPixx /3D Lite (VPixx Technologies Inc., referred to as VIEWPixx hereafter), a 3D-capable display. To operate the VIEWPixx, a system of hardware is necessary (Figure 7.1.1). Below, the apparatus are each explained, organised as functional categories: computer (Chapter 7.1.1), display sub-system (Chapter 7.1.2), and response sub-system (Chapter 7.1.3).

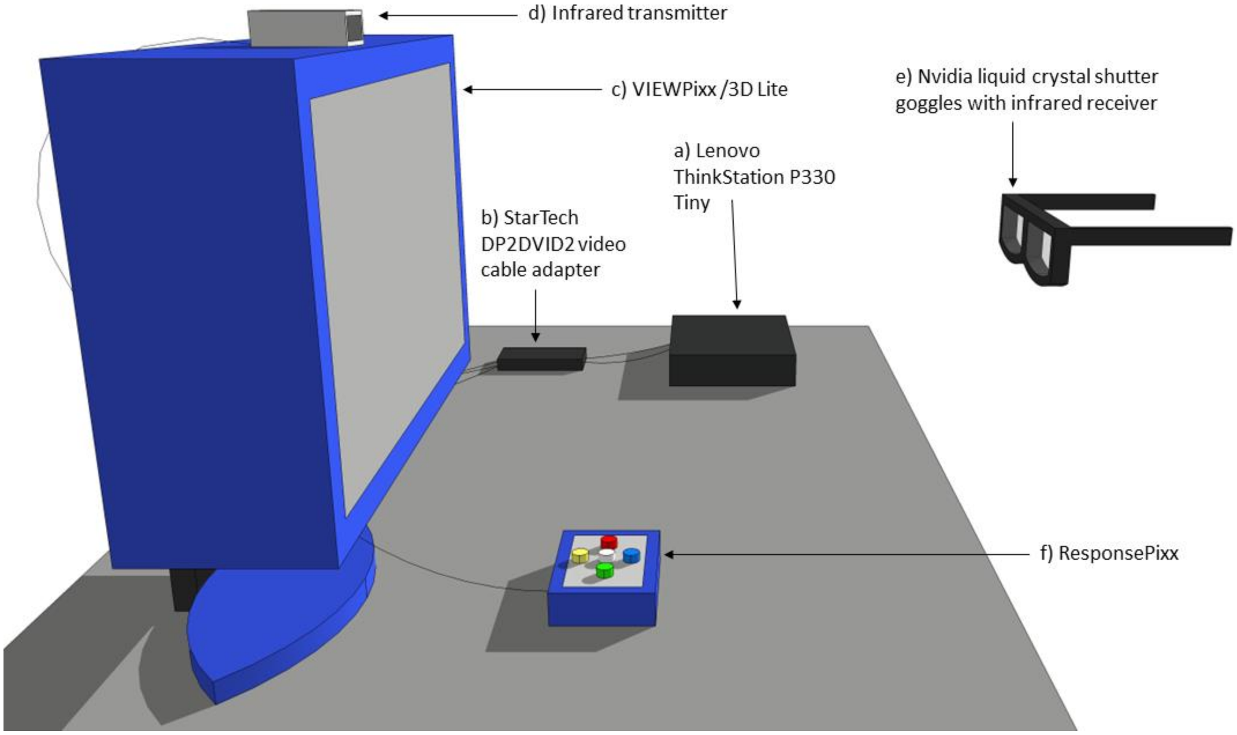

d) Infrared transmitter
c) VIEWPixx /3D Lite
e) Nvidia liquid crystal shutter goggles with infrared receiver
a) Lenovo ThinkStation P330 Tiny
b) StarTech DP2DVID2 video cable adapter
f) ResponsePixx

Figure 7.1.1. The apparatus for the short-term monocular patching pilot study. a) Lenovo ThinkStation P330 Tiny, the computer controlling the apparatus, and running the experimental script. b) StarTech DP2DVID2, a USB-powered video cable adapter which converts the computer’s native DisplayPort 60Hz video output to DVI-D 120Hz for use with the VIEWPixx. c) VIEWPixx /3D Lite, a 3D-capable LCD display with a refresh rate of 120 Hz. d) Infrared transmitter, which communicates with the e) Nvidia liquid crystal shutter goggles, which gates the interleaved images for presentation to each eye. f) ResponsePixx, a handheld button box with five illuminated colour buttons, to make responses on. Note: figure not to scale, researcher’s display, mouse and keyboard not shown.

7.1.1 Computer

A desktop workstation (Lenovo ThinkStation P330 Tiny, Figure 7.1.1a) running Windows 10 (Microsoft Corporation), Matlab 2019b (The MathWorks, Inc.) and Psychophysics Toolbox Extension version 3.0.16 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) was used to conduct the experiments in this project. The computer had a dedicated graphics card (Nvidia Quadro P620, set to built-in 3D profile for optimal performance) outputting a 60Hz video signal through the mini-DisplayPort protocol. This 60 Hz video signal was doubled to 120Hz and converted to the dual-link DVI protocol using a USB-powered signal-repeating video cable adapter (StarTech

DP2DVID2, Figure 7.1.1b), for use with the VIEWPixx. Prior to experimentation, the compatibility of this computer system with the VIEWPixx was checked using the Psychophysics Toolbox Extension test scripts *BitsPlusImagingPipelineTest* and *BitsPlusIdentityClutTest*. The computer system passed both tests, and the optimal settings obtained during the tests were saved for use in experiments.

7.1.2 Display sub-system

The VIEWPixx (Figure 7.1.1c) features a 24-inch (diagonal) LCD panel with a resolution of 1920 (horizontal) by 1080 (vertical) pixels. The key feature of this display panel is its 120Hz refresh rate (i.e., 120 images per second), which is double that of ordinary computer displays. This rapid succession of images is then used to interleave images for the left and right eyes (i.e., image 'A' and image 'B', Figure 7.1.2a), either for binocular rivalry, or 3D presentation.

If one were to look at the 120Hz interleaved images with the naked eye, one would see a fusion of both images (Figure 7.1.2b). Crucially, the interleaved images need to be gated, such that each image is only presented to one eye (i.e., image 'A' only to the left eye, image 'b' only to the right eye; Figure 7.1.2c). Image gating is achieved using the system of infrared transmitter (Figure 7.1.1d) and a pair of liquid crystal shutter goggles (Nvidia, Figure 7.1.1e, termed '3D goggles' hereafter). The 3D goggles resemble sunglasses, but in fact the two eyepieces can individually and rapidly turn from transparent to opaque, and vice versa, using electricity. As an example, when image A is presented onscreen, the infrared transmitter communicates with the 3D goggles to make the right eyepiece opaque, the left eyepiece transparent, such that image A is only presented to the left eye (Figure 7.1.2c). Similarly, when image B is presented onscreen, the infrared transmitter communicates with the 3D goggles to make the left eyepiece opaque, the right eyepiece transparent, such that image B is only presented to the right eye. In this system of VIEWPixx and 3D goggles via infrared communication, the 120Hz interleaved images were separated into two 60Hz channels for each eye. There are two main uses for differential image presentation to each eye: binocular rivalry using two incompatible images for each eye (see also Chapters 7.3 and 7.4), and 3D presentation by interleaving two images that are parallax-shifted versions of each other (see also Chapter 7.5).

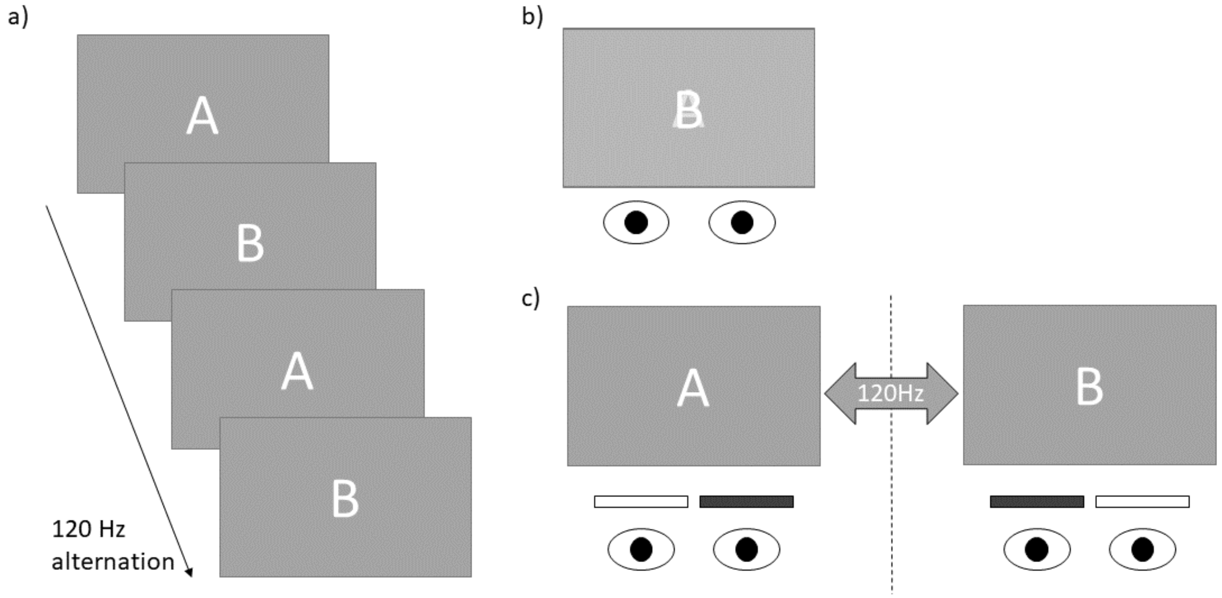


Figure 7.1.2. The principles of the display sub-system, for the presentation of binocular rivalry or 3D imagery. a) Two images, ‘A’ and ‘B’ in this Figure, are interleaved in presentation on the VIEWPixx display, at the VIEWPixx’s 120Hz refresh rate. b) Viewing the 120Hz interleaved images with the naked eye, one would only see a fusion of both images. c) To perceive binocular rivalry or three dimensional imagery, the interleaved images need to be gated for each eye, i.e., image ‘A’ to the left eye only, image ‘B’ to the right eye only. Image gating was achieved using the 3D goggle system, which is synchronised with the imagery on the VIEWPixx via infrared communication.

As an evaluation, this image-interleave technique using a 3D display system is not the only technique for binocular rivalry and 3D presentation. The image-interleave technique potentially suffers from cross-talk due to imperfect gating, resulting in traces of the left-eye image also visible to the right eye, and vice versa, degrading stereovision and interfering with binocular rivalry by encouraging binocular alternations and fusion (Baker, Kaestner, & Gouws, 2016; Woods, 2012). Crosstalk typically originates in either the 3D goggles (e.g., the eyepiece is not 100% opaque when required to obstruct the view, the goggles not in perfect synchrony with the display) or the display latency, resulting in image remnants visible to the non-intended eye (Baker et al., 2016; Woods, 2012). Crosstalk is an inherent issue of 3D displays, but first, the combination of VIEWPixx and Nvidia 3D goggles produced some of the lowest luminance crosstalk in comparison to other systems (although there were some contrast crosstalk artefacts; Baker et al., 2016). Second, the alternative, the mirror-stereoscope technique, which uses two displays to present different imagery to the participant’s eyes via a mirror array (Baker et al., 2016; Carmel, Arcaro, Kastner, & Hasson, 2010; see Figure 7.1.3), has its own challenges in setup and operation. With the mirror-stereoscope, due to individual differences, setup and adjustment is specific to each participant, and head stabilisation is required to maintain alignment with the image pathways (Carmel et al., 2010). The second alternative is virtual reality headsets (Baker et al., 2016), which avoids crosstalk by using physically separate image pathways, but these were not available to use. On balance, the VIEWPixx system was easy to setup and use, and is most likely adequate for the purposes of this study.

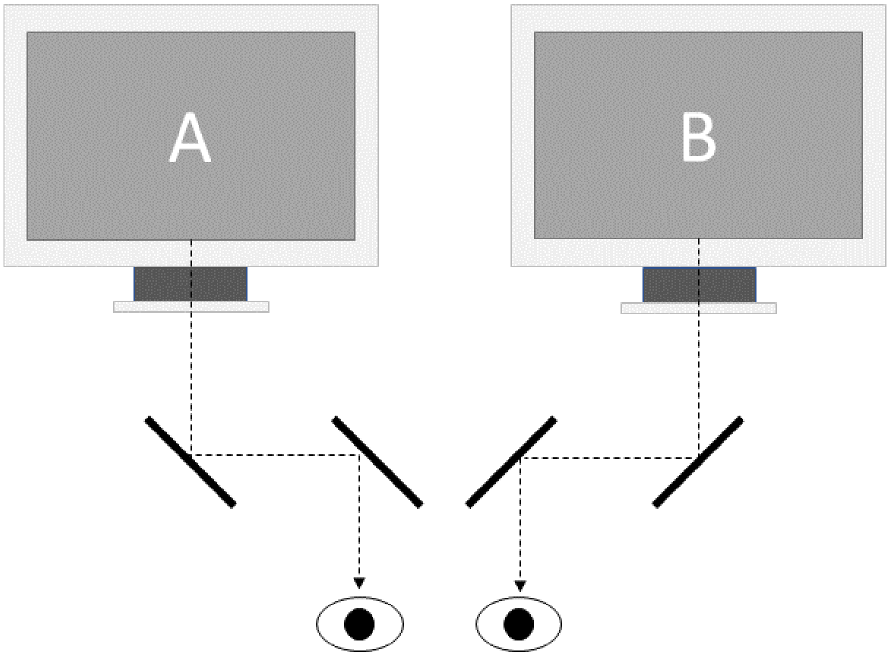


Figure 7.1.3. The mirror stereoscope setup. Two images are each presented on separate displays, and brought together to view using an array of mirrors (angled thick black lines). The image pathways are shown with the dotted arrowed lines. The physically separate image pathways for each image should avoid crosstalk, but its use requires head stabilisation and many participant-specific adjustments (Baker et al., 2016; Carmel et al., 2010; Woods, 2012).

7.1.3 Response sub-system

The response sub-system consists of ResponsePixx and Datapixx. ResponsePixx is a handheld button box with five illuminated colour buttons (green, yellow, red, blue, white; Figure 7.1.4) which the participant presses to make a response. The ResponsePixx is wired to the Datapixx enclosed within the VIEWPixx unit. Control of the response sub-system is performed through Datapixx commands, such as switching the system on, and timestamping responses incoming from the ResponsePixx. Below, more details about the response sub-system are explained through their implementation for this study.

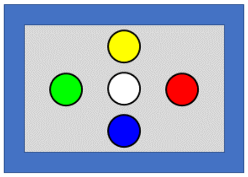


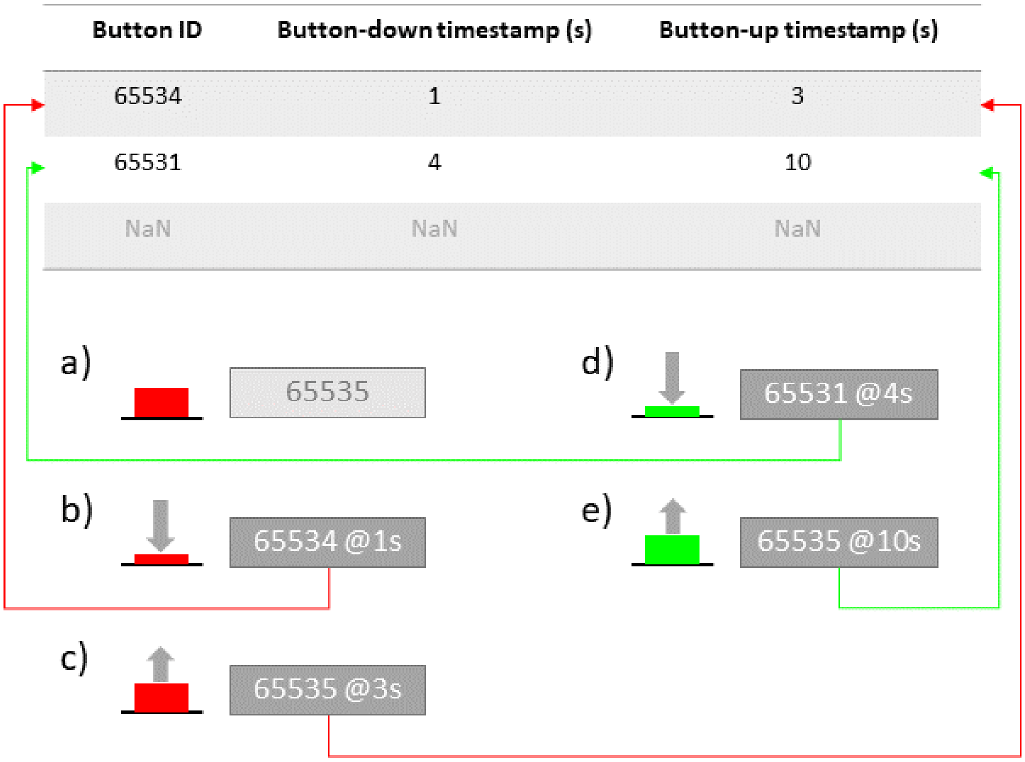
Figure 7.1.4. Diagram of the ResponsePixx handheld button box. There are five selectively illuminated buttons: white, green, yellow, red and blue. Only the buttons valid for the visual test were illuminated.

First, in the experimental script, one of the first steps is to set a time marker on the Datapixx. The time marker is the anchor point from which all button events are timestamped within the response sub-system. The time marker was set right before stimulus presentation began.

Button	**Status Code**
Green	65531
Yellow	65533
Red	65534
Blue	65527
White	65519

Table 7.1.1. Table of status codes for each of the five buttons when pressed.

Each button press produces a status code unique to the button (Table 7.1.1). When no buttons are pressed, the status code is 65535. The status code *changes* are checked at a frequency of 120 Hz, e.g., from 65535 (no buttons pressed) to 65534 (red button pressed), and this button event is recorded. Below, the exact utilisation of this button event system is explained, for each visual test.



Button ID	Button-down timestamp (s)	Button-up timestamp (s)
65534	1	3
65531	4	10
NaN	NaN	NaN

Figure 7.1.5. Schematic of the recording of button press events, for the binocular rivalry visual test. a) When the visual test starts, no buttons are being pressed, so the status code is 65535. b) One second in, the red button is pressed: the status code 65534 and its timestamp of 1 second is recorded in a single row of button event log. c) Three seconds in, the red button is released, meaning the status code changes back to 65535; the timestamp of this change is recorded in the final column on the current row, in the button event log. d) Later, the green button was pressed: the status code of 65531 and the timestamp of 4 seconds is recorded on a new row in the button event log. e) When the green button is released, the status changes back to 65535, and the timestamp of this change is recorded, to complete the button event entry.

Binocular rivalry visual test

For the binocular rivalry visual test (Chapter 7.3), the response duration is one minute, during which the participant can freely press the designated response button, as they see fit. Hence, for this visual test, a 500-row log was pre-loaded in the experimental script, to record each button event (status code, button-down time, button-release time; see schematic in Figure 7.1.5). At first, no buttons are being pressed, so the status code is 65535 (Figure 7.1.5a). When a new button event (non-65535) was detected, then a button is being pressed, and the experimental script records the new status code (the identity of the button), and its timestamp, on the same row (Figure 7.1.5b). When the button is released, the status code changes back to 65535, and the timestamp of this button release is recorded on the current row of the button event log (Figure 7.1.5c). This process occurs anew for each button press, on a new row of the button event log (Figure 7.1.5d, e). After the visual test, the duration of each button press can simply be calculated by subtracting the button-down timestamp from the button-up timestamp.

Contrast letters and stereopsis visual tests

The contrast letters and stereopsis visual tests (Chapters 7.4 and 7.5 respectively) both operate on a multi-trial, staircase principle. On each trial, only a single button press is needed, without the need for timestamping. Hence, for these visual tests, on each trial, only the first change in status code from 65535 is taken, and interpreted as the response.

Evaluation

A question one may have is the necessity of using specialist equipment such as the ResponsePixx and Datapixx, when standard computer input devices like a keyboard has the same function. The answer is that standard computer input devices are not designed for accurate and precise timing, the keyboard in particular would introduce a timing variability of up to 70 milliseconds (Li et al., 2010). In the audio-visual looming project, the problem of inaccurate keyboard timing was addressed by using the RTbox v5/6, a specialised response-collection apparatus with its own precise button mechanism, logic board and accurate onboard clock (Li et al., 2010). ResponsePixx and Datapixx fulfils a similar role: the ResponsePixx has dedicated buttons with tactile actuation, and timestamping is managed by Datapixx. Only the binocular rivalry test (Chapter 7.3) needs precise response timing. However, as the three visual tests were conducted one after another, it was convenient to have only one response device to handle – no need to juggle devices, and less surfaces to touch and clean. Also, only the ResponsePixx buttons valid for the visual test were illuminated, to minimise button confusion. A typical keyboard has over 100 keys, none illuminated, making them difficult to use in the dimly lit experiment room.

7.2 General procedures

7.2.1 Participants

Participants were recruited by word-of-mouth, and through advertisements emailed to general addresses of eligible groups. Due to then-current covid rules, only University of St Andrews' research postgraduates and staff were eligible to participate, as they had access to the University buildings.

For people who replied with interest to the advertisements, an email was sent, containing the information sheet, consent form, an optional participant questionnaire, and a coronavirus declaration form. These documents were so that the potential participant had information about the experiment and the risks involved, such that they could provide informed consent. To proceed with participation, the participant needed to complete and return the consent form and coronavirus declaration form. An experiment time was then arranged with the participant. This remote information and consent is a departure from my usual practice of verbally explaining and letting the participant try the experiment first before asking for consent, but it was necessary to minimise face-to-face contact at the time.

On the day of the experiment session, just before the session began, the participant was still given the opportunity to try the visual tests first, and the researcher gave instructions if necessary. Before experimentation, participants practiced on a demo version of the visual tests, which were shorter in duration or trials compared to the experiment version. The demo version does not collect data. The practice runs were an opportunity for the participant to become familiar with the VIEWPixx viewing experience, and the response logic using the buttons on the ResponsePixx. Any experimental issues the participant may have could be flagged up here, and addressed. The experiment began once the participant indicated that they were satisfied with the experimental arrangement and were ready to begin.

After the experiment session, the participant was emailed a debrief form, and the room and apparatus were sanitised to comply with then-current covid protocols. Personal data collected in the coronavirus declaration form for track-and-trace was deleted 28 days after the experiment session.

All procedures complied with the *Code of Human Research Ethics* (The British Psychological Society, 2014). The experiment has received internal ethical approval from the *University Teaching and Research Ethics Committee* (UTREC, University of St Andrews, PS14148, Appendix D).

7.2.2 Experimental arrangement

The aim of this pilot study was to explore how monocular patching might change visual functioning. Thus, as the basic principle in the experiments, visual functioning was measured before and after monocular patching. Visual functioning was measured using three visual tests: binocular rivalry (Chapter 7.3), contrast letters (Chapter 7.4) and stereopsis (Chapter 7.5). These three visual tests were grouped into a battery (Figure 7.2.1).

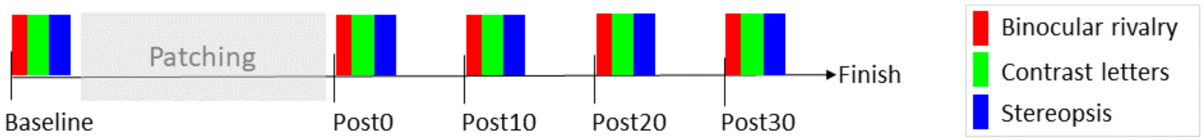


Figure 7.2.1. The visual test battery, consisting of the binocular rivalry visual test (red), the contrast letters visual test (green), and the stereopsis visual test (blue). The visual test battery was conducted once before patching, and four times after patching at 10 minute intervals.

Pre-patching, the experiment session began with a battery of the visual tests, to take a baseline reading of the participant’s visual functioning (Figure 7.2.1). After the baseline visual test battery, the participant removed the 3D goggles, and put on the translucent eye patch to watch their choice of television program (the exact duration and method of patching depends on the experiment; see Chapters 8 to 10). The time spent watching television was kept by the researcher. Once the time for watching television was up, the participant was asked to remove the eye patch, and put on the 3D goggles to do the visual test battery at specific timepoints.

Post-patching, the participant did the visual test battery immediately after (‘Post0’), 10 minutes after (‘Post10’), 20 minutes after (‘Post20’), and 30 minutes after patching (‘Post30’) – the researcher kept time. Thus, there were measurements of the participants’ visual functioning at ten minute intervals post-patching. As the battery of visual tests typically takes around seven minutes to complete, the participant had around three minutes of rest between the visual test batteries (Figure 7.2.1). With 30 minutes of patching, the experiment session in its entirety took approximately 90 minutes to complete. With 10 minutes of patching, the experiment session took approximately 70 minutes to complete.

The visual test battery put the binocular rivalry as the first visual test, followed by the contrast letters, and lastly the stereopsis visual test (Figure 7.2.1). Binocular rivalry was first in the test battery because a) this test has a fixed duration of one minute and b) it is the test Lunghi and colleagues used (Lunghi et al., 2015a; Lunghi et al., 2011; Lunghi et al., 2013; Lunghi et al., 2015b; Lunghi & Sale, 2015), so it was given priority to start on time. The contrast letters and stereopsis visual tests come afterwards, because both tests were exploratory in the sense that they have not been used before to measure the effects of monocular patching, and both tests are variable in duration (both use an adaptive staircase).

7.3 Visual test – Binocular rivalry

Binocular rivalry is when the observer is presented with incompatible images across the two eyes, thus producing visual conflict. Like other forms of visual conflict, binocular rivalry does not result in a stable combined percept of both images; instead there are transient periods of perceiving one view over the other, in a dominance-suppression dynamic which alternates between the two eyes (Blake & Logothetis, 2002; Carmel et al., 2010). In this visual test, binocular rivalry was implemented with Gabor patches at conflicting orientations between the eyes; dominance was measured using percept durations (e.g., Blake & Logothetis, 2002). More details on the stimuli, procedures and data analysis below.

7.3.1 Stimuli

The stimulus consisted of two parts: the Gabor patch, and a white ring encircling the Gabor patch. All parts were presented against a colour-neutral mid-grey background, making the light and dark of the Gabor patches equally visible. In more detail, the Gabor patch was an area where, in one spatial direction, a band of white gradually turned into a band of black, then gradually to white, and so forth (i.e., a sinusoidal wave; see Figure 7.3.1). The Gabor patch was generated using Psychtoolbox-3 commands 'CreateProceduralGabor' and 'Screen('DrawTexture',…)', which together, produced a Gabor patch over 1000 pixels square (on the 1980 x 1020 VIEWPixx display), 0.8 wave amplitude, and 7 black-white cycles. To produce binocular rivalry, the left-eye Gabor was oriented at 45° clockwise from horizontal (9 o'clock position, '\'), while the right-eye Gabor was oriented at 135° clockwise from horizontal (9 o'clock position, '/'). These incompatible Gabor patches were interleaved at 120Hz, and gated to each eye using the system of VIEWPixx and 3D goggles (see Figure 7.3.1, also Chapter 7.1). As a side note, rivalry gratings are commonly horizontal and vertical (e.g., Blake & Boothroyd, 1985; Carmel et al., 2010; Lunghi et al., 2011), but here, orthogonal-diagonals were used, to avoid potential artefacts during binocular rivalry associated with visual system preferences towards horizontals and verticals. The set of 45° and 135° rivalry gratings has been used before (e.g., Dieter, Sy, & Blake, 2017; Lunghi et al., 2013).

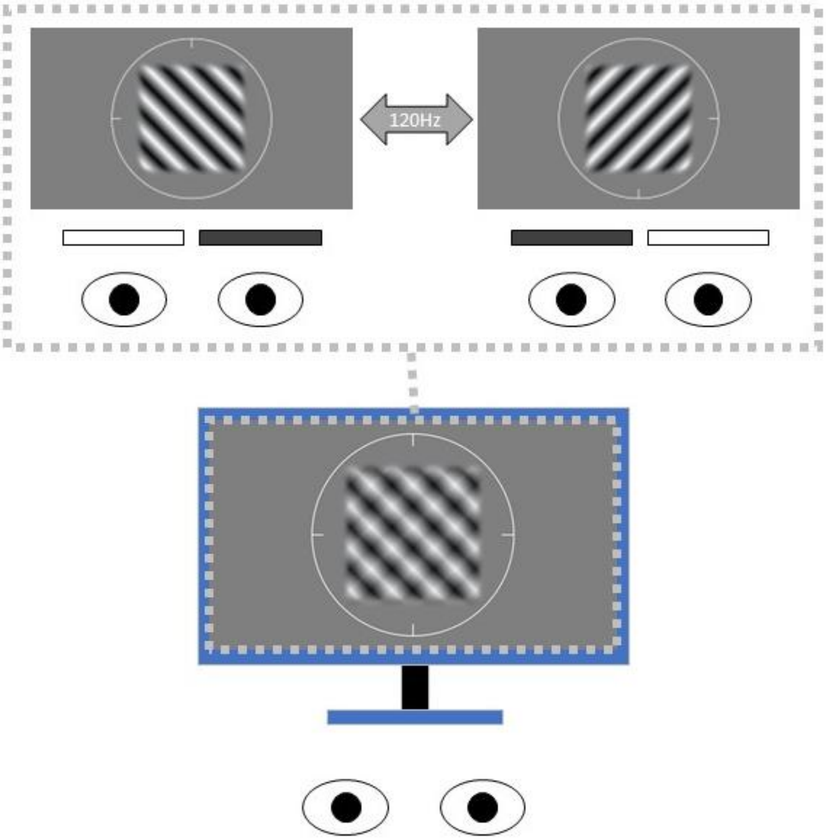
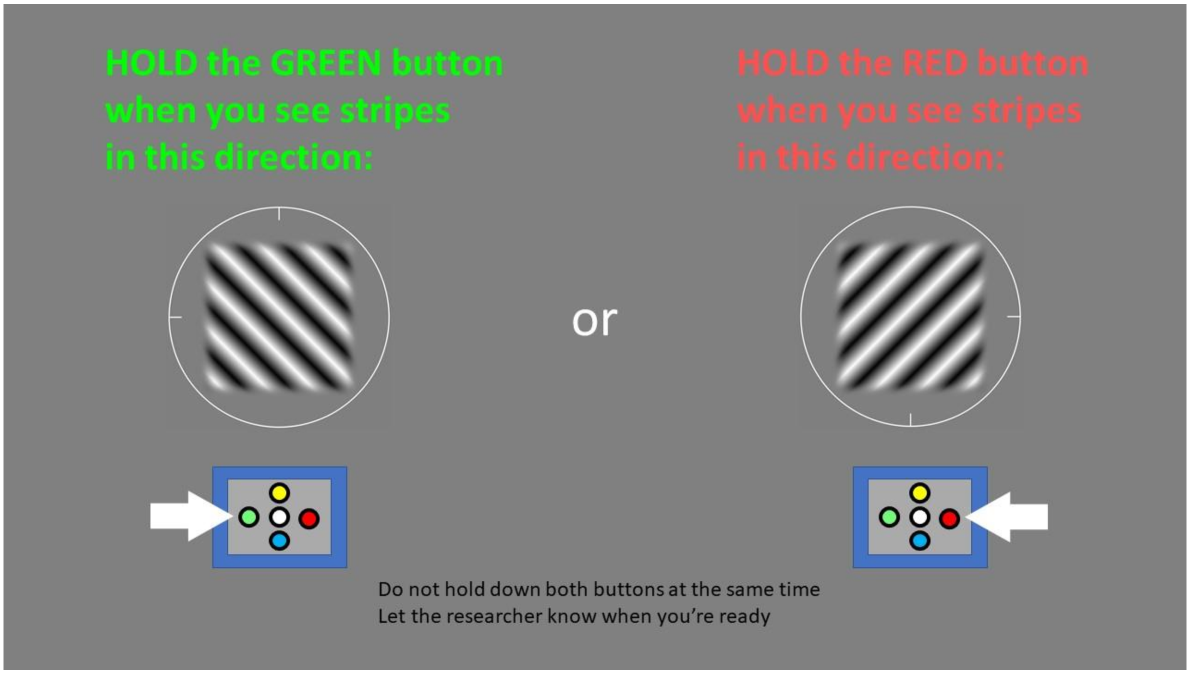


Figure 7.3.1. Presentation of conflicting Gabor patches between the eyes, using the system of VIEWPixx and 3D goggles. The 45° Gabor (left eye) and 135° Gabor (right eye) were interleaved at 120Hz, and gated to each eye using the 3D goggles. The presentation of conflicting Gabor patches produces binocular rivalry.

Second, the white ring encircling the Gabor patches. Both the left-eye and right-eye Gabor patches were drawn centred to their respective image areas, but it is not a given that the two Gabor patches will therefore align in perceptual space during binocular rivalry. The issue is that with conflicting imagery between the eyes, the visual system does not have a basis to maintain binocular alignment on the imagery (vergence; Carmel et al., 2010). To promote stable vergence onto the rivalrous Gabor patches (i.e., to keep both Gabor patches overlapping and therefore in binocular rivalry), a white ring was drawn around each Gabor patch, as a salient marker common between both eyes, which helps anchor both eyes' views together in alignment (Carmel et al., 2010; Lack, 1974). This white ring was used as a vergence device because it was not situated within the Gabor patch itself (Figure 7.3.1); the alternative of using a centrally-positioned vergence device on the

Gabor patches (e.g., a fixation cross) would promote binocular fusion in that visual area (Blake & Boothroyd, 1985), thus interfering with the rivalry on the Gabor patches.

7.3.2 Procedures



HOLD the GREEN button
when you see stripes
in this direction:

or

HOLD the RED button
when you see stripes
in this direction:

Do not hold down both buttons at the same time
Let the researcher know when you’re ready

Figure 7.3.2. Instruction screen before the visual test began. Using text and graphics, the participant was reminded of the experimental task: indicate which percept was seen, by pressing and holding onto the relevant button for the duration of the percept.

First, the participant puts on the 3D goggles and switches it on, if the participant has not already done so. Before the visual test began, an instruction screen was shown (Figure 7.3.2). The instruction screen explained using graphics and text that, for as long as a striped pattern in a ‘\’ orientation was seen, then the green button on the Responsepixx should be pressed and held, while for as long as a striped pattern in the ‘/’ orientation was seen, the red button on the Responsepixx should be pressed and held. There was no mention that the two Gabor patches were presented simultaneously, one to each eye, and that it was a visual test involving binocular rivalry. Also not mentioned was that one may see a fusion of both striped patterns. To simplify task demands, in this visual test’s response logic, only the left- or right-eye view needs to be indicated. Apart from left- or right-eye views, fusion is the only other possibility, so the absence of button press was taken as fusion. A further instruction on the instruction screen was that the two response buttons should not be held down together at the same time.

When the participant was ready, they indicated this to the researcher, who then started the visual test. For one minute exactly, the two Gabor patches were presented in binocular rivalry, and the participant gave button presses to indicate their percept. At the end of the minute, the Gabor patches were replaced by a screen informing the participant that the visual test has finished. The experiment code then moves onto the next visual test in the battery, the contrast letters visual test (Chapter 7.4).

7.3.3 Data collection

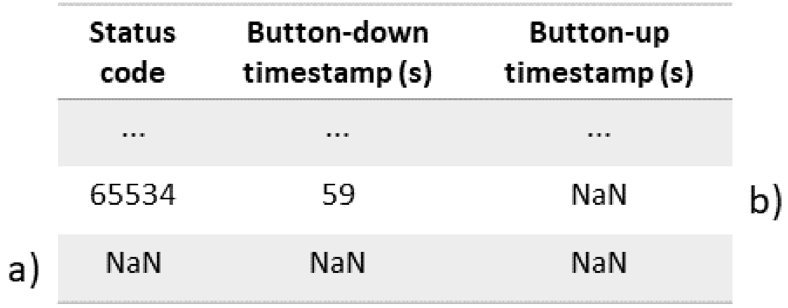
During the one-minute of binocular rivalry, all the button events were collected in the button event log, a three-column Matlab matrix, where each row is one button event. The first column was the status code (i.e., which button was pressed), the second column was the button-down timestamp, and the third column was the button-up timestamp. The matrix was predetermined with 500 rows, to allow for a maximum 500 button events during the one minute – a limit unlikely to be reached.

7.3.4 Data tidying

Tidying the button event log

Directly from the visual test was the button event log, in all its 500 rows. Before this data can be used, some tidying was required. There were three types of data tidying, performed in sequence: tidying the button event log, processing the button identities, and calculating the duration of each button press event.

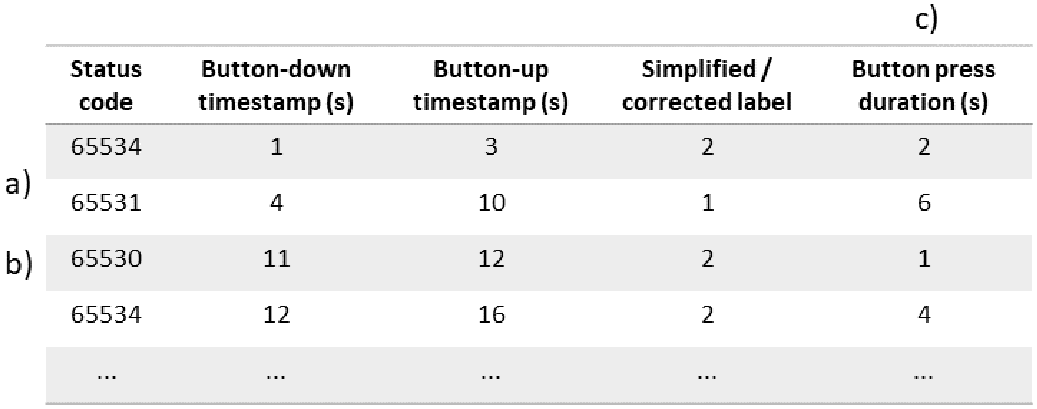
There were two steps in tidying the button event log. First, the button event log was pre-configured with 500 rows, allowing for 500 button events. However, binocular rivalry only lasted for one minute, and it is unlikely that all 500 rows were used. Hence, all the unused rows (filled with NaN placeholders) were removed (Figure 7.3.3a). Second, in case the last button event was cut off by the time limit, then that button press event would not have a button-up timestamp (an incomplete row, with a placeholder NaN on the final column; Figure 7.3.3b). As such, the duration of the final button event cannot be calculated. The remedy for a cut-off button event was to use the time limit (60 seconds) as the button-up timestamp. An alternative solution for a cut-off button event is to extend the response window until the last button press is released (as implemented by Dieter et al., 2017), however this solution comes with the complication of having irregular test durations.



Status code	Button-down timestamp (s)	Button-up timestamp (s)
...	...	...
65534	59	NaN
NaN	NaN	NaN

Figure 7.3.3. The raw button event log, with 500 rows, unlikely to all be used. The unused rows need to be trimmed. The trimming steps work at the end of the log, so here in this Figure, the earlier entries are represented with ellipses. a) The first row of placeholders (NaN; unused row) was identified, and together with the rows below (all NaN), were all deleted from the button event log. b) In case the final button event was cut by the time limit, that button event would be missing a button-up timestamp (i.e., filled with a NaN placeholder). The solution was to replace the NaN with the time limit of the visual test (60 seconds).

Processing the button identities



Status code	Button-down timestamp (s)	Button-up timestamp (s)	Simplified / corrected label	Button press duration (s)
65534	1	3	2	2
65531	4	10	1	6
65530	11	12	2	1
65534	12	16	2	4
...	...	...	...	...

Figure 7.3.4. The dataset also needs to be simplified, corrected, and processed. Ellipses in the last row represent further button presses. a) The raw status codes were re-labelled in a new column, using an easier coding system (‘1’ for green, ‘2’ for red). b) Simultaneous red-green button press produces a status code of 65530. This 65530 code is relabelled as green (‘1’) or red (‘2’) depending on the button press immediately before it. In this example, as the previous button event was for the green button, the 65530 simultaneous button event was interpreted and recoded as red (‘2’). c) Once simplification and re-coding was complete, the duration of each button event was calculated, by subtracting the button-down timestamp from the button-up timestamp.

To process the button identities, there were another two steps. First, the raw status codes 65531 (Green) and 65534 (Red) are unintuitive, and may cause errors during manual processing and analysis. Hence, in a new column on the event log, the status codes were relabelled: 65531 (Green) became ‘1’, 65534 (Red) became ‘2’, which, intuitively for the author, corresponds to the green button being on the left, the red button being on the right of the ResponsePixx (see the two examples in Figure 7.3.4a).

Second, despite clear instructions not to simultaneously press both response buttons, simultaneous green-red button presses occurred. To this visual test’s response logic, the simultaneous green-red button press is non-sensical. Yet, the simultaneous green-red button press has a unique status code of 65530, with timestamps, and is recorded in the button event log. One interpretation for a 65530 event is that in switching between the two valid response buttons, there was a slight temporal overlap. Referring to the ResponsePixx button layout, the green button is on the left, while the right button is on the right. Naturally, one holds the ResponsePixx in both hands, with thumbs on either response button, so the occasional temporal overlap in button press is perhaps unavoidable. Taking that 65530 events were indeed accidental temporal overlaps when switching between response buttons, the 65530 events were re-coded depending on the button event immediately prior. If the prior button event was for a green button, then the simultaneous press was likely the temporal overlap in switching to the red button, so 65530 was re-coded as red ('2'; Figure 7.3.4b). If the prior button event was for a red button, then the simultaneous press was likely the temporal overlap in switching to the green button, so 65530 was re-coded as green ('1'). Altogether, the two processing steps on the button press identities yield an entirely new column with simplified and corrected labels for the button pressed.

Calculating button event durations

In the final step, the button event durations were calculated. In the button event log, each row is a button event, containing button-down and button-up timestamps. To calculate the button event duration, the button-down timestamp was subtracted from the button-up timestamp. This calculation was performed for all button events. The button event durations were placed in a new column on the button event log (Figure 7.3.4c).

7.3.5 Data analysis

Using the button event durations as the basis, this data analysis stage determined eye dominance using two metrics. Two metrics of eye dominance were identified in past research on binocular rivalry (e.g., Dieter et al., 2017; Lunghi et al., 2011). Owing to the exploratory nature of this pilot study, both metrics were included, to determine if and where their similarities and differences are.

The first metric, termed dominance ratio, directly compares the duration indicated for each eye. In detail, the individual event durations for the red button (i.e., indicating the right-eye Gabor) were summed together, and similarly for the green button (i.e., indicating the left-eye Gabor). The dominance ratio is therefore the total right-eye duration (red) divided by the total left-eye duration (green). If there was neutral binocular balance, the total durations would be the same between the two eyes, so the ratio would be one. If there was a right-eye dominance, the total duration to the right-eye would be longer than to the left-eye, so the ratio would be larger than one. Similarly, if the dominance was on the left eye, then the ratio would be smaller than one. The ratio was constructed in this fashion to put right-eye dominance on a positive scale, because the right-eye was patched, and it was the patched eye that became dominant after short-term monocular patching (e.g., Lunghi et al., 2011). The dominance ratio metric was taken from Lunghi et al. (2011), who found a dominance ratio of up to 2.6 to the patched eye.

The second metric of eye dominance, termed dominance proportion here, uses the proportion of time indicated for each eye, and was taken from Dieter et al. (2017). For this pilot study, the interest is on the patched right eye, so the dominance proportion calculation was formulated as the proportion of total time indicating the right eye, minus the proportion of total time indicating the left eye, all divided by the proportion of total time spent indicating (Equation 7.3.1; Dieter et al., 2017).

$$\frac{Right\ eye_{proportion} - Left\ eye_{proportion}}{Right\ eye_{proportion} + Left\ eye_{proportion}} \times 100$$

Equation 7.3.1. Formula for calculating eye dominance. Formula taken from Dieter et al. (2017).

In addition to the two measures of eye dominance described above, the button event durations also contain information on the time spent perceiving fusion. During binocular rivalry, there may be brief moments where both the left- and right-eye views are simultaneously perceived. Perceptual fusion is thought to be linked to binocular suppression, with longer fusion possibly suggesting weak binocular suppression, hence weak dominance (Dieter et al., 2017). In the dataset, the fused percept was determined from the periods where no buttons were pressed (i.e., neither the left- nor right-eye Gabor were indicated). Each period of perceptual fusion were summed together to determine the duration of perceptual fusion during the one minute of binocular rivalry.

Altogether, three metrics were applied to each test result, as a wide-ranging exploration of the effects during binocular rivalry for this pilot study. Hence for each participant, there were metrics of binocular dominance and suppression over the course of the experiment (Baseline, Post0, Post10, Post20, Post30). The progression of these metrics were analysed using one-way ANOVAs and paired-samples t-tests, with the alpha level at 0.05.

7.4 Visual test – Contrast letters

The contrast letters visual test was another technique for measuring eye dominance. The rationale for including this test, in addition to the binocular rivalry visual test, is multi-fold. On one level, the originators of the contrast letters visual test have claimed that it is a robust measure of eye dominance (Bossi et al., 2018; Bossi et al., 2017). On another level, there are potential issues with the binocular rivalry visual test. The binocular rivalry visual test asks the observer to make button presses in response to a momentary percept. First, volitional reports on such a momentary visual percept may not be accurate, reliable or repeatable (Aleshin, Ziman, Kovács, & Braun, 2019). Second, there is no way of knowing if the observer responds true to their percept. Third, with extended periods of binocular rivalry, the observer can actively influence the percept they perceive (Blake & Logothetis, 2002; Lack, 1974) – from personal experience, during the one minute of binocular rivalry, with the intention and some effort, it was indeed possible to hold percept onto a Gabor patch for several seconds at a time, rather than allowing the natural course of percept alternation. These potential issues could distort the determination of eye dominance. Altogether, the contrast letters visual test, by using a different test paradigm to the binocular rivalry visual test, is both interesting to include in this explorational pilot study, and necessary as a backup measure of eye dominance in case issues arise with the binocular rivalry visual test.

7.4.1 Stimuli

In basic terms, the stimuli consists of two images, one for each eye. Both images contain the letters ‘Y’ and ‘B’ arranged vertically in the middle. There is a white frame surrounding the letters, operating as a vergence device to maintain binocular alignment on the images (Carmel et al., 2010; see also Chapter 7.3.1). All these elements were on a neutral grey background (50 cd/m^2). The two images are interleaved and presented on the VIEWPixx at 120Hz, with the system of 3D goggles gating the letters image to each eye (Figure 7.4.1).

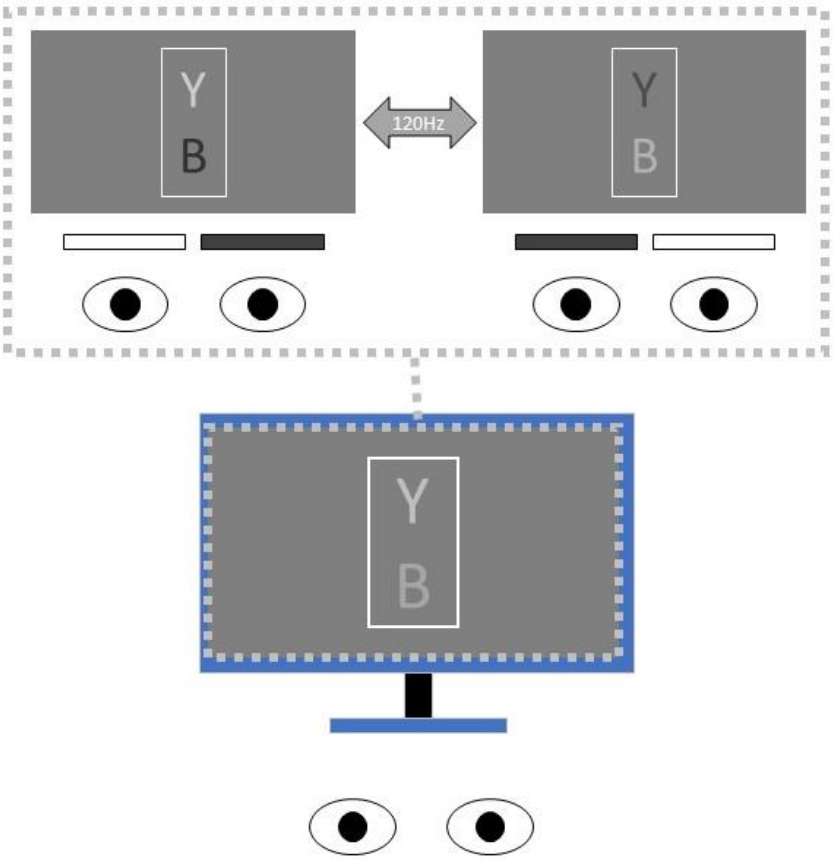


Figure 7.4.1. Presentation of letters to each eye, conflicting in luminance. The VIEWPixx interleaves the two images at 120Hz, and the images are gated to each eye using the 3D goggle system.

The crucial part of this visual test are the letters, which systematically change in their luminance, to create binocular rivalry on the letters. The participant's choice of letter as the brightest is informative of the participant's eye dominance. The degree of binocular contrast was varied on each trial, to determine the extent of eye dominance. Below, the principles of this visual test are explained in more detail.

Luminance map

As the first step, the actual luminance onscreen must be mapped to an RGB triplet (red-green-blue, e.g., [255 255 255], for pure white), such that in the experiment, the precise luminance values required can be produced on the letters using the RGB triplet specifier in the letter drawing commands. Out of the box, it is unknown what RGB input is needed to produce a given luminance output on the VIEWPixx.

To create the luminance map, a Matlab script was created, which takes an RGB input to present a single colour over the entire VIEWPixx display panel. A handheld photometer (Minolta LS-110) was positioned 500 millimetres away and centred to the VIEWPixx display panel, just as a participant would be in the experiment session, to measure the luminance on the colour presented on the VIEWPixx panel. Starting with a desired luminance output of 1 cd/m^2, it was a process of trial and error to find an RGB input value which produces the 1 cd/m^2 output. This trial and error input-output mapping was performed in 1 cd/m^2 increments, up to the maximum luminance of the VIEWPixx (95 cd/m^2). This input-output map was recorded in a Matlab matrix for use in the experimental script, and ranges from 1 cd/m^2 to 95 cd/m^2.

General design

As the basic idea, two letters conflicting in luminance (one dark, one light) are binocularly presented, such that the one eye sees the letter as dark, the other eye sees the letter as light. In this visual test, there are two sets of such binocularly conflicting letters, but in opposite light-dark directions between the eyes (compare top and bottom sets, Figure 7.4.1). Thus, there are four elements to the stimulus: left eye top letter, right eye top letter, left eye bottom letter, right eye bottom letter (Figure 7.4.1). The luminance on these four elements is determined by a mathematical rule, such that one element is brightest of the four (i.e., visible to only one eye). In binocular presentation, the four elements are perceived only as two letters, top and bottom. The choice between the top and bottom letters as brightest (2AFC) is indicative of the participant's binocular balance. If the participant 'correctly' determines the brightest letter, then they are either binocularly balanced, or dominant in the eye that happened to be presented with the brightest letter. If the participant chooses the other letter as brightest, then their dominance is to the other eye. Below are some worked examples to illustrate the principle in more detail.

Worked examples

The key to the contrast letters visual test is a contrast multiplier, ranging from 0 to 1, which controls the luminance on all four elements, in effect controlling the contrast between the letters. The contrast multiplier, C, works with the background luminance, B, to determine the luminance, LUM, for each of the four letters, in the following way:

$\text{LUM}_{(\text{left eye, top letter})} = \text{B} + \text{BC}$

$\text{LUM}_{(\text{right eye, top letter})} = B – B(1 – C)$

$\text{LUM}_{(\text{left eye, bottom letter})} = B – BC$

$\text{LUM}_{(\text{right eye, bottom letter})} = B + B(1 – C)$

To put some numbers to better illustrate the principle, the background luminance, B, is 50 cd/m^2, and in the experiment, the contrast multiplier starts at 0.7. Hence, the luminance values, LUM, for the four letters:

$\text{LUM}_{(\text{left eye, top letter})} = 50 + 50(0.7) = 85\ \text{cd/m}^2$

$\text{LUM}_{(\text{right eye, top letter})} = 50 – 50(0.3) = 35\ \text{cd/m}^2$

$\text{LUM}_{(\text{left eye, bottom letter})} = 50 – 50(0.7) = 15\ \text{cd/m}^2$

$\text{LUM}_{(\text{right eye, bottom letter})} = 50 + 50(0.3) = 65\ \text{cd/m}^2$

Figure 7.4.2 shows how the letters look when the contrast multiplier is at 0.7. Note that the two letters are binocularly conflicting by the same amount (50 cd/m^2), just in opposite directions. To the left eye, ‘Y’ is brighter than ‘B’, by 70 cd/m^2. To the right eye, ‘B’ is brighter than ‘Y’, by 30 cd/m^2. The brightest letter of the four is the upper left ‘Y’, at 85 cd/m^2. The experimental task is to choose which letter appears brightest. Which letter the participant chooses as brightest is indicative of their binocular balance in the moment. If the top letter, ‘Y’ is indicated as the brightest (using the yellow button on the ResponsePixx), then the participant was either binocularly balanced (both eyes’ views seen), or biased towards the left eye. If the bottom letter, ‘B’ is indicated as brightest (blue button on the ResponsePixx), then the participant was seeing more with their right eye, despite there being a brighter letter in the other eye’s view.

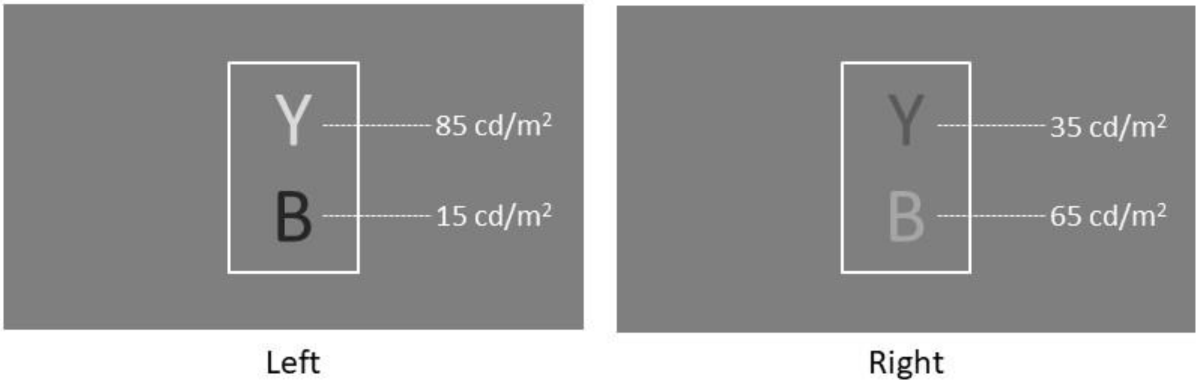


Figure 7.4.2. The contrast letters display, with contrast multiplier at 0.7, and background luminance at 50 cd/m^2. This display is good for detecting right-eye bias, because it is always the right-eye view with the lower contrast and dimmer bright element. To indicate the bottom letter, ‘B’, as brightest, one must be right-eye biased, because there is concurrently a brighter bright element in the left-eye view.

To make the test display more difficult, the contrast multiplier can be reduced. Below is a worked example, using 0.6 as the contrast multiplier, and illustrated in Figure 7.4.3.

$\text{LUM}_{\text{(left eye, top letter)}} = 50 + 50(0.6) = 80\ \text{cd/m}^2$

$\text{LUM}_{\text{(right eye, top letter)}} = 50 – 50(0.4) = 30\ \text{cd/m}^2$

$\text{LUM}_{\text{(left eye, bottom letter)}} = 50 – 50(0.6) = 20\ \text{cd/m}^2$

$\text{LUM}_{\text{(right eye, bottom letter)}} = 50 + 50(0.4) = 70\ \text{cd/m}^2$

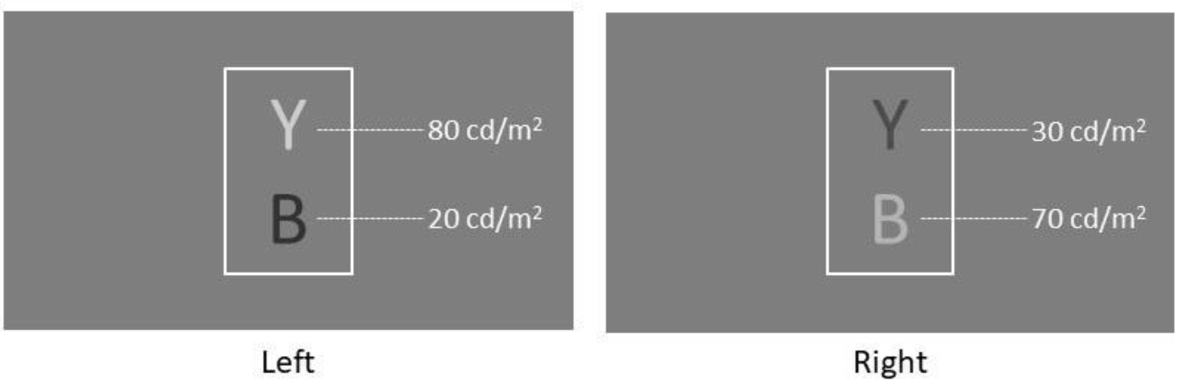


Figure 7.4.3. The contrast letters display, with a contrast multiplier of 0.6 against a background of 50 cd/m2. Compared to the contrast multiplier at 0.7, here, the left-eye contrast is lower, and the right-eye contrast is higher. The top letter, ‘Y’, in the left-eye view is still brightest, but only by 10 cd/m^2 against the bottom letter, ‘B’, in the right-eye view. The smaller contrast multiplier reduced binocular contrast, making the task of determining the brightest letter more difficult.

With the contrast multiplier at 0.6, the left and right eye versions of each letter again differs in luminance by 50 cd/m^2, in opposite directions. The reduced contrast multiplier tends to equalise the contrast in each eye. In the left eye, the top and bottom letters differ by 60 cd/m^2, compared to 70 cd/m^2 when the contrast multiplier was 0.7. In the right eye, the top and bottom letters differ by 40 cd/m^2, an increase from the 30 cd/m^2 when the contrast multiplier was 0.7. Moreover, the brightest element, ‘Y’, in the left-eye view, is only brightest by 10 cd/m^2 against the bright element ‘B’ in the right-eye view. If the participant selects the top letter, ‘Y’, as brightest, then they are either binocularly balanced, or left-eye biased. If the participant selects the bottom letter, ‘B’, as brightest, despite the brighter element in the left-eye view, then they must be right-eye biased. Altogether, the contrast multiplier at 0.6 is still good for detecting a right-eye dominance, but due to the equalisation of binocular contrast, this test display is more difficult than with the contrast multiplier at 0.7.

The contrast multiplier reaches its endpoint at 0.5. To illustrate, a worked example (see Figure 7.4.4):

$\mathrm{LUM}_{\text{(left eye, top letter)}} = 50 + 50(0.5) = 75$ cd/m^2

$\mathrm{LUM}_{\text{(right eye, top letter)}} = 50 – 50(0.5) = 25$ cd/m^2

$\text{LUM}_{(\text{left eye, bottom letter})} = 50 – 50(0.5) = 25\ \text{cd/m}^2$

$\text{LUM}_{(\text{right eye, bottom letter})} = 50 + 50(0.5) = 75\ \text{cd/m}^2$

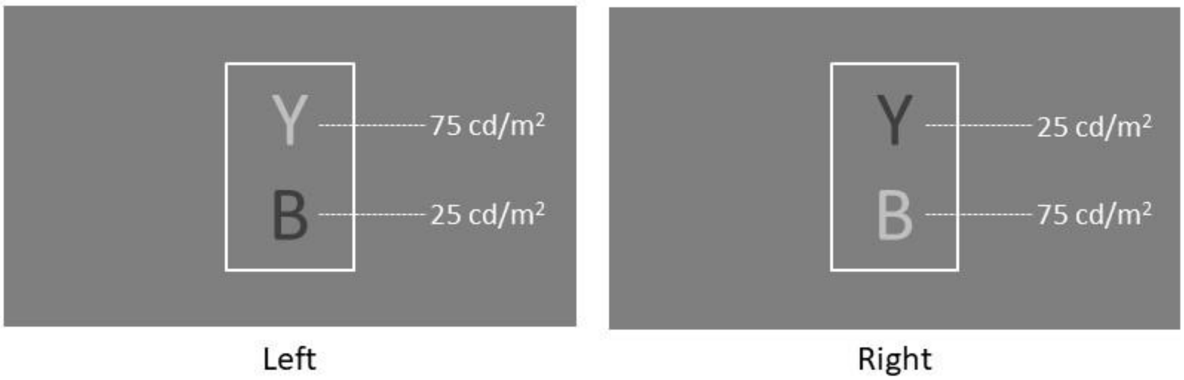


Figure 7.4.4. The contrast letters visual test reaches its endpoint when the contrast multiplier is at 0.5, because both letters are made of the same binocular luminance conflict (75 cd/m^2 versus 25 cd/m^2).

With the contrast multiplier at 0.5, the left- and right-eye views are exact opposites. Binocularly, each letter differs by 50 cd/m^2. In each view, the top and bottom letters differ by 50 cd/m^2. There is no absolute brightest among the four elements. With this binocularly-equal display, there should be three options in response to the question of which letter was brightest: 1) top button, meaning the participant sees more from the left-eye, 2) bottom button, meaning the participant sees more from the right-eye, or 3) equal button, meaning the participant sees both views equally. However, this test operates on a 2AFC paradigm.

In summary, when the contrast multiplier is greater than 0.5, the test is good for measuring right-eye dominance. The larger the contrast multiplier, the greater the binocular contrast. If the participant were to make a right-eye biased response at high contrast multipliers, then that would suggest a strong right-eye dominance. The contrast multiplier was varied over multiple trials to determine the contrast threshold for making a right-eye biased response, as a quantification for right-eye dominance.

Left-eye dominance

If it is left-eye dominance that needs quantification, then a contrast multiplier between 0 and 0.5 is required. Below is a worked example, using 0.3 as the contrast multiplier (illustrated in Figure 7.4.5).

$\text{LUM}_{(\text{left eye, top letter})} = 50 + 50(0.3) = 65\ \text{cd/m}^2$

$\text{LUM}_{(\text{right eye, top letter})} = 50 – 50(0.7) = 15\ \text{cd/m}^2$

$\text{LUM}_{(\text{left eye, bottom letter})} = 50 – 50(0.3) = 35\ \text{cd/m}^2$

$\text{LUM}_{(\text{right eye, bottom letter})} = 50 + 50(0.7) = 85\ \text{cd/m}^2$

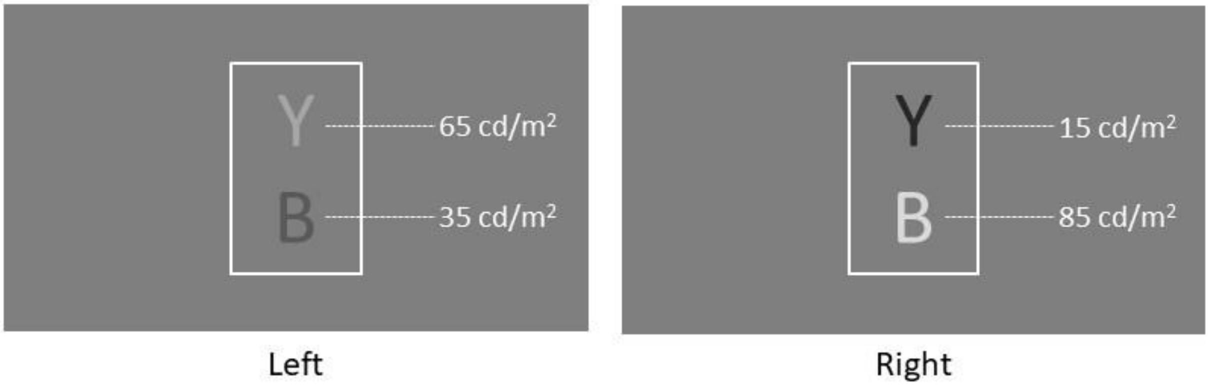


Figure 7.4.5. The contrast letters display, when the contrast multiplier is set at 0.3. The right-eye view features higher contrast than the left-eye view. Also, it is the bottom element in the right-eye view that is brightest of all. If the bottom letter is chosen, then the participant must either be binocularly balanced, or right-eye biased. If the top letter is chosen, then the participant must be left-eye biased. Hence, contrast multipliers between 0 and 0.5 are good for determining left-eye dominance.

When the contrast multiplier is between 0 and 0.5, it is the right-eye view that has the higher contrast between top and bottom letters (70 cd/m^2), compared to the left-eye view (30 cd/m^2). Also, the brightest element of all is the bottom letter, in the right-eye view (85 cd/m^2). Hence, if the bottom letter, ‘B’, was indicated as brightest, then the participant was either binocularly balanced, or right-eye biased. If the top letter, ‘Y’, was indicated as brightest, despite the presence of a brighter letter in the right-eye view, then the participant must be left-eye dominant. In short, when the contrast multiplier is between 0 and 0.5, the contrast letters displays are good for determining left-eye dominance. The contrasts here are reversed from when the contrast multiplier is between 0.5 and 1.

7.4.2. Procedures

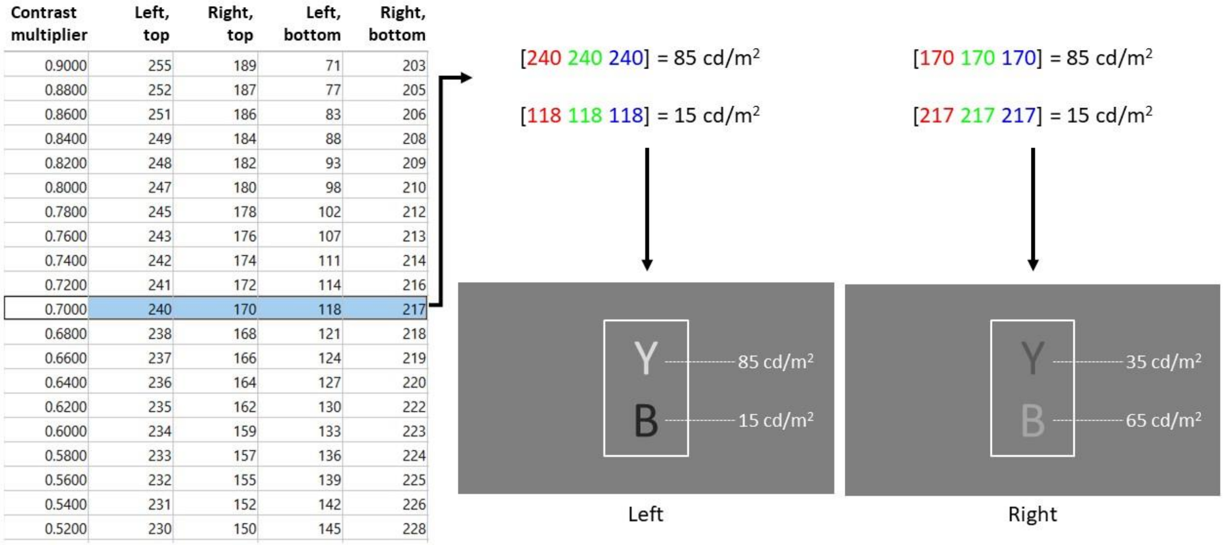
The visual tests were arranged as a battery (see Chapter 7.2). The contrast letters visual test comes immediately after the binocular rivalry visual test, so the participant should already be wearing the 3D goggles. Before the test began, an instruction screen was presented onscreen, as a reminder of the experimental task. The experimental task was to indicate which letter was brighter, yellow button for top letter, blue button for bottom letter, matching with the button arrangement on the ResponsePixx (see also Figure 7.1.4). When the participant was ready, they indicated this to the researcher, who then started the contrast letters visual test.

The contrast letters visual test used an adaptive staircase to determine the contrast multiplier for each trial. The adaptive staircase was implemented using the Palamedes toolbox running-fit function (Kingdom & Prins, 2016a; Prins & Kingdom, 2018). The running-fit function operated on four parameters that characterise the participant’s responding (Kingdom & Prins, 2016b): alpha (contrast multiplier thresholds, set at 0.52 to 0.90 in 0.02 increments), beta (contrast multiplier slope, set at 3.5), gamma (guess rate, set at 0.5, i.e., the participant guesses out of the two options), and lambda (lapse rate, set at 0.02, i.e., the rate at which the participant occasionally misses regardless of stimulus level e.g., momentary lapse of attention). The beta and gamma values were taken directly from the originator of this test paradigm, Bossi et al. (2018). The lambda value was also meant to be the same as that used by Bossi et al. (2018), but was accidentally set as 0.02 instead of the original’s 0.01. The running-fit functions keep track of performance on each trial, and uses prior performance to determine the contrast threshold for the upcoming trial (Kingdom & Prins, 2016a, 2016b).

On the first trial, 0.7 was chosen as the contrast multiplier. The experimental script searched in the luminance map (see Chapter 7.4.1), looking for the values to determine the luminance of the four letters in the test display (Figure 7.4.6). There were four values, one for each letter, and each value was simply expanded out to form the RGB triplet, which then went into the letter-drawing commands. Thus, the correct binocular contrast was shown for the selected contrast multiplier of 0.7.

The contrast multiplier for subsequent trials depended on the response in the previous trial. If the participant 'correctly' responds that the top letter is brighter, then the running-fit function recommends a smaller contrast multiplier for the next trial (more difficult to discern the brightest letter). If the participant 'incorrectly' responds that the bottom letter is brighter, then the running-fit function recommends a larger contrast multiplier on the next trial (easier to discern the brightest letter). If the participant does not respond in time (1.5 second response window), then this was also taken as an 'incorrect' response, and the running-fit function recommends a larger contrast multiplier on the next trial. The experimental script takes the recommended contrast multiplier and seeks the closest match on the luminance map, using the associated luminance values for the next trial (Figure 7.4.6). All trials presented the letters for one second, followed by a 1.5 second response window with the onscreen prompt 'Which letter was brighter?'. On making a response on the ResponsePixx, or timing out, the next trial was immediately presented.

The criteria for terminating the test was 12 reversals. A reversal is when the staircase changes direction (Kingdom & Prins, 2016a). For example, a series of trials were answered 'correctly', so the contrast multiplier was on a downward staircase (more difficult). Once an 'incorrect' response was made, the contrast multiplier changes to an upward staircase (less difficult). This change in staircase direction is a reversal, and the participant's contrast threshold is somewhere in the reversal, between the 'correctly' answered contrast multiplier, and the 'incorrectly' answered contrast multiplier (Kingdom & Prins, 2016a, 2016b), The first two reversals were not used, so the final threshold value was taken from the last 10 reversals. At the end of the contrast letters visual test, the experiment script moved to the next visual test in the battery: stereopsis visual test (Chapter 7.5).



Contrast multiplier	Left, top	Right, top	Left, bottom	Right, bottom
0.9000	255	189	71	203
0.8800	252	187	77	205
0.8600	251	186	83	206
0.8400	249	184	88	208
0.8200	248	182	93	209
0.8000	247	180	98	210
0.7800	245	178	102	212
0.7600	243	176	107	213
0.7400	242	174	111	214
0.7200	241	172	114	216
0.7000	240	170	118	217
0.6800	238	168	121	218
0.6600	237	166	124	219
0.6400	236	164	127	220
0.6200	235	162	130	222
0.6000	234	159	133	223
0.5800	233	157	136	224
0.5600	232	155	139	225
0.5400	231	152	142	226
0.5200	230	150	145	228

Figure 7.4.6. Once the contrast multiplier has been recommended by Palamedes' running-fit function (Kingdom & Prins, 2016a; Prins & Kingdom, 2018), the nearest match is sought for in the luminance map, and the four associated luminance values are each expanded out into a RGB triplet. The four RGB triplets are used in the letter drawing commands, so that each of the four letters on the test display are of the correct luminance.

Special procedure: trick trials

In addition to the regular trials determined by the running-fit adaptive staircase (Kingdom & Prins, 2016a; Prins & Kingdom, 2018), trick trials were also occasionally inserted within. The issue is that the staircase is limited to a range of 0.52 to 0.90, which means the top letter in the left-eye view is always brightest, and the 'correct' response is always the yellow button on the ResponsePixx. The staircase also goes from an intermediate contrast multiplier down to a difficult contrast multiplier. Therefore, to a participant with fairly balanced visual functioning, they might realise that only the yellow button needs to be pressed. On the regular staircase, if one keeps pressing the yellow button, then one will quickly reach the most difficult contrast multiplier, regardless of their true visual functioning.

As a countermeasure, a trick trial was inserted after every five regular trials which were below the contrast multiplier of 0.6. Trick trials feature either a 0.8 contrast multiplier (same direction as the regular trials, 'correct' response on the yellow button), or a 0.2 contrast multiplier (opposite direction to the regular trials, ‘correct’ response on the blue button). The two trick contrast multipliers were used in alternation.

The trick trials served two functions. First, the occasional insertion of a trial strongly in the opposite direction is to catch any mindless pressing of the response buttons. On the regular trials, pressing the yellow button to indicate top letter is ‘correct’. However, on the opposite direction trick trial, pressing the blue button to indicate bottom letter is correct. If the participant was mindlessly pressing the yellow button, then the trick trials will catch and record this error.

The second function of the trick trials was to motivate the participant. Once the fine contrast multipliers had been reached (in this case, below 0.6), then occasional trials of an ‘easy’ 0.8 contrast multiplier can help the participant stay motivated in the visual test.

7.4.3 Data analysis

There were two steps to analysing the data from the contrast letters visual test. First, responses to the trick trials with the opposite-direction contrast multiplier (0.2) were checked. Participants who had answered incorrectly to these opposite-direction trick trials were rejected, as they would appear to be pressing away on the yellow button and not truly responding to the stimuli.

Second, the staircase data on each test session was analysed for the contrast multiplier threshold, using the analysis scripts in the Palamedes toolbox (Prins & Kingdom, 2018). Each participant was tested five times (Baseline, Post0, Post10, Post20, Post). Hence, for each participant, there were five threshold values of the contrast multiplier, which were charted to examine the progression of contrast multiplier thresholds (i.e., eye dominance) over the course of the experiment. Where applicable, one-sample ANOVA and paired-samples t-tests were performed, with the alpha level at 0.05. Statistical analyses were performed using SPSS (IBM).

7.5 Visual test – Stereopsis

The final visual test examined the participant's perception of visual depth, stereopsis. For this pilot study, stereopsis was of interest, for two reasons. First, if monocular patching produces a dominance effect on the patched eye (e.g., Lunghi et al., 2011), is this an imbalance between the two eyes, that could be detrimental to stereopsis, which requires the use of both eyes? Second, alternately patching both eyes makes up one of the knowledge gaps (see Chapter 10), and the key question for that experiment was whether alternately patching both eyes produces binocular benefits, thus improving stereopsis. Hence, the stereopsis performance before and after the patching procedure needs to be measured. Following are the details of the stereopsis visual test, covering the stimuli, procedures and data analysis.

7.5.1 Stimuli

In basic terms, the stimuli of the stereopsis visual test were presented binocularly, i.e., one image to each eye. Each image contained two large rectangular patches (220 x 400 pixels, width x height) populated by 400 randomly positioned small white squares (each measuring 8 x 8 pixels). One patch was on the left half of the image, the other patch was on the right half of the image. These two images were interleaved at 120Hz in presentation on the VIEWPixx, and the two images were gated to the correct eye using the 3D goggles system (Figure 7.5.1).

To generate binocular disparity, thus apparent depth, the left- and right-eye views were horizontally offset to each other. The manner in which both eyes' views are offset to each other determines whether the patch appears closer or further away than the display plane. This apparent depth varied trial by trial. Binocular disparity was only applied to the left-side patch of squares (see slight shift in the left-side patch of squares, between left- and right-eye views; Figure 7.5.1). The right-side patch did not receive the binocular disparity treatment, so it always appeared to be at the depth of the display plane. With two patches at once, left-side patch with binocular disparity, right-side patch without binocular disparity, the participant was faced with a choice of two patches (2AFC), in responding to the experimental task of determining which patch appeared closer.

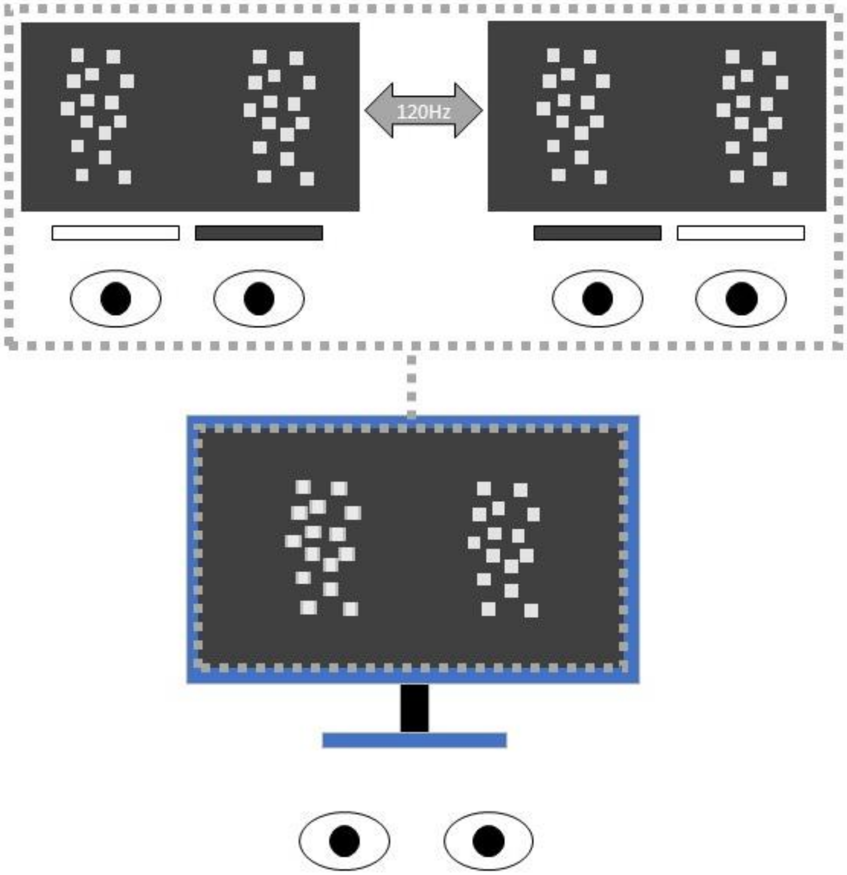


Figure 7.5.1. In the stereopsis visual test, two patches of small squares were shown binocularly, one patch without binocular disparity, one patch with a variable degree of binocular disparity to simulate different apparent depths. The task was to determine which patch of squares appeared closer to oneself. The performance towards various levels of binocular disparity determines the participant's stereopsis threshold. Figure not drawn to scale.

Creating binocular disparity

If the object had no horizontal offsets applied between both eyes' views (see Figure 7.5.2a), i.e., no binocular disparity, then the left and right eyes converge onto the display plane. The object therefore appears to be on the display plane, which in the experiment is 500 millimetres away (Figure 7.5.2b). The right-side patch never had horizontal offsets applied (see Figures 7.5.2, 7.5.3, 7.5.4).

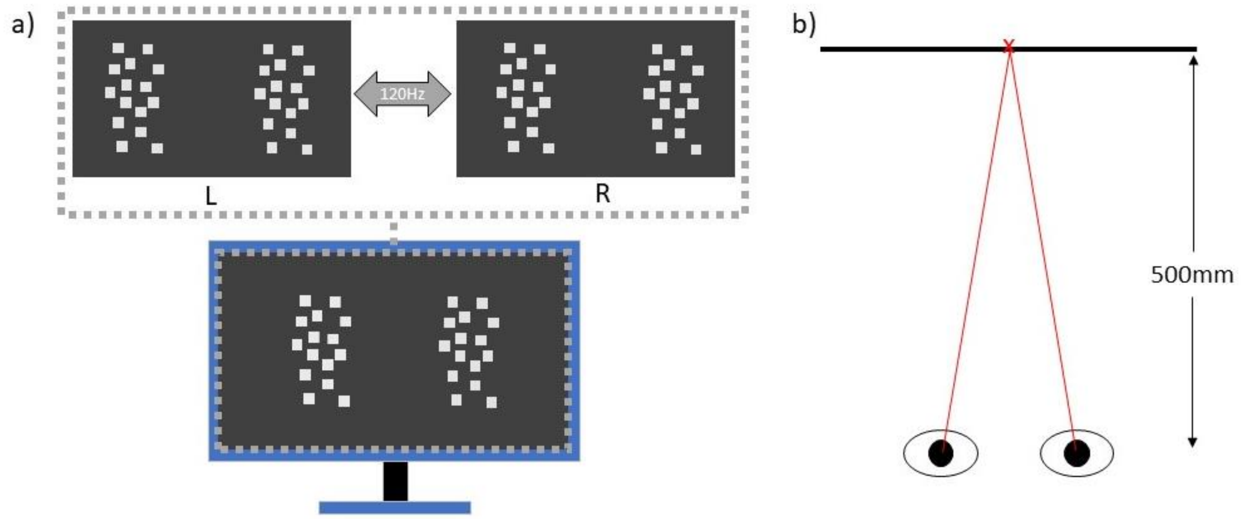


Figure 7.5.2. The stimuli with no binocular disparity applied (via binocular horizontal offsets). a) The system of VIEWPixx and 3D goggles present the left- and right-eye views to each eye. Neither patch were horizontally offset to each other; both were drawn in the same positions in both eyes' views. b) When the left- and right-eye views are in the same position, in effect, the eyes converge onto the display plane, and the object is perceived at the physical 500mm depth.

The visual stimuli can be made to appear either closer or further away than the physical distance (i.e., 500 millimetres to the display plane), depending on the manner of binocular disparity, applied via the binocular horizontal offsets. The binocular horizontal offset can be made by shifting only one eye's view, or shifting both eyes' views. In this visual test, only the right eye's view was shifted. Shifting one eye's view means the offset can be made in single pixel increments, useful given that the combination of display resolution and viewing distance meant each pixel shift produced about 0.03° of binocular disparity.

To create an illusion that the visual object was further away than the display plane, the visual object in the right-eye view was shifted rightwards, whilst the position of the same visual object in the left-eye view was unchanged (Figure 7.5.3a, shifted patch of squares highlighted in green). In effect, this manner of binocular horizontal offset meant that the eyes converged at an illusory point behind the display plane: the visual object appears to be further away (Figure 7.5.3b).

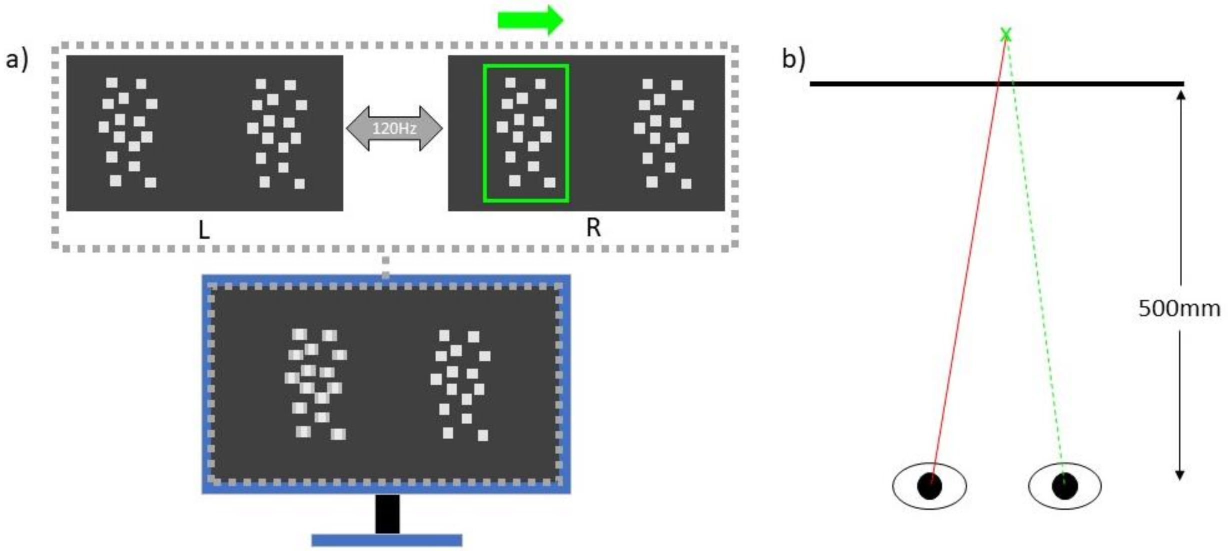


Figure 7.5.3. Generating apparent depth behind the display plane. a) The system of VIEWPixx and 3D goggles presents left- and right-eye images to each eye. Binocular horizontal offset is applied to the left-side patch. In the left-eye view, the left-side patch does not shift. In the right-eye view, the left-side patch is shifted rightwards (patch highlighted with green box, shift shown with arrow). b) With the right-eye view shifted rightwards, in effect, both eyes converge at an illusory point behind the display plane, making the stimuli appear to be further away than the display plane.

To create an illusion that the visual object is in front of the display plane, the visual object in the right-eye view was shifted leftwards, whilst the position of the same object in the left-eye view was unchanged (Figure 7.5.4a, shifted patch highlighted in blue, arrow shows shift direction). In effect, this manner of binocular horizontal offset meant the eyes converge at an illusory point in front of the display plane: the object appears to be closer (Figure 7.5.4b).

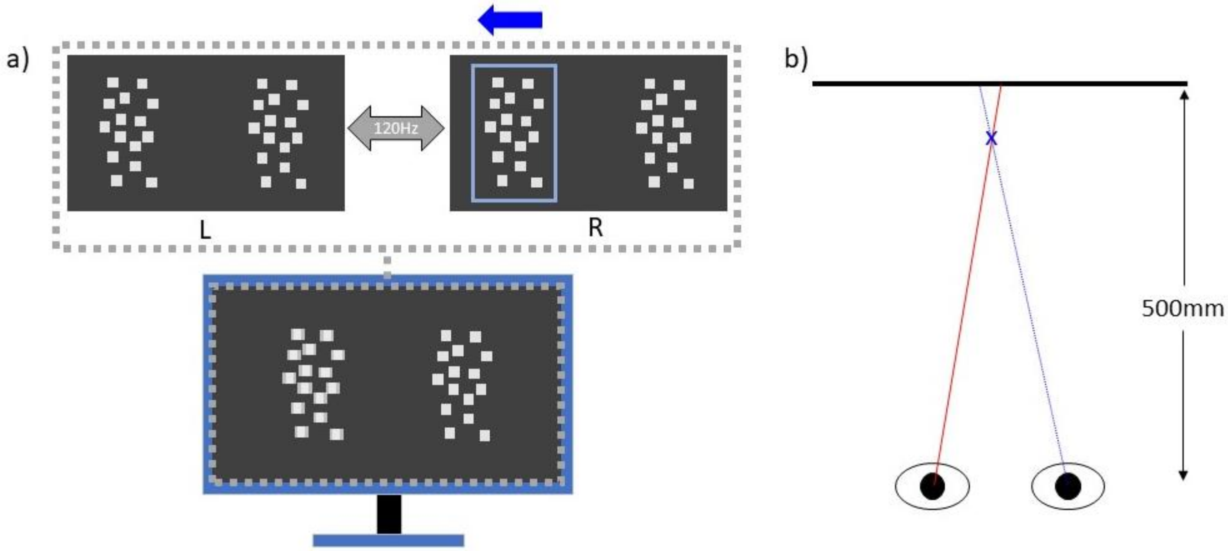


Figure 7.5.4. Generating apparent depth in front of the display plane. a) The system of VIEWPixx and 3D goggles presents left- and right-eye views of the stimuli to each eye. Binocular horizontal offset is applied to the left-side patch. In the left-eye view, the left-side patch is unchanged. In the right-eye view, the left-side patch is shifted leftwards. b) In effect, this manner of binocular horizontal offset means both eyes converge at an illusory point in front of the display plane; the object appears closer.

Depth map

Using geometry, a depth map was created to map each pixel increment in horizontal offset between left- and right-eye views, to binocular disparity (as a visual angle). The basic measurements were 500 millimetres from display to participant, and 63 millimetres between each eye. The display resolution was 1920 x 1080, on a 24 inch panel, so the pixel pitch was 0.2715 millimetres. Altogether, it was determined that each pixel of shift approximated to an increase in binocular disparity by 0.03° in either direction (apparent further/nearer). Hence, five levels of binocular disparity were used, for both apparent further (Table 7.5.1) and apparent nearer (Table 7.5.2), with a maximum of ±0.15° of binocular disparity. Higher levels of binocular disparity were not used, because they risk diplopia, whereby the left- and right-eye views are so far apart that one no longer perceives depth, but rather perceives a double image, and can cause discomfort for the observer. The depth map was encoded in a Matlab matrix, for use in the experimental scripts.

Binocular disparity (°)	Rightwards shift (pixel)
+0.15	5
+0.12	4
+0.09	3
+0.06	2
+0.03	1

Table 7.5.1. Depth map for apparent depth behind the display plane. There are five levels of binocular disparity, shown here with the associated rightwards shift in the right-eye view.

Binocular disparity (°)	Leftwards shift (pixel)
-0.15	5
-0.12	4
-0.09	3
-0.06	2
-0.03	1

Table 7.5.2. Depth map for apparent depth in front of the display plane. There are five levels of binocular disparity, shown here with the associated leftwards shift in the right-eye view.

7.5.2 Procedures

The visual tests were arranged as a battery (see Chapter 7.2). The stereopsis visual test comes immediately after the contrast letters visual test, so the participant should already be wearing the 3D goggles. Before the test began, instructions were presented onscreen, as a reminder of the experimental task. The experimental task was to indicate which patch of squares appeared closer, green button for the left-side patch, red button for the right-side patch, matching with the button arrangement on the ResponsePixx (see also Figure 7.1.4). When the participant indicated that they were ready, the researcher started the test.

The stereopsis visual test used an adaptive up/down staircase to determine the binocular disparity to use for the next trial. The up/down staircase was implemented with the Palamedes toolbox, using the up/down functions (Kingdom & Prins, 2016a; Prins & Kingdom, 2018). At first, the adaptive staircase works on a 1-down rule, meaning one correct response is enough to decrease the binocular disparity on the next trial (i.e., more difficult). This initial 1-down rule is such that the participant's approximate threshold level can be quickly reached.

Once an incorrect response has been made, a 1-up/3-down rule was used. So, if an incorrect response was made, the next trial moves one step up the staircase, using a higher binocular disparity (easier). To move one step down the staircase, three consecutive correct responses at the same binocular disparity are needed. If no response was made, then it was treated as an incorrect response, and the next trial steps up the binocular disparity. This adaptive up/down staircase blends speed with accuracy: the initial 1-down rule quickly fixes the subsequent trials on near-threshold values, while the later 1-up/3-down rule has a greater threshold accuracy of 79% compared to 50% for the 1-up/1-down rule (Kingdom & Prins, 2016a).

In use, the test was split into two parts. Part 1 tested for sensitivity towards visual depth behind the display plane, using the depth map in Table 7.5.1. Part 2 tested for sensitivity towards visual depth in front of the display plane, using the depth map in Table 7.5.2. For both Parts, the first trial started with the maximum binocular disparity, 0.15° (+/- depending on the Part). The binocular disparity for the next trial depended on current performance, according to the adaptive up/down staircase described above (Kingdom & Prins, 2016a; Prins & Kingdom, 2018). The binocular disparity recommended by the adaptive up/down staircase was taken for the next trial, which using the relevant depth map, the corresponding amount of pixel shift was applied to the stimulus-drawing commands (the depth map was stored as a Matlab matrix, in effect Tables 7.5.1 and 7.5.2). The stimulus patches had freshly randomised square positions on each trial.

Each trial consisted of one second of stimulus presentation, followed by a 1.5 second response window with the onscreen prompt 'Which side was nearer?'. The experimental script advanced to the next trial immediately after a valid button press (red or green, within the response window), or after the response window timed out. The test terminated after 12 reversals on the staircase. A reversal is a change in staircase direction, e.g., from downwards to upwards after an incorrect response (Kingdom & Prins, 2016a; see also Chapter 7.4 on the staircase procedure). The participant's stereopsis threshold is somewhere in the reversal, between the correctly answered binocular disparity, and the incorrectly answered binocular disparity (Kingdom & Prins, 2016a, 2016b). The first two reversals were not taken; only the latter 10 were used to determine the participant's stereopsis threshold.

Special procedures: trick trials

In addition to the regular trials determined by the adaptive up/down staircase, trick trials were also periodically inserted within. A key issue with this stereopsis test is that each part tests for one direction of apparent visual depth, e.g., Part 1 for apparent depth behind the display plane. So, to a participant with fairly good stereopsis, they may realise that in Part 1, only the red button needs to be pressed (to indicate the right-side patch as appearing nearer), or that in Part 2, only the green button needs to be pressed (to indicate the left-side patch as appearing nearer). Such responding would result in apparently very low stereopsis thresholds, regardless of the participant's true stereopsis functioning.

As a countermeasure, a trick trial was inserted after every four trials with a binocular disparity below 0.10°. The trick trial could either be +0.15° (5 pixels rightward shift on the right-eye image, object appears further), or -0.15° in binocular disparity (5 pixels leftward shift on the right-eye image, object appears closer). The two types of trick trials were used in alternation.

These trick trials served two functions. First, by inserting trials with binocular disparity opposite to the regular trials, it was possible to catch any mindless pressing on the buttons. For example, in Part 1, the red button was 'correct'. Yet, in the trick trials with binocular disparity in the opposite direction, the green button was correct. Hence, a participant mindlessly pressing on the red button would answer incorrectly on this type of trick trials. The errors on this type of trick trial was recorded.

The second function of trick trials was to motivate the participant. Once the staircase had reached fine binocular disparities (in this case, below 0.10°), then occasional trials with an 'easy' 0.15° binocular disparity (maximal binocular disparity in this test) could help the participant stay motivated during the visual test.

7.5.3 Data analysis

Data analysis consisted of two parts. First, the responses on trick trials were examined, and participants who had incorrectly responded to the trick trials were excluded from further analysis.

Next, the staircase on each test session was analysed for the stereopsis threshold, using the Palamedes toolbox functions (Prins & Kingdom, 2018). Each participant was tested five times

(Baseline, Post0, Post10, Post20, Post30). Hence, for each participant, there were five threshold values, which were then charted to examine the progression of stereopsis over the course of the experiment. Where applicable, one-way ANOVAs and paired-samples t-tests were used, with an alpha level of 0.05. Statistical analyses were performed using SPSS (IBM).

Chapter 8: Answering knowledge gap 4: Are there other positive effects of short-term monocular patching?

For the first step of this pilot study, the interest was on discovering more effects of short-term monocular patching. Three visual tests were employed (see Chapter 7), to test for visual functioning before and after 30 minutes of patching the right eye.

8.1 Methods for Experiment 1

8.1.1 Participants

Four participants (two female, all with normal or corrected-to-normal eyesight, all right-handed, mean age: 29 years ± 5 years) were recruited for this experiment, in accordance with the participant recruitment procedures explained in Chapter 7.2.1.

8.1.2 Patching

Eye patching was performed using a translucent eye patch (Bernell BTEP+), resulting in 20/400 vision or worse, on the participant's right eye, for 30 minutes. During this time, the participants watched television. The number 20/400 is an indication of visual acuity: a person with such vision can see an object only 20 feet away, when the same object can typically be seen from 400 feet afar (normal vision is 20/20; American Optometric Association, n.d.; Cleveland Clinic, 2022).

8.2 Results from the binocular rivalry visual test

The binocular rivalry visual test presented conflicting imagery between the eyes. The image perceived depends on eye dominance. On a group level, three metrics were analysed from the test data: dominance ratio, dominance proportion, and fusion. See Chapter 7.3 for details about the test.

8.2.1 Dominance ratio

The dominance ratio is the total time indicating the right-eye image, divided by the total time indicating the left-eye image. Thus, a ratio of one means equal time between right- and left-eye views, whereas a ratio larger than one means more time spent viewing with the right eye (i.e., dominance). The ratio was constructed in this way to put right-eye dominance on a positive scale;

participants were patched on the right eye, so it should be the right eye that becomes more dominant (patched eye more dominant; e.g., Lunghi et al., 2011).

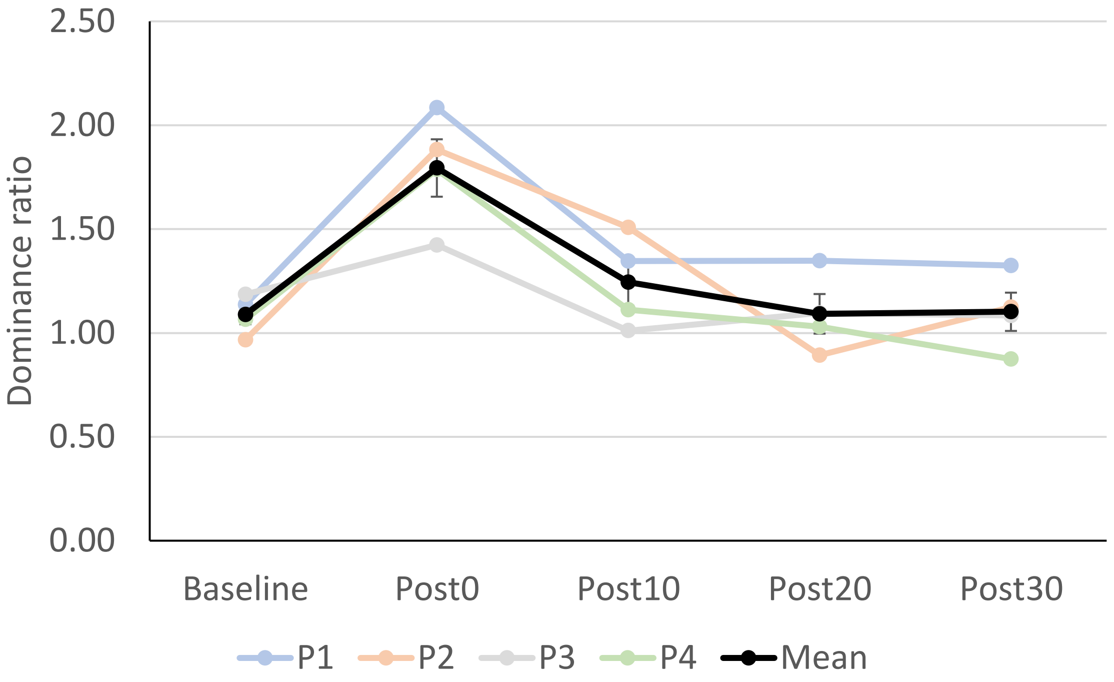


Figure 8.2.1. Dominance ratio over the course of the experiment, on an individual level (coloured line for each individual, see legend), and as a group-mean (black line, SEM error bars).

Looking at the dominance ratios (Figure 8.2.1), all four participants started by being almost perfectly balanced binocularly (dominance ratio close to 1, baseline), and became more right-eye biased (increase in dominance ratio) immediately after patching (Post0). However, there was a general decline in dominance ratio, hence right-eye dominance, thereafter, with variability between participants. By 10 minutes after patching, Participant 1's (blue; Figure 8.2.1) right-eye dominance reduced from the peak at Post0, appearing to hold steady at a dominance ratio only slightly more right-eye biased than at baseline. Participant 2's (orange) right-eye dominance reduced from the peak at Post0, down to around neutral binocular balance by 20 to 30 minutes after patching (Post20, Post30). Participants 3 and 4 (grey, green respectively) seem to have neutralised their Post0 peak right-eye dominance already by 10 minutes after patching (Post10). Participant 4 seems also to have transitioned to a slight left-eye dominance at Post30.

To check if the above observations reached statistical significance, first, examining the dominance ratio as a function of time, a one-way ANOVA was performed, which revealed a significant effect of time on the dominance ratio ($F_{(4,12)} = 12.424$, $p<0.001$, $\eta_p^2 = 0.805$). Then, to examine if the significant changes in binocular ratio referred to a significant post-patching shift towards right-eye dominance, multiple paired-sample t-tests were performed. The dominance ratio at each 'Post-' timepoint was compared to the dominance ratio at baseline. The dominance ratio immediately after patching (Post0) was significantly larger than baseline ($t_{(3)} = 4.312$, $p = 0.023$, mean difference: 0.705 ± 0.163), meaning there was a significant shift to right-eye dominance after patching the right eye. However, 10 minutes after patching (Post10), the right-eye bias was no longer significant ($p = 0.380$). The non-significance of the right-eye bias continued into 20 minutes after patching (Post20; $p = 0.969$) and 30 minutes after patching (Post30; $p = 0.901$).

Altogether, it appears that the binocular balance shifted to the right eye after patching the right eye, concurring with previous research on monocular patching (e.g., Lunghi et al., 2011). While the size of the dominance effect is similar to that found in Lunghi et al. (2011), the effect found here was not durable, seemingly lasting less than 10 minutes.

8.2.2 Dominance proportion

Eye dominance was also determined from the Dieter et al. (2017) formula based on the proportion of time indicating each eye's view (see Chapter 7.3.5). Neutral binocular balance is indicated with 0%, while a positive percentage indicates a right-eye bias, a negative percentage indicates a left-eye bias.

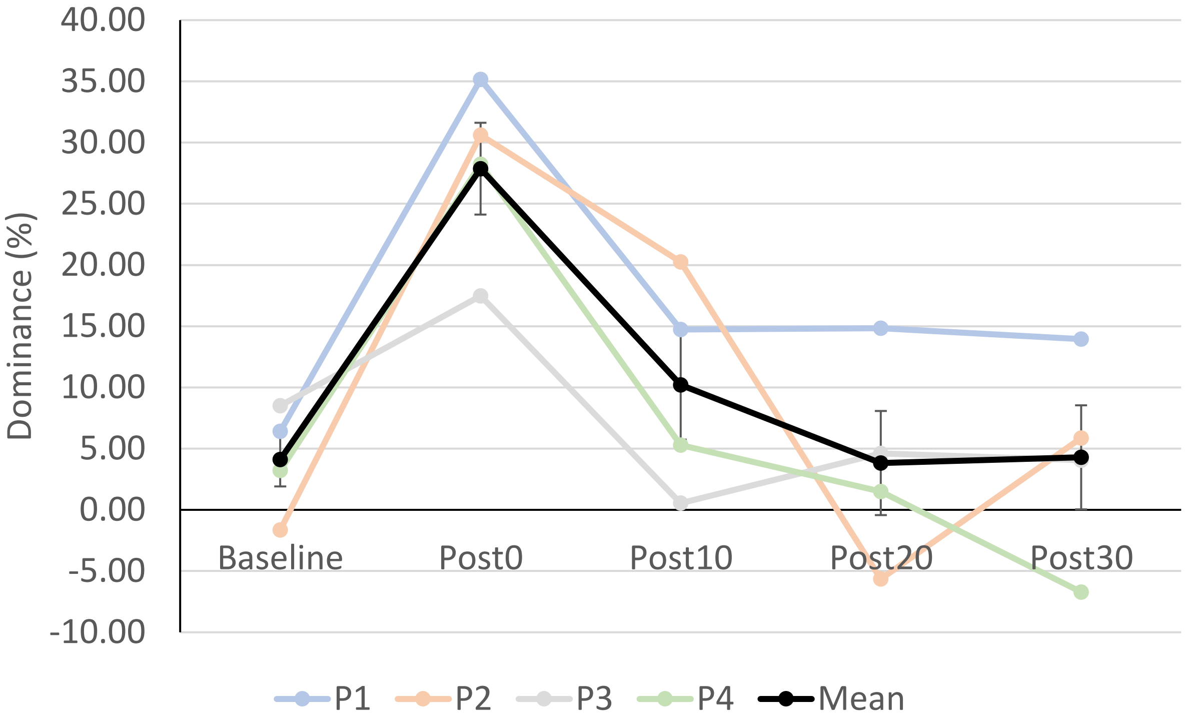


Figure 8.2.2. Eye dominance proportion over the course of the experiment. The individual eye dominance proportion over the course of the experiment is charted on separate lines, using different colours. The participant-averaged progression of eye dominance is shown in black (SEM error bars).

Visually, the dominance proportion shows a similar pattern as that found using the dominance ratio (compare Figure 8.2.2 with Figure 8.2.1). All participants were close to neutral binocular balance at the baseline (near 0%). Immediately after patching the right eye (Post0), the dominance proportion percentage increased, meaning the right eye became dominant. However, the percentage decreased by 10 minutes after patching: the right-eye dominance declined. There appears to be participant variability in the dominance decline from Post0 onwards, perhaps more so than that found using the dominance ratio metric (compare Figure 8.2.2 with Figure 8.2.1, note greater heterogeneity in participant trends with the dominance proportion than with dominance ratio). From Post0, Participant 1's (blue line, Figure 8.2.2) right-eye dominance appears to have reduced and held steady at a lower level, albeit seemingly still more right-eye dominant than at baseline. Participant 2's (orange line) right-eye dominance appears to have reduced to around neutral by Post20. Participants 3 and 4 (grey and green lines respectively) both show a reduction in right-eye dominance to around neutral by 10 minutes after patching (Post10). Participant 4 seems also to have transitioned into a slight left-eye dominance by 30 minutes after patching (Post30).

To check if the progression of dominance proportion was statistically significant, a one-way ANOVA was performed, revealing that time was a significant factor on the dominance proportion ($F_{(4,12)} = 9.404$, $p=0.001$, $\eta_p^2 = 0.758$). To check if this time effect meant a change in dominance proportion compared to baseline, the dominance proportion at each 'Post' timepoint was compared to baseline, using paired-samples t-tests. At Post0, the dominance proportion was significantly larger than the baseline dominance ($t_{(3)} = 4.619$, $p = 0.019$, mean difference: 23.748% ±5.142%), meaning a significant shift to right-eye dominance. However, 10 minutes after patching, the dominance proportion was no longer significantly larger than the baseline dominance proportion ($p = 0.402$), and this non-significance continued into 20 minutes after patching ($p = 0.927$) and 30 minutes after patching ($p = 0.972$).

Altogether, this dominance proportion analysis showed that right eye patching was followed by increased right eye dominance, concurring with past research (e.g., Lunghi et al., 2011). However, this patching-induced dominance seemed not to last more than 10 minutes post-patching. Similar findings were obtained using the dominance ratio (see Chapter 8.2.1).

8.2.3 Fusion

During binocular rivalry, there may be moments where both eyes' views are simultaneously perceived as a fused percept. The amount of fusion could be an indication of binocular suppression, with more fusion associated with weaker binocular suppression (Dieter et al., 2017). The moments of fusion were summed, for each timepoint, each participant, and analysed.

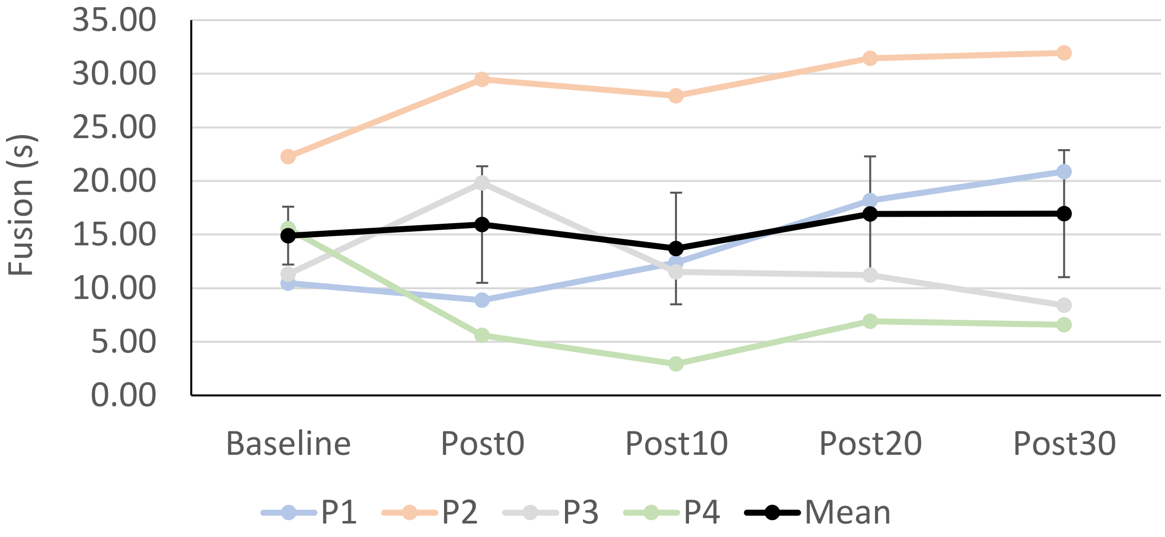


Figure 8.2.3. Duration of fused percept, at each timepoint. Individual participant data on separate coloured lines. Participant-averaged data in black, SEM error bars.

Visually (see Figure 8.2.3), there appears to be individual variability in the duration of fusion perceived by the four participants. At baseline, Participants 1 and 3 (blue and grey lines respectively, Figure 8.2.3) had around 10 seconds of fusion during 60 second binocular rivalry. Participant 4 (green) had 15.558 seconds. Participant 2 (orange) had 22.275 seconds.

The progression of the fused percept post-patching also showed individual variability: Participants 1 and 2 (blue and orange respectively) appear to have steadily perceived more fusion after patching, Participant 3 (grey) had a singular increase at Post0 followed by a downward trend in fusion, while Participant 4 (green) seemed to have steadily perceived less fusion after patching. The progression of the fused percept post-patching is interesting. First, more fused percept could indicate reduced binocular suppression (Dieter et al., 2017); the increase in fusion immediately post-patching (hence a decrease in binocular suppression) exhibited by Participants 2 and 3 (orange and grey) seems to run counter to the significant increase in right-eye dominance immediately post-patching (see Chapters 8.2.1, 8.2.2 for an analysis on dominance ratio and dominance proportions, respectively). Second, the trend in the fused percept at later Post- timepoints could suggest some possible recovery styles from short-term monocular patching. Participants 1 and 2 (blue and orange) show steady upward trends in fusion, so one style of recovery from monocular patching could be an increase in fusion. Participant 4 (green) showed a downward trend in fusion (i.e., increased suppression; see Dieter et al., 2017), and also seemed to have become slightly left-eye dominant at Post30 (see Chapters 8.2.1, 8.2.2), hence another style of recovery from monocular patching could be in switching dominance to the non-patched eye.

On a group level (black line, Figure 8.2.3), a one-way ANOVA was performed to examine if the duration of fused percept changed significantly over the course of the experiment: no effect of time was found ($F_{(4,12)} = 0.311$, $p = 0.865$, $\eta_p^2 = 0.094$). As a further step, to check if there were changes to fusion after patching, at each 'Post' timepoint, the duration of fused percept was compared to baseline, using paired-samples t-tests. At no point was the fused percept duration significantly different to baseline (all $p = 1.000$). The group-level statistics were null, but this could be partly attributable to the aforementioned individual variability in perceiving the fused percept.

8.3 Results from the contrast letters visual test

The contrast letters visual test presents letters conflicting in contrast between each eye. The participant's choice of letter as brightest indicates their eye dominance. This visual test works by varying the binocular contrast on an adaptive staircase to determine a contrast threshold. More details about this visual test in Chapter 7.4.

8.3.1 Absolute contrast threshold

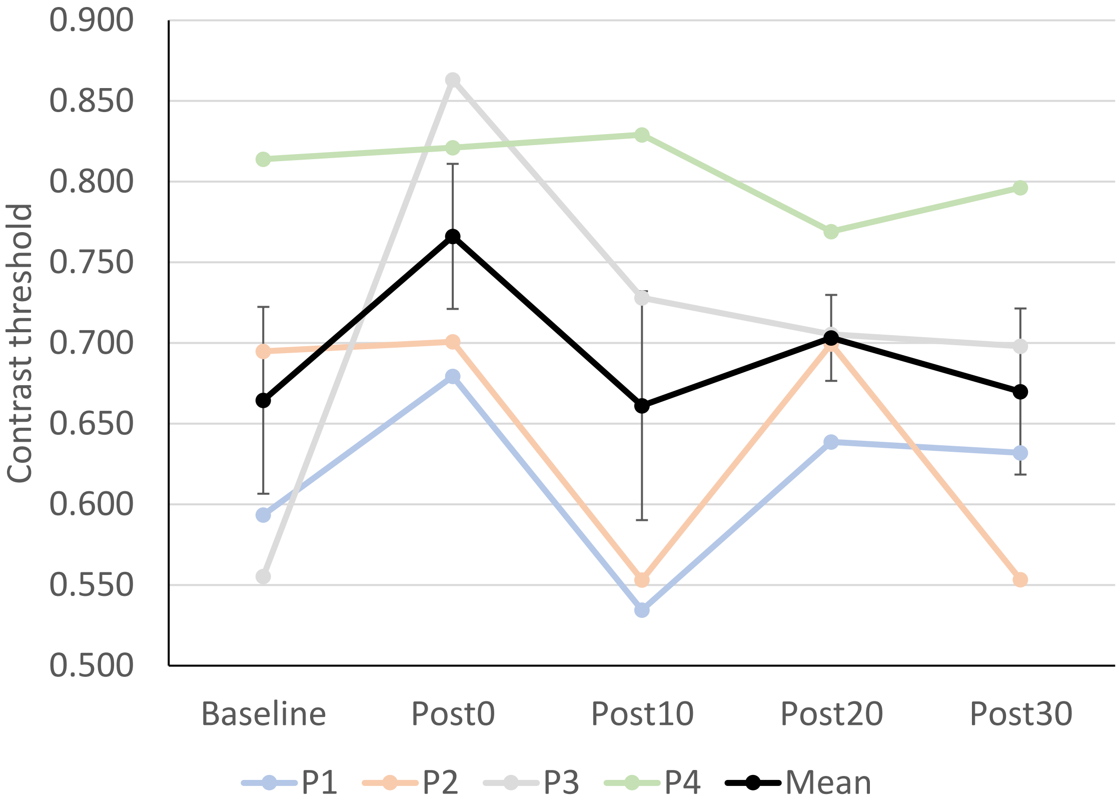


Figure 8.3.1. Absolute contrast threshold as a function of time. Individual participant data on separate coloured lines. Group-averaged data in black, SEM error bars.

A contrast threshold larger than 0.5 indicates a right-eye bias. The larger the contrast threshold above 0.5, the more right-eye biased the participant was. Visually (Figure 8.3.1), it appears that only Participants 1 and 3 (blue and grey) produced an increase in contrast threshold after patching the right eye. Hence, it appears that only these two participants became more right-eye biased after patching.

On a group level, to examine if the contrast threshold changed as a function of time, a one-way ANOVA was performed: there was no effect of time on the contrast threshold ($F_{(4,12)} = 1.663$, $p = 0.223$, $\eta_p^2 = 0.357$). As a further check, the contrast threshold at each 'Post' timepoint was compared to baseline, to examine if patching had brought about significant changes to contrast threshold. At no point was the contrast threshold different to baseline: Post0 ($p = 0.250$), Post10 ($p = 0.962$), Post20 ($p = 0.422$), Post30 ($p = 0.935$).

8.3.2 Contrast threshold change

It was noted that absolute contrast thresholds were highly variable between participants. Notably, the baseline values were quite disparate, with two participants below 0.6 in contrast threshold (weak right-eye bias), while one participant was over 0.8 in contrast threshold (strong right-eye bias). Hence, an analysis on absolute contrast threshold may not yield any significant results.

To reduce the heterogeneity in the data, the baseline was equalised to zero for all participants. For each participant, the 'Post' contrast thresholds were subtracted with the participant's baseline, in effect converting the data into contrast threshold *changes* (see Figure 8.3.2). The group-average was calculated by averaging through the participants' contrast threshold changes. Hence, if patching produced a right-eye bias, then the contrast threshold change was positive. If patching produced a left-eye bias, then the contrast threshold change was negative.

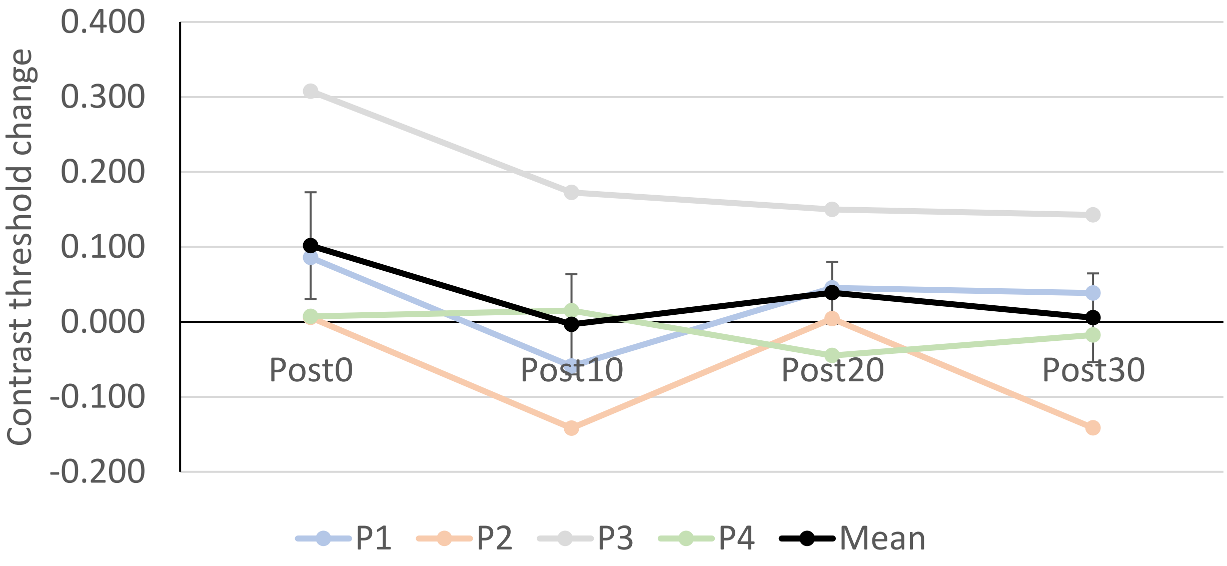


Figure 8.3.2. Contrast threshold change, compared to baseline, for each timepoint post-patching. Individual data on separate coloured lines, group-averaged data in black (SEM error bars).

A more consistent picture emerged with the contrast threshold changes (Figure 8.3.2), compared to the absolute contrast thresholds (Figure 8.3.1). Visually, there was an increase in contrast threshold after patching for Participants 1 and 3 (Participants 2 and 4 remain unchanged), so these two participant became more right-eye dominant. 10 minutes and later into the post-patching period, there seems to be a downward trend in contrast threshold, suggesting that the participants either became less right-eye dominant, or became left-eye dominant.

A one-way ANOVA on the contrast threshold changes was performed, to check for the effect of time. No effect of time was found ($F_{(3,9)} = 3.164$, $p = 0.078$, $\eta_p^2 = 0.513$). To check if the contrast threshold changes were significantly different to 0 (baseline), a one-sample t-test against 0 was performed on the contrast threshold changes at each timepoint. At no point was the contrast threshold change significantly different to 0: Post0 ($p = 0.249$), Post10 ($p = 0.964$), Post20 ($p = 0.422$), Post30 ($p = 0.935$).

8.4 Results from the stereopsis visual test

The stereopsis visual test determines the participant's stereopsis threshold by varying the binocular disparity (hence the percept of apparent depth) on a staircase. There were two parts to this test (more details in Chapter 7.5): Part 1 (apparent depth behind the display plane) and Part 2 (apparent depth in front of the display plane). There were only two valid datasets (Participants 1 and 2) for this test. The other two participants responded correctly to every trial; the test terminated due to time constraints, with no stereopsis threshold measurable for these participants in this test's design.

<u>8.4.1 Part 1: apparent depth behind the display plane</u>

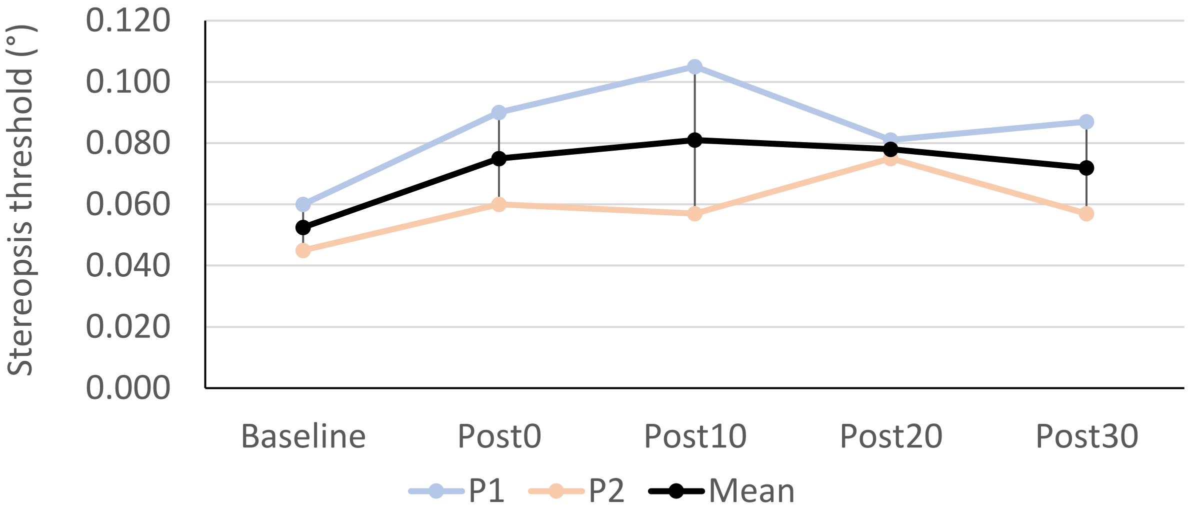


Figure 8.4.1. Stereopsis threshold as a function of time (apparent depth behind the display plane). Individual data on separate coloured lines. Group-averaged data in black, SEM error bars.

Looking at the two participants' stereopsis thresholds over the course of the experiment (Figure 8.4.1), there appears to be a general trend of stereopsis threshold increase, i.e., worsening, after 30 minutes of right-eye patching. For Participant 1 (blue), the stereopsis threshold increased immediately after patching, and continued to increase until 10 minutes after patching. Thereafter, Participant 1's stereopsis threshold tended to decrease, albeit not to baseline levels even at the end of the experiment. For Participant 2 (orange), there appears to be a slight increase in stereopsis threshold immediately after patching, with a sudden and unique peak at 20 minutes after patching. Participant 2's stereopsis threshold reduces thereafter, but again, not to baseline levels by the end of the experiment. Averaging the two participants' progression in stereopsis thresholds (black), there was a steady increase (worsening) in stereopsis threshold after patching, with the peak at 10 minutes after patching, followed by only a slight lowering (improvement) of stereopsis threshold thereafter, without returning to baseline levels by the end of the experiment.

Despite the small sample size, a one-way ANOVA was performed to examine if the observed changes in stereopsis thresholds over the course of the experiment were significant: no effect of time on stereopsis threshold was found (p = 0.267). As a further check, the stereopsis threshold at each 'Post' timepoint was compared to baseline, using paired-samples t-tests. At no point was the stereopsis threshold significantly different to baseline: Post0 (p = 0.205), Post10 (p = 0.334), Post20 (p = 0.111), Post30 (p = 0.234). The group-level statistics did not support the above observations of worsened stereopsis after patching, but the sample size of two is perhaps not the most conducive to statistical testing.

8.4.2 Part 2: apparent depth in front of the display plane

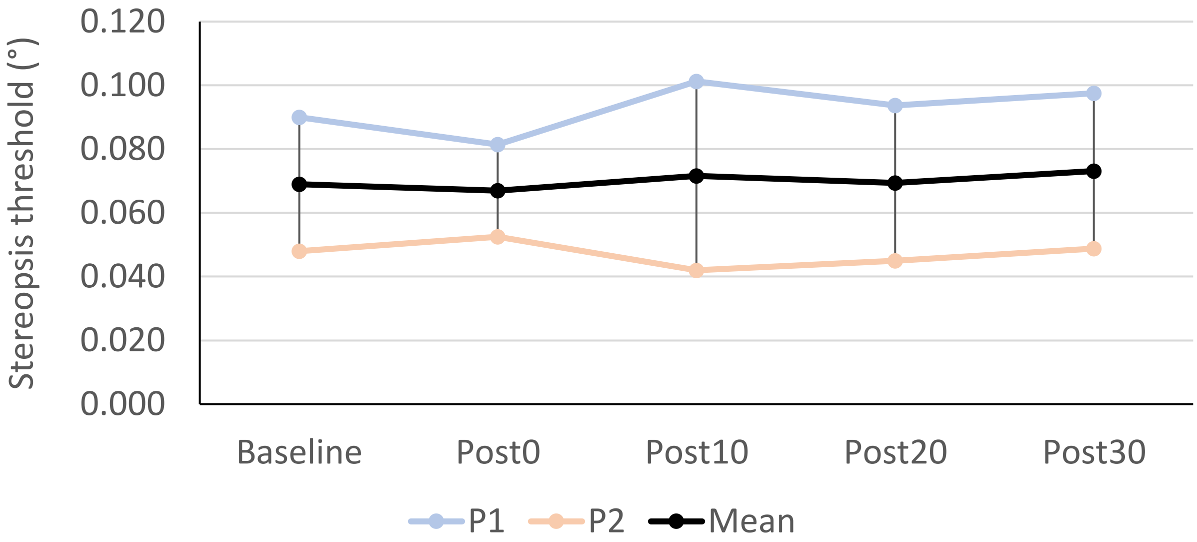


Figure 8.4.2. Stereopsis thresholds as a function of time (apparent depth in front of the display plane). Individual data on separate coloured lines. Group-averaged data in black, SEM error bars.

On visual inspection (Figure 8.4.2), for both participants (blue and orange), stereopsis thresholds appeared to hover around baseline levels throughout the course of the experiment; no consistent trends were detected. The average of both participants' stereopsis thresholds (black) yielded a consistent stereopsis threshold of around 0.070° throughout the course of the experiment. To check if the stereopsis thresholds differed as a function of time, a one-way ANOVA was performed; there was no effect of time on the stereopsis thresholds (p = 0.933). To follow up, the stereopsis threshold at each 'Post' timepoint was compared against the baseline, using paired-samples t-tests. At no point was the 'Post' stereopsis threshold different to the baseline: Post0 (p = 0.823), Post10 (p = 0.818), Post20 (p = 0.910), Post30 (p = 0.421). Altogether, the visual inspection and statistical analyses (albeit with only two datasets) did not reveal changes to stereopsis during the course of the experiment; 30 minutes of right-eye patching did not impact the sensitivity to visual depth in front of the display plane.

8.5 Chapter Discussion

This chapter attempted to explore the effects of short-term monocular patching, the first knowledge gap of this pilot study. Unlike previous studies, which patched for two or more hours and focused on measuring eye dominance (e.g., Lunghi et al., 2011; Lunghi et al., 2013; Lunghi & Sale, 2015), this first experiment patched for 30 minutes, and examined if there were patching effects in addition to changes in eye dominance. Three visual tests were employed, measuring eye dominance, suppression and stereopsis. The results were mixed, though with the caveat of a limited sample size (n=4) due to covid-related limitations, there were still statistically significant effects, and as a pilot study, the results are arguably interesting to examine nonetheless.

First, the binocular rivalry visual test, which measures eye dominance using two metrics (dominance ratio, dominance proportion) found a significant shift in eye dominance towards the patched eye after patching, thus concurring with past research (e.g., Lunghi et al., 2011). After patching the right eye for 30 minutes, all participants perceived the right-eye image more than the competing left-eye image, and both metrics picked up on this right-eye dominance. This right-eye dominance after patching was statistically significant in both metrics, in terms of a time effect on the eye dominance (ANOVA), and in comparing the post-patching dominance metric to the baseline dominance (paired-samples t-test). The dominance effect found here was also of similar magnitude to the dominance effect found by the originators of this patching paradigm (e.g., Lunghi et al., 2011). However, unlike earlier studies (e.g., Lunghi et al., 2011) which found durable dominance effects post-patching, here, the patching-induced dominance effect did not seem durable, lasting no more than 10 minutes after patching.

Some individual variability in the binocular rivalry visual test was also noted. When again tested on the binocular rivalry visual test, by 10 minutes after patching and later, participants diverged in how their eye balance recovered: reduction in right-eye dominance but still above baseline, reduction down to neutral balance, or reduction in right-eye dominance with an eventual transition to slight left-eye dominance. Although both dominance metrics revealed significant dominance effects, the dominance proportion metric seems more sensitive to individual variability than the dominance ratio metric.

Individual variability in the progression of eye dominance was also seen in the fusion metric. The fusion metric looks into the total time perceiving a fusion of left- and right-eye images; more time perceiving fusion is thought to indicate weaker binocular suppression (Dieter et al., 2017). Two participants appeared to perceive more fusion immediately after patching (i.e., weaker binocular suppression), which seems at odds with their increase in right-eye dominance according to the dominance metrics. Nonetheless, the fusion metric is most interesting at the later 'Post-' timepoints, as they seem to offer an insight into the recovery from the monocular patching experience. Specifically, two participants eventually perceived more fusion at the later timepoints, while one participant perceived less fusion and this coincided with a transition to left-eye dominance. Hence, it appears that the imbalance caused by monocular patching could eventually be counteracted either by allowing more contribution from both eyes (perhaps a reduction in suppression/dominance), or by switching dominance to the non-patched eye. Although speculative and from a small sample size, it was interesting to see how eye balance might recover after monocular patching. To my knowledge, this is the first look into eye balance recovery after monocular patching.

Intriguingly, the contrast letters visual test, which was recommended as a robust test of eye dominance (Bossi et al., 2018), did not seem to concur with the positive results found using the binocular rivalry visual test (i.e., patched eye becomes more dominant). Instead, the results from the contrast letters test seemed quite noisy. Even at baseline, the contrast letters visual test found participants to range from near-neutral binocular balance to strongly right-eye biased; the binocular rivalry visual test found all four participants to be tightly grouped around neutral at baseline. No dominance effects were also found using the contrast letters visual test, despite the revelation of significant dominance effects using the binocular rivalry visual test. If not a problem of the test itself, the apparent failure of the contrast letters visual test may be partly attributable to it being second in the battery (after the 1-minute binocular rivalry visual test, so some of the dominance effects may have faded by the time of the contrast letters visual test).

Outside of binocular dominance and suppression, stereopsis was also tested, because the dominance effect from monocular patching might be detrimental to the usage of both eyes together.

The limited sample size for this test (n=2), the resulting reliance on visually judging the data, and the stereopsis test being last in the test battery (a delay from the timepoint reaching five minutes), makes it challenging to produce conclusive statements on the effect of monocular patching on stereopsis. From the results that were obtained, there does not seem to be complete support for the notion of weakened stereopsis after monocular patching. First, two participants had stereopsis at a finer level than could be tested, at all timepoints, suggesting that monocular patching may not significantly impact stereopsis for some people. Second, for the other two participants, there seems to be some worsening of sensitivity to apparent depth behind the display plane after patching, but the sensitivity to apparent depth in front of the display plane appear unaffected. The examination of stereopsis after monocular patching is novel, and it is interesting to see that monocular patching does not necessarily upset binocular functions like stereopsis, or for some people, there is only a slight and temporary worsening of stereopsis after patching.

On the whole, this first experiment presents a preliminary look at how short-term monocular patching affects visual functioning. A drawback to the experiment, and this pilot study in general, is the small samples sizes, which was the result of covid-related restrictions on experimentation at the time. The small sample sizes are unlikely to offer enough power to discover all effects. Nonetheless, with only four participants, a statistically significant dominance effect on the patched eye was found, concurring with past research (e.g., Lunghi et al., 2011); the dominance effect seems robust. A more contentious issue with this first experiment is perhaps the patching duration. This experiment patched the right eye for 30 minutes. Earlier studies patched for two hours or more (e.g., Lunghi et al., 2011; Lunghi et al., 2013; Lunghi & Sale, 2015). Hence, it could be that robust dominance effects only arise from long patching durations. However, herein lies the problem: if this novel 'short-term' monocular patching takes two hours or more, then as a treatment, it is in fact no quicker to perform than the conventional patching treatment for amblyopia (typically 1-2 hours daily; Holmes & Clarke, 2006). It was therefore important to test an appreciably shorter patching duration, firstly because the effects of shorter patching durations have not been extensively studied, and also because one possible justification for this novel patching technique is speed and convenience. There was a significant dominance effect here already with 30 minutes of patching. In the next chapter, the issue of an even shorter patching duration was tested.

Chapter 9: Answering knowledge gap 5: How are the effects of short-term monocular patching when monocular patching is very short-term?

As explained in the introduction (Chapter 1.3.3) and the previous chapter, there is an interest to discover the effects of patching short periods of time. Conventional amblyopic patching can be lengthy, uncomfortable and inconvenient (Gambacorta et al., 2018; Holmes & Clarke, 2006; Maconachie & Gottlob, 2015; Webber & Wood, 2005). If this novel patching technique can be very short-term, then perhaps there is some scope for its usage. Furthermore, there is some evidence to suggest that the dominance effect on the deprived eye could be elicited with much shorter periods of monocular deprivation (i.e., 15 minutes, even 3 minutes; Kim et al., 2017). In this second experiment, the effects of patching for 10 minutes were tested.

9.1 Methods for Experiment 2

9.1.1 Participants

One participant was recruited (male, normal eyesight, right-handed, 25 years old) in accordance with the participant recruitment procedures outlined in Chapter 7.2.1. This participant also participated in Experiment 1 (see Chapter 8, P1).

9.1.2 Patching

Eye patching was performed with a translucent eye patch, the same as that used in Experiment 1 (Bernell BTEP+, resulting in 20/400 or worse vision, see Chapter 8.1.2 for definition), on the participant's right eye, for 10 minutes. During this time, the participant watched television.

9.2 Results from the binocular rivalry visual test

9.2.1 Dominance ratio

In binocular rivalry, the dominance ratio is the total time indicating the right-eye view, divided by the total time indicating the left-eye view. Hence, ratios larger than one mean more time perceiving the right-eye view, thus the right eye is dominant. See Chapter 7.3 for more details.

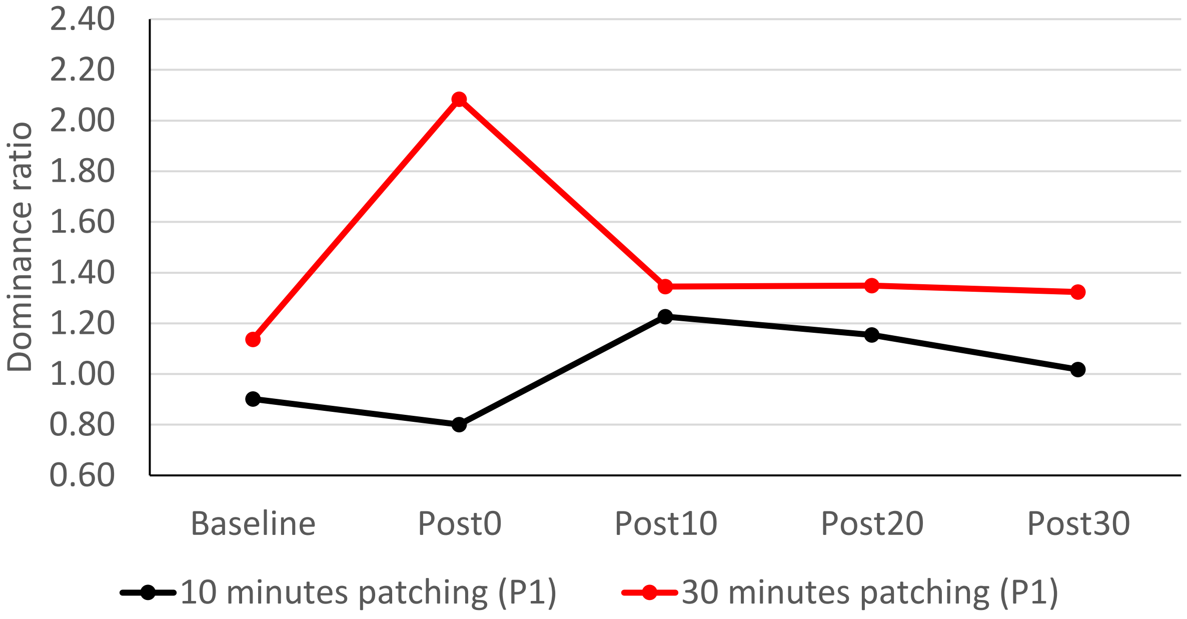


Figure 9.2.1. Comparison of dominance ratios over the course of the experiment, after 10 minutes of patching (black, this experiment), and after 30 minutes of patching (red, previous experiment). Both datasets originate from the same participant.

Visually examining the results (Figure 9.2.1, black), the participant started with a dominance ratio slightly below 1, meaning the participant was slightly left-eye dominant at baseline. After 10 minutes of patching to the right eye, the dominance ratio decreased further, suggesting an increase in left-eye dominance. Only by 10 minutes after patching (Post10) was there a shift to the right-eye dominance, as shown by a dominance ratio greater than 1. However, this right-eye dominance gradually faded thereafter, reaching neutral binocular balance by 30 minutes after patching (Post30).

In comparison, with 30 minutes of right-eye patching (Figure 9.2.1, red), the same participant immediately had a large shift to right-eye dominance (Post0, to a dominance ratio of over 2), which was followed by a reduction in right-eye dominance, but was still more right-eye dominant than baseline for the remainder of the experiment session. Thus, it appears that 10 minutes of right-eye patching induced a smaller and less durable shift to right-eye dominance than with 30 minutes of patching.

9.2.2 Dominance proportion

The dominance proportion metric compares the time proportions indicated right versus left (Dieter et al., 2017). Here, a positive percentage means right-eye dominance, whereas a negative percentage means left-eye dominance. See Chapter 7.3 for details. In Experiment 1, the dominance proportion metric was found to largely concur with the dominance ratio metric.

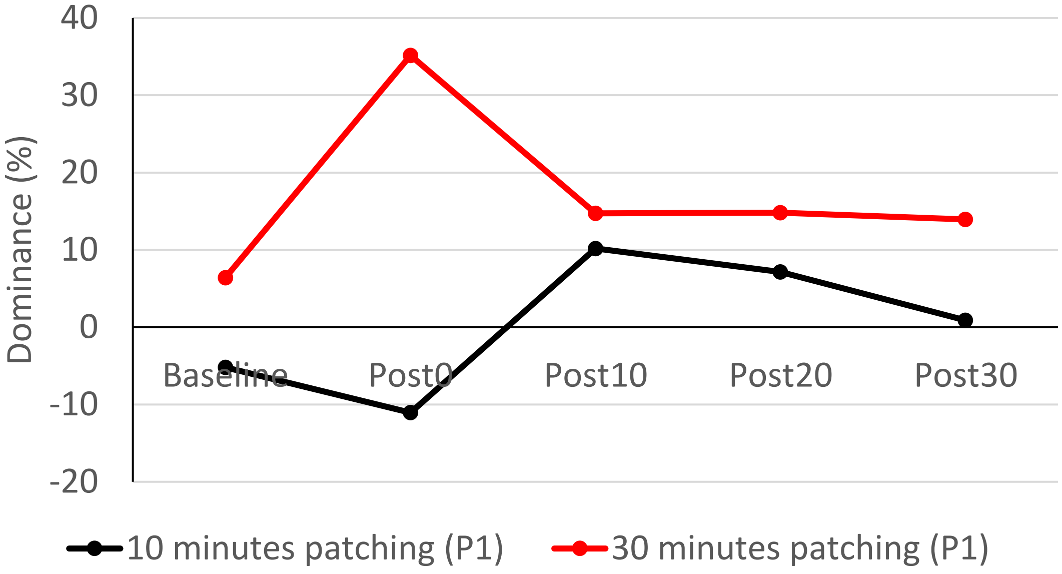


Figure 9.2.2. A comparison of eye dominance over the course of the experiment, between 10 minutes of patching (black) and 30 minutes of patching (red). This data originates from the same participant.

Visually inspecting the results (Figure 9.2.2, black), the participant had a small negative dominance percentage at baseline, suggesting a slight left-eye dominance. Immediately after patching, the left-eye dominance increased, as seen in the percentage becoming more negative. Only by 10 minutes after patching was there a shift to right-eye dominance, shown by the dominance percentage becoming positive. Thereafter, the right-eye dominance reduced, reaching neutral binocular balance by 30 minutes after patching. In comparison, with 30 minutes of right-eye patching (Figure 9.2.2, red), the same participant had an immediate and large shift to right-eye dominance after patching, which faded, but still remained more right-eye dominant than at baseline. Hence, both the dominance ratio and dominance proportion metrics agree; 10 minutes of patching seems to produce a smaller and less durable dominance effect than with 30 minutes of patching.

9.2.3 Fusion

During binocular rivalry, there can be moments when both eyes' views are simultaneously perceived, as a fused percept. Fusion has been linked to weak binocular suppression (Dieter et al., 2017). This experiment also examined the duration of fusion, as a purported suppression metric, to see if there is a relation to the dominance metrics. In Experiment 1 (Chapter 8), the fusion metric also seemed to reveal recovery styles from monocular patching.

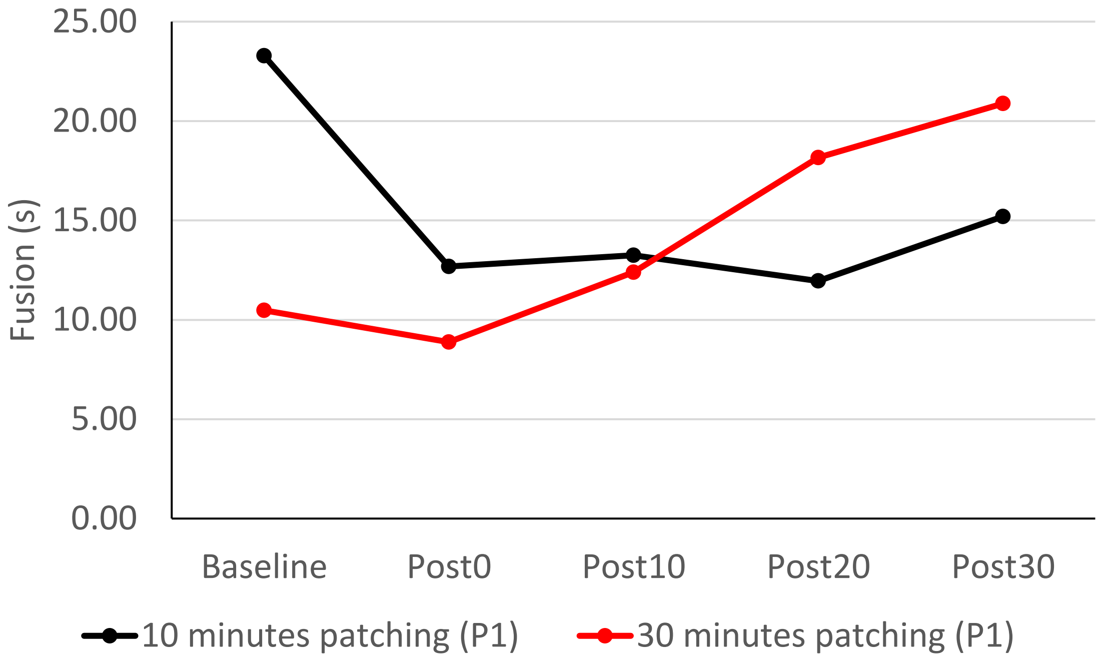


Figure 9.2.3. The amount of time (seconds) perceiving fusion, over the course of the experiment, for 10 minutes of patching (black) and 30 minutes of patching (red). Both datasets originate from the same participant.

Visually, it appears that 10 minutes of patching reduced the duration of fusion (Figure 9.2.3, black). This reduction in fusion persisted to 20 minutes after patching. By 30 minutes after patching, there was a small increase in fusion, though about eight seconds short of the baseline. The reduction in fusion could be linked to an increase in binocular suppression (Dieter et al., 2017), seemingly matching with the finding of increased right-eye dominance after patching (see dominance metrics, Chapters 9.2.1 and 9.2.2). The increase in fusion between 20 to 30 minutes after patching, i.e., reduced binocular suppression, seems to relate to the decrease in right-eye dominance, arriving at neutral binocular balance, as found in the dominance metrics (Chapters

9.2.1 and 9.2.2). Hence, for this participant, there appears to be a loose inverse relationship between the dominance and fusion metrics, with increased dominance related to decreased fusion (increased suppression), and vice versa. For this participant, the recovery from monocular patching also appears to be through an increase in fusion.

Curiously, the same participant after 30 minutes of right-eye patching (Figure 9.2.3, red) only showed a steady and large upward trend in fusion. There was only a slight reduction in fusion after 30 minutes of patching, despite a strong shift to right-eye dominance after patching (more so than with 10 minutes of patching; see Chapters 8.2, 9.2.1, 9.2.2). Hence, it appears that the loose inverse relationship between dominance and fusion is not apparent with 30 minutes of patching. This participant was consistent in recovery style from right-eye patching: in both patching durations, the recovery was through increasing fusion.

9.3 Results from the contrast letters visual test

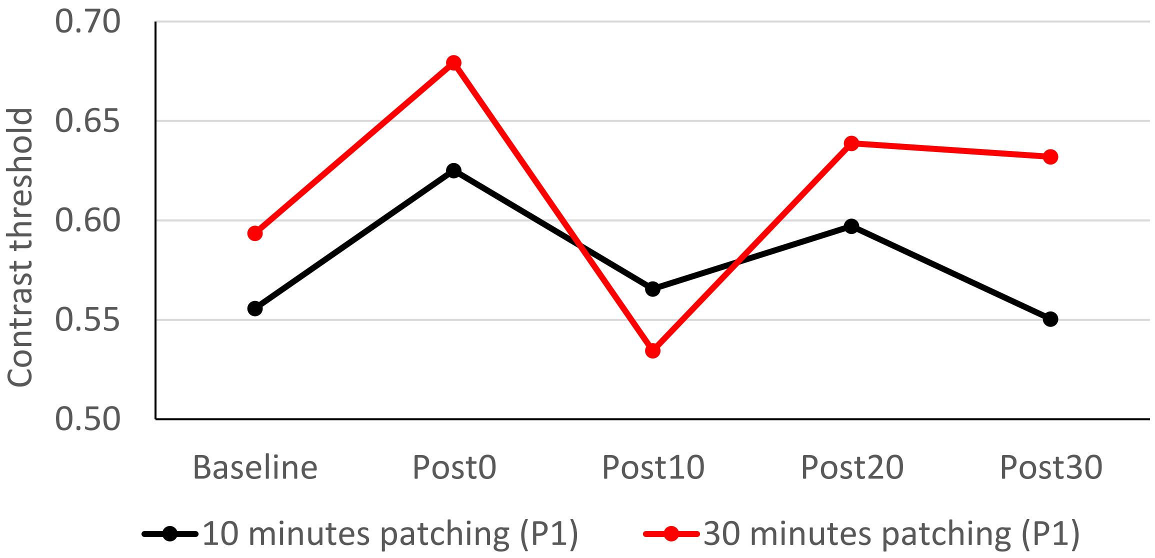


Figure 9.3. Contrast thresholds over the course of the experiment, for 10 minutes of patching (black), and 30 minutes of patching (red). The two datasets originate from the same participant.

The contrast letters visual test uses a binocular array of letters differing in luminance to obtain a contrast threshold, which is indicative of binocular balance (Bossi et al., 2018; see Chapter 7.4 for details). On visual inspection (Figure 9.3, black), the participant had a contrast threshold slightly above 0.5 at baseline, meaning a slight right-eye dominance. Immediately after 10 minutes of patching the right eye, the contrast threshold increased, meaning an increase in right-eye dominance. Thereafter, the contrast threshold was on an overall downward trend (including a dip in contrast threshold at Post10, followed by a rise in contrast threshold at Post20), reaching baseline levels by 30 minutes after patching.

In comparison, with 30 minutes of right-eye patching (Figure 9.3, red), the contrast threshold also increased after patching, and to a higher level of contrast threshold (i.e., greater right-eye dominance) than with 10 minutes of patching. However, for the experiment with 30 minutes of right-eye patching, the baseline was more right-eye biased. In reality, looking at the change in contrast threshold, the increase was similar for both 10 and 30 minutes of right-eye patching. 30 minutes of patching also exhibits large variability in the progression of contrast threshold, notably a large dip in contrast threshold at Post10, which is not as pronounced with 10 minutes of patching. Both patching durations seem to suggest an overall trend of returning to baseline levels by 30 minutes after patching.

9.4 Results from the stereopsis visual test

Stereopsis was tested in two parts, Part 1 for apparent depth behind the display plane, Part 2 for apparent depth in front of the display plane. Stereopsis functioning was of interest in case it would be affected by the dominance effect from monocular patching. Details of the stereopsis visual test are in Chapter 7.5.

9.4.1 Part 1

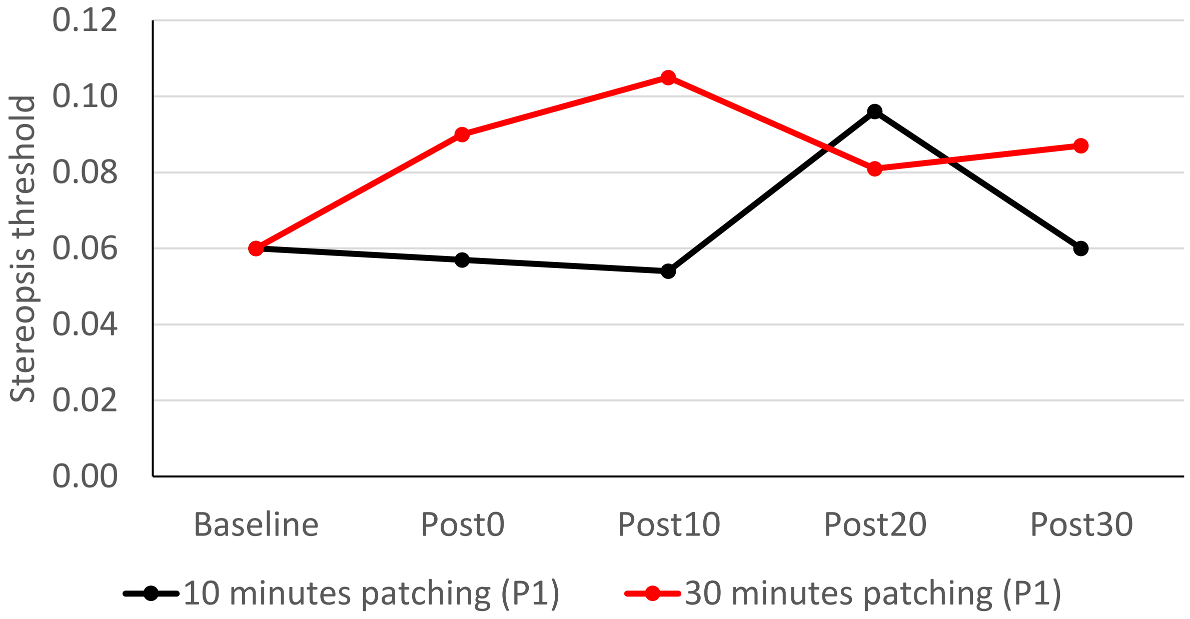


Figure 9.4.1. Stereopsis threshold over the course of the experiment (apparent depth behind the display plane). A comparison between 10 minutes of patching (black) and 30 minutes of patching (red).

On visual inspection (Figure 9.4.1, black), the stereopsis threshold remained at baseline levels after patching. It was only until 20 minutes after patching was there an increase in stereopsis threshold, i.e., a worsening of sensitivity to visual depth. However, by 30 minutes after patching, the stereopsis threshold returned to baseline levels. A worsening of stereopsis after monocular patching is perhaps not unexpected, given the dominance on the patched eye (see previous chapters), but there is no clear explanation for the singular worsening at 20 minutes after patching – perhaps it is an outlier. In comparison, 30 minutes of right-eye patching produced a steady increase in stereopsis threshold, peaking at 10 minutes after patching (Figure 9.4.1, red). Thereafter, the stereopsis threshold decreased, but not to baseline levels by 30 minutes after patching. Hence, contrary to 10 minutes of patching, with 30 minutes of patching, there seems to be a coherent picture of worsened stereopsis after patching, but from which there is also a gradual recovery.

9.4.2 Part 2

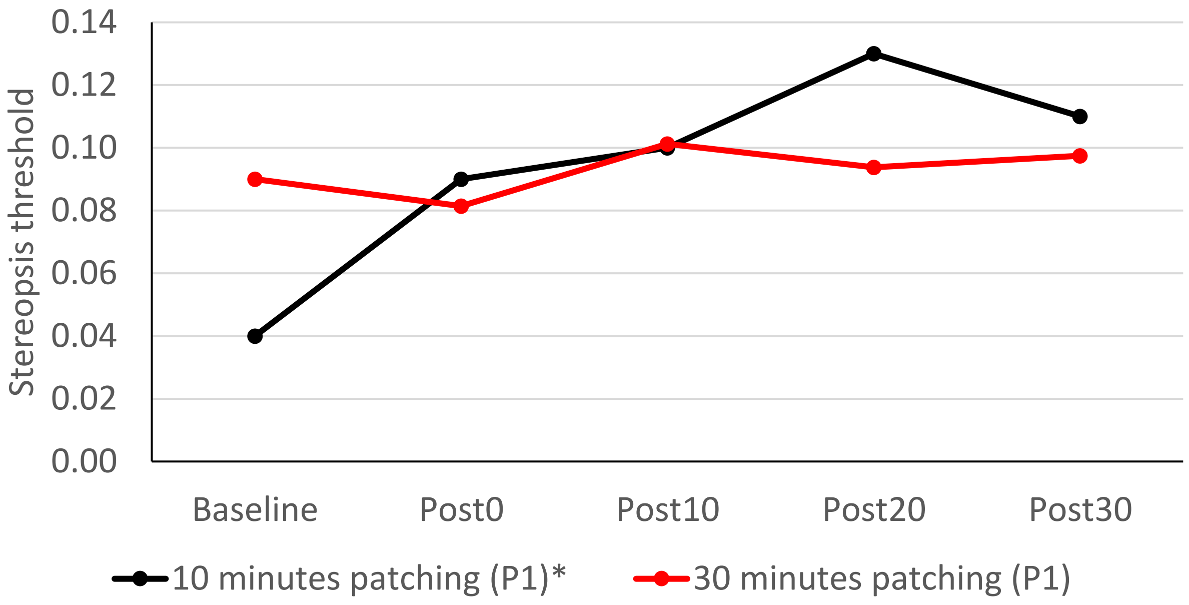


Figure 9.4.2. Stereopsis thresholds over the course of the experiment (apparent depth in front of the display plane). A comparison between 10 minutes of patching (black) and 30 minutes of patching (red). * Note that the baseline for 10 minutes of patching was calculated using the last response reversal – the test had unexpectedly terminated without enough reversals to perform the usual threshold calculation.

First, a caveat about the baseline for 10 minutes of patching: the visual test had unexpectedly terminated before the pre-set number of response reversals, so the usual threshold calculation could not be performed. Instead, the baseline stereopsis threshold was determined from the last response reversal. Visually inspecting the data (Figure 9.4.2, black), with 10 minutes of patching, there seems to be a steady increase in stereopsis threshold from baseline. It appears that the stereopsis threshold peaked at 20 minutes after patching, but this could be an outlier, given that in Part 1, there was an unexplained singular peak in stereopsis threshold also at Post20. Looking past the performance at 20 minutes after patching, there still appears to be an upward trend in stereopsis threshold. Hence, it appears that sensitivity to apparent depth in front of the display plane worsened after 10 minutes of right-eye patching. In comparison, after 30 minutes of patching, the stereopsis threshold held steady at around the baseline level throughout the course of the experiment (Figure 9.4.2, red line). For sensitivity to visual depth in front of the display plane, it seems as though 10 minutes of right-eye patching had more impact than with 30 minutes of right-eye patching

9.5 Chapter discussion

This chapter looked into the effects of patching the right eye only for 10 minutes, in an attempt to answer this pilot study's second knowledge gap on very short-term monocular patching. The interest in examining such short patching durations stem from the practical issues associated with patching: patching is uncomfortable and inconvenient, whilst prescribed patching durations can be lengthy, so adherence to patching prescriptions can be low (Gambacorta et al., 2018; Holmes & Clarke, 2006; Maconachie & Gottlob, 2015; Webber & Wood, 2005). Hence, the novel patching procedure examined in this project could be more attractive if it produces effects already from very short patching durations. Here, one participant was tested on 10 minutes of patching, who also participated in Experiment 1 which had 30 minutes of patching (see Chapter 8). Thus, there is a direct comparison of how one responds differently to 10 or 30 minutes of patching. With only one participant in this experiment (due to covid-related restrictions at the time), the results are individual and offer a preliminary indication of the effects from very short-term monocular patching, but cannot be generalised.

For this participant, the results from 10 minutes of right-eye patching were rather complicated. From Experiment 1, the binocular rivalry visual test was most promising of the three visual tests, as it had offered a coherent picture on the progression of eye dominance after 30 minutes of right-eye patching. Hence, taking the binocular rivalry visual test as a point of reference, 10 minutes of right-eye patching seems to have produced weaker effects than with 30 minutes of right-eye patching. Both dominance metrics (dominance ratio, dominance proportion) indicated that with 10 minutes of patching, the dominance effect on the patched eye was smaller, delayed by up to 10 minutes after patching, and less durable compared to the strong and immediate dominance effect elicited by 30 minutes of patching. It seems as though 10 minutes of patching is insufficient to produce a robust dominance effect.

However, the fusion metric and the other two visual tests (contrast letters, stereopsis), which Experiment 1 found to be rather non-indicative of patching effects, seemed to have picked up on several coherent trends after just 10 minutes of patching. First, a loose inverse relationship between the duration of fusion and eye dominance was seemingly found in this experiment. The duration of fused percept during binocular rivalry is thought to relate to the degree of binocular suppression, with longer durations of fusion associated with weak suppression (Dieter et al., 2017).

Here, after 10 minutes of patching, eye dominance shifted to the patched eye and the duration of fusion decreased; thereafter, eye dominance neutralised and the duration of fusion increased. This relationship between dominance and suppression seems coherent, yet was not apparent with 30 minutes of right-eye patching (no appreciable reduction in fusion despite the strong dominance effect after 30 minutes of patching). Most curiously, for the same participant, the duration of patching impacted stereopsis differently: 10 minutes of right-eye patching appeared to have worsened sensitivity to apparent depth *in front of*, not behind, the display plane, whereas 30 minutes of right-eye patching did the opposite, seemingly worsening sensitivity to apparent depth *behind*, not in front of, the display plane. Meanwhile, as measured with the contrast letters visual test, for the same participant both 10 and 30 minutes of right-eye patching seemingly produced similar trends in eye dominance: a similar magnitude of dominance shift to the patched eye, followed an eventual neutralisation of the induced dominance.

Overall, from this preliminary experiment, it is rather difficult to arrive at any definitive conclusions to fully answer the knowledge gap presented in this chapter. On one level, there was only one participant. This participant also participated in Experiment 1 (Chapter 8), and thus offered a direct comparison of how one responds differently to 10 or 30 minutes of patching, but is only an individual example. Ideally, the other participants in Experiment 1 would also participate in this experiment, such that there are more examples of the differences between 10 and 30 minutes of patching, and thus produce a more comprehensive picture of very short-term monocular patching. On another level, from just one participant, the difference in effects between 10 and 30 minutes of patching is quite inconsistent. The binocular rivalry visual test suggests weaker effects from 10 minutes rather than 30 minutes of patching. The contrast letters visual test suggests similar effects with both patching durations. The stereopsis test suggests a reversal in effect direction between the two patching durations. Based on Experiment 1, the binocular rivalry visual test was most robust of the three tests, so perhaps a weakening of effect from shorter patching durations is most likely. As a final level of complication, the participant had different baseline eye dominance statuses between the two experiments: slightly left-eye dominant for this experiment, slightly right-eye dominant in Experiment 1. This variability within the participant may have also contributed to the variability in patching outcomes, and comes on top of the variability between participants found in Experiment 1.

Chapter 10: Answering knowledge gap 6: What are the effects of alternately patching both eyes, and does it improve visual functions that require both eyes, such as stereopsis?

If the dominance effect on the patched eye constitutes a temporary monocular improvement, then would patching the left and right eyes in alternation improve visual functioning on both eyes, benefitting stereopsis? Previous studies have only patched one eye in a singular experimental episode. In this chapter, an attempt is made to answer if alternately patching the left and right eyes for 30 minutes (termed 'switch patching' hereafter for brevity) produces improvements in visual functioning (especially stereovision), more so than with 30 minutes of purely monocular patching.

10.1 Methods for Experiment 3

10.1.1 Participants

One participant (male, normal vision, right-handed, 25 years old) was recruited using the participant recruitment procedures outlined in Chapter 7.2.1. This participant also participated in Experiment 1 (30 minutes of right-eye patching; Chapter 8), and Experiment 2 (10 minutes of right-eye patching; Chapter 9). Hence, there was the opportunity here to compare between 30 minutes of switch patching and 30 minutes of exclusively right-eye patching.

10.1.2 Patching

First, the participant's right eye was patched for 15 minutes. Afterwards, the participant's left eye was patched for 15 minutes. Thus, the total patching duration was 30 minutes. As with Experiments 1 and 2, the eyepatch was a translucent piece of plastic enclosing the entire eye (Bernell BTEP+, resulting in 20/400 or worse vision; see Chapter 8.1.2 for definition). During patching, the participant watched television.

10.2 Results from the binocular rivalry visual test

10.2.1 Dominance ratio

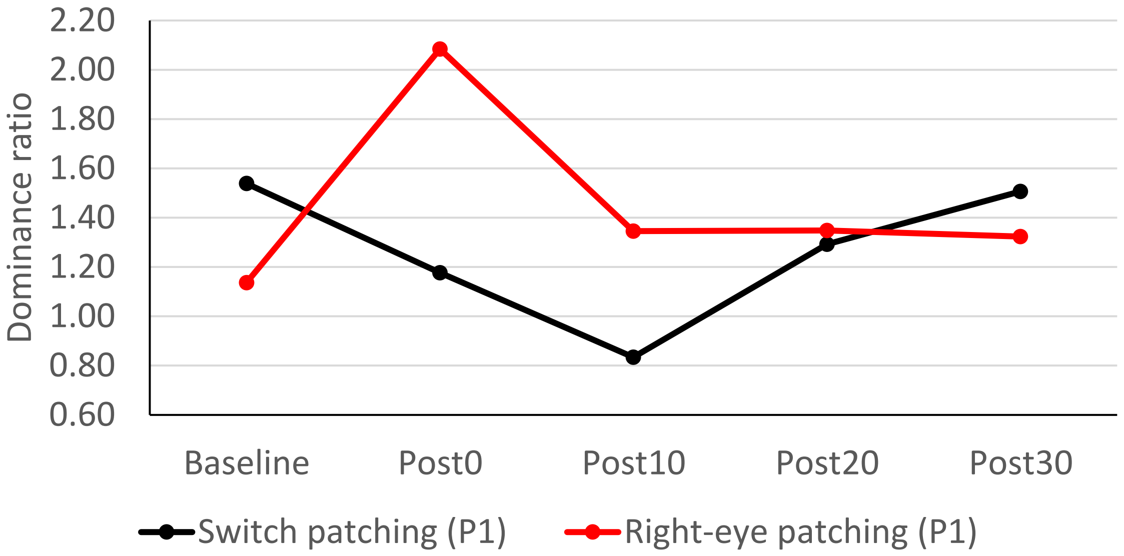


Figure 10.2.1. Dominance ratios over the course of the experiment, after 30 minutes of switch patching (black) or 30 minutes of right-eye patching (red). Both datasets originate from the same participant.

In binocular rivalry, the dominance ratio is the total time indicating the right-eye image, divided by the total time indicating the left-eye image. Hence, right-eye dominance is indicated by a ratio larger than one (more details in Chapter 7.3). Visually inspecting the dominance ratios (Figure 10.2.1, black), the participant started with a dominance ratio of almost 1.6, suggesting that the participant was noticeably right-eye dominant at baseline. After switch patching, the participant exhibited a steady downward trend in dominance ratio, eventually going below the binocular-neutral ratio of one by 10 minutes after patching. Thus, switch patching seemed to have redressed the initial right-eye dominance, and possibly induced a left-eye dominance. However, beyond 10 minutes after patching, the participant exhibited an upwards trend in dominance ratio, reaching baseline levels by 30 minutes after patching. Hence, switch patching appears to have produced a transient neutralising or leftwards dominance effect; the participant quickly returned to the initial right-eye dominance thereafter. In comparison, for the same participant, 30 minutes of right-eye patching produced a strong dominance effect on the right eye, which faded but seemingly persisted at a low level for at least 30 minutes after patching.

10.2.2 Dominance proportion

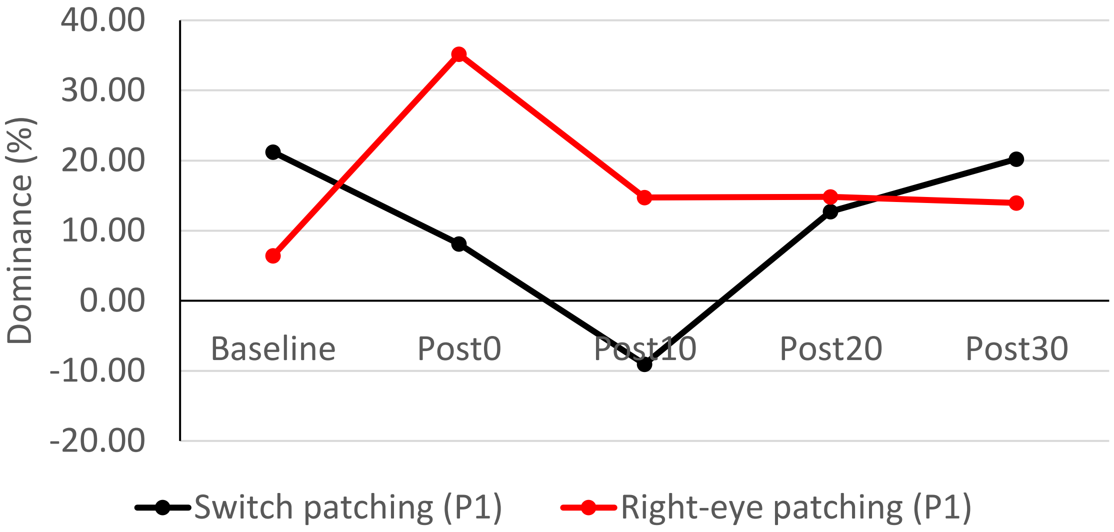


Figure 10.2.2. Eye dominance over the course of the experiment, for 30 minutes of switch patching (black), or 30 minutes of exclusively right-eye patching (red). Both datasets originate from the same participant.

The dominance proportion metric uses the proportion of time indicating each eye's view to determine eye dominance (Dieter et al., 2017; more details in Chapter 7.3). A positive percentage indicates right-eye dominance, a negative percentage indicates left-eye dominance. In Experiments 1 and 2, the dominance proportion metric largely concurs with the dominance ratio. Looking at the data from this experiment (Figure 10.2.2, black), the participant had a percentage of 20% at baseline, indicating a strong right-eye dominance. After switch patching, the percentage steadily decreased, reaching negative percentages by 10 minutes after patching, thus suggesting a neutralisation and possibly a shift to left-eye dominance. Thereafter, the percentage steadily increased, returning to the baseline right-eye dominance by 30 minutes after patching. In comparison, for the same participant, 30 minutes of exclusively right-eye patching (Figure 10.2.2, red) seemed to have induced a strong dominance effect to the right eye, which fades but apparently continued at a low level for the remainder of the experiment session. The trends found with the dominance proportion metric are similar to those found with the dominance ratio metric (Chapter 10.2.1).

10.2.3 Fusion

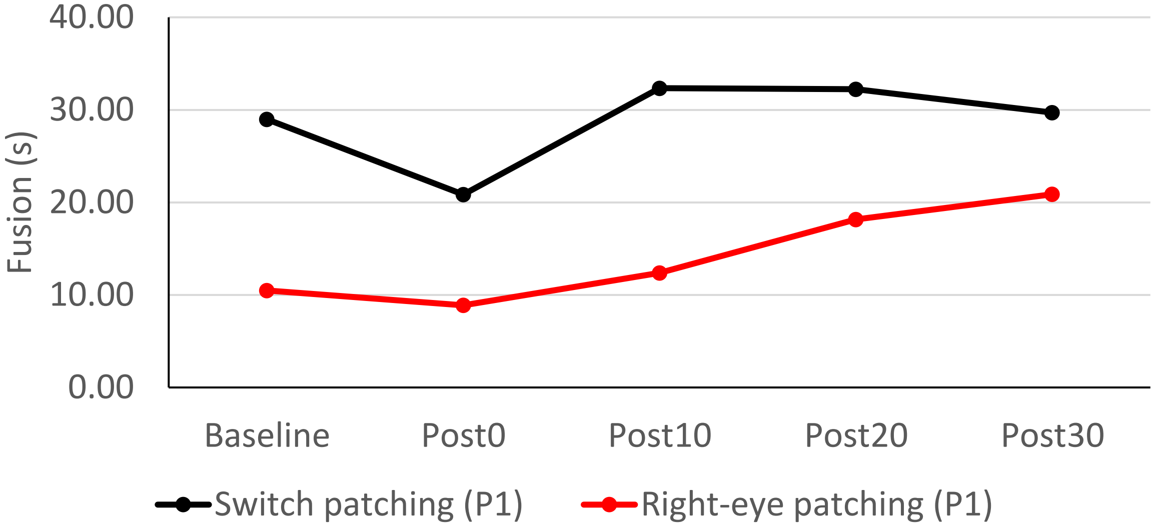


Figure 10.2.3. Duration of fused percept over the course of the experiment, with 30 minutes of switch patching (black), and 30 minutes of exclusively right-eye patching (red).

During binocular rivalry, there may be periods when both eyes' views are perceived simultaneously, as a fused percept. The fused percept is thought to be related to weak binocular suppression (Dieter et al., 2017). Looking at the durations of fused percept (Figure 10.2.3, black), there appears to be a reduction in fusion after switch patching. However, by 10 minutes after patching, the amount of fusion returned to baseline levels, which continued to the end of the experiment. In comparison, for the same participant, 30 minutes of right-eye patching (Figure 10.2.3, red) produced a negligible decrease in duration of fusion, but was followed by a steady increase thereafter, culminating in 10 seconds more fusion than at baseline. Note, even with the same participant, the duration of fusion (including at baseline) was always more in this experiment, than in the experiment with right-eye patching (Experiment 1).

10.3 Results from the contrast letters visual test

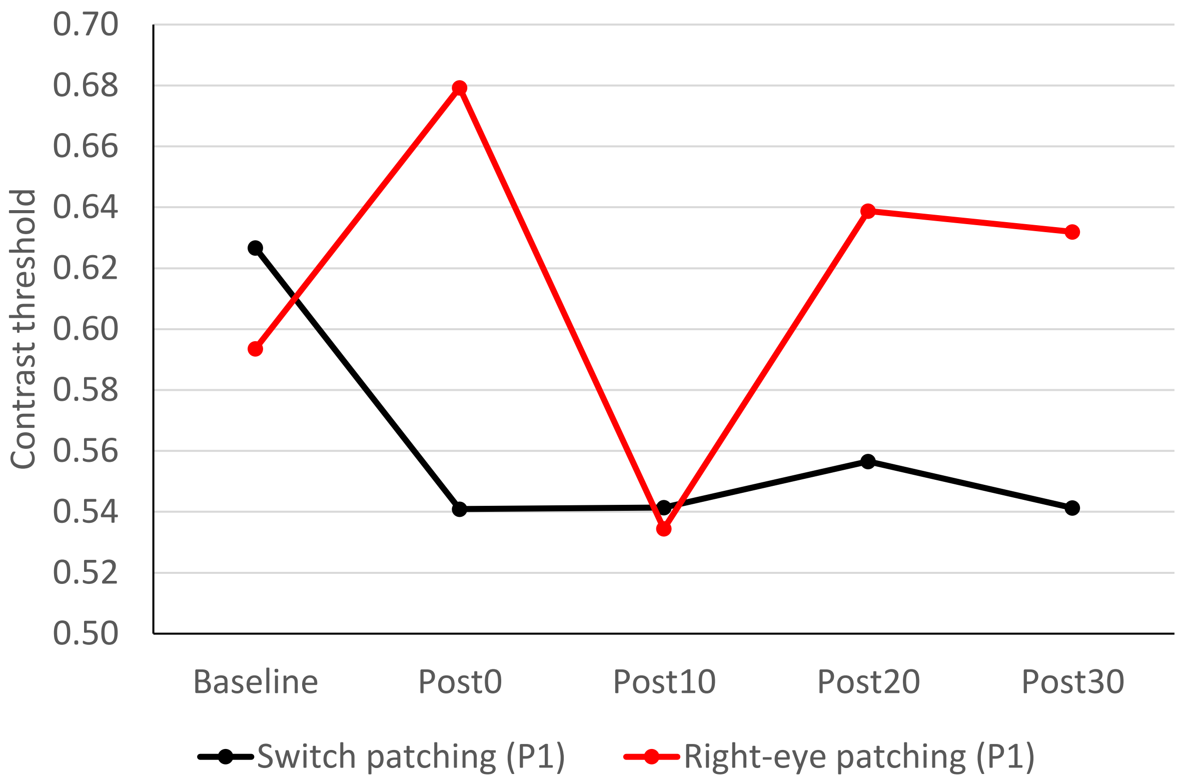


Figure 10.3. Contrast thresholds over the course of the experiment, for 30 minutes of switch patching (black) and 30 minutes of right-eye patching (red).

The contrast letters visual test presents letters conflicting in luminance across both eyes, to determine eye dominance (see Chapter 7.4). A larger contrast threshold above 0.5 indicates a stronger right-eye dominance. Visually inspecting the data (Figure 10.3, black), the participant started with a contrast threshold of almost 0.63, indicating a slight right-eye dominance. After 30 minutes of switch patching, the contrast threshold dropped to around 0.54 for the remainder of the experiment. Note, the test only measures for right-eye dominance, with a lower limit of 0.52. Hence, contrast thresholds arriving near the test's lower limit could indicate that the participant's right-eye dominance was neutralised or converted into left-eye dominance by switch patching. In comparison, 30 minutes of right-eye patching (Figure 10.3, red) induced an initial strong right-eye dominance, which then fluctuates; perhaps contrast sensitivity after monocular patching is unstable.

10.4 Results from the stereopsis visual test

Key to this chapter was whether switch patching would benefit both eyes, thereby improving stereopsis. Stereopsis was tested by varying the binocular disparity of the visual stimulus, in search of the participant's stereopsis threshold. The test was divided into two parts: Part 1 checks for sensitivity to apparent depth behind the display plane, Part 2 checks for sensitivity to apparent depth in front of the display plane. Details about this visual test are in Chapter 7.5.

10.4.1 Part 1

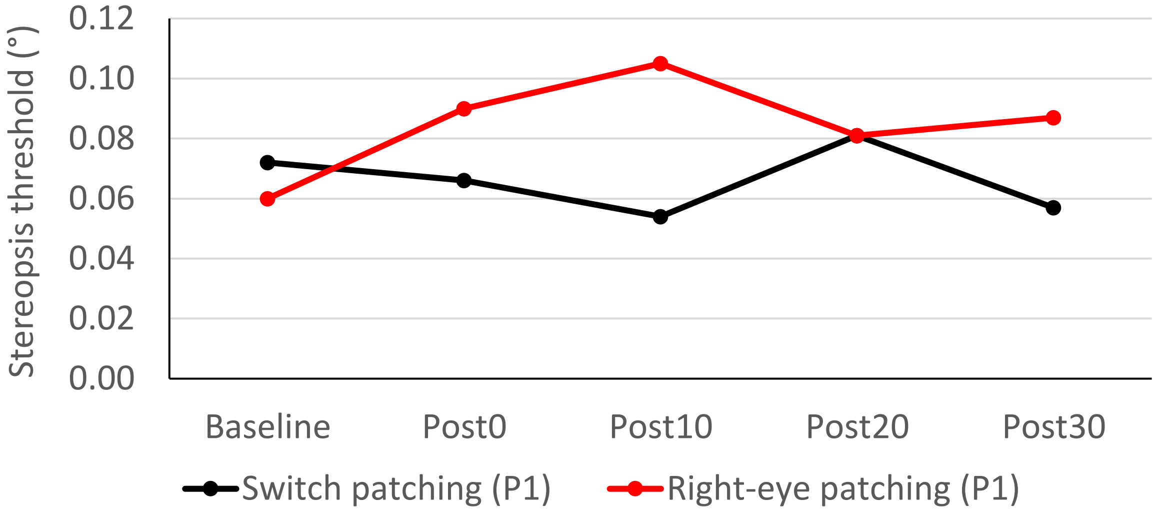


Figure 10.4.1. Stereopsis thresholds over the course of the experiment (apparent depth behind the display plane), for 30 minutes of switch patching (black), and 30 minutes of exclusively right-eye patching (red). Both datasets originate from the same participant.

Visually examining the data (Figure 10.4.1, black), it appears that after 30 minutes of switch patching, there was a slight downward trend in stereopsis threshold to the end of the experiment session, which could signify a small improvement in sensitivity towards visual depth. However, within this downward trend was also a singular rebound in stereopsis threshold at 20 minutes after patching, suggesting instability in stereopsis after patching. In comparison, for the same participant, after 30 minutes of right-eye patching (Figure 10.4.1, red), there was a steady increase in stereopsis threshold, peaking at 10 minutes after patching, and slightly decreasing thereafter. Increase in stereopsis threshold suggests a worsening of sensitivity to visual depth. Altogether, it appears that switch patching offered slight improvements in stereopsis, while the same duration of monocular patching worsened stereopsis.

10.4.2 Part 2

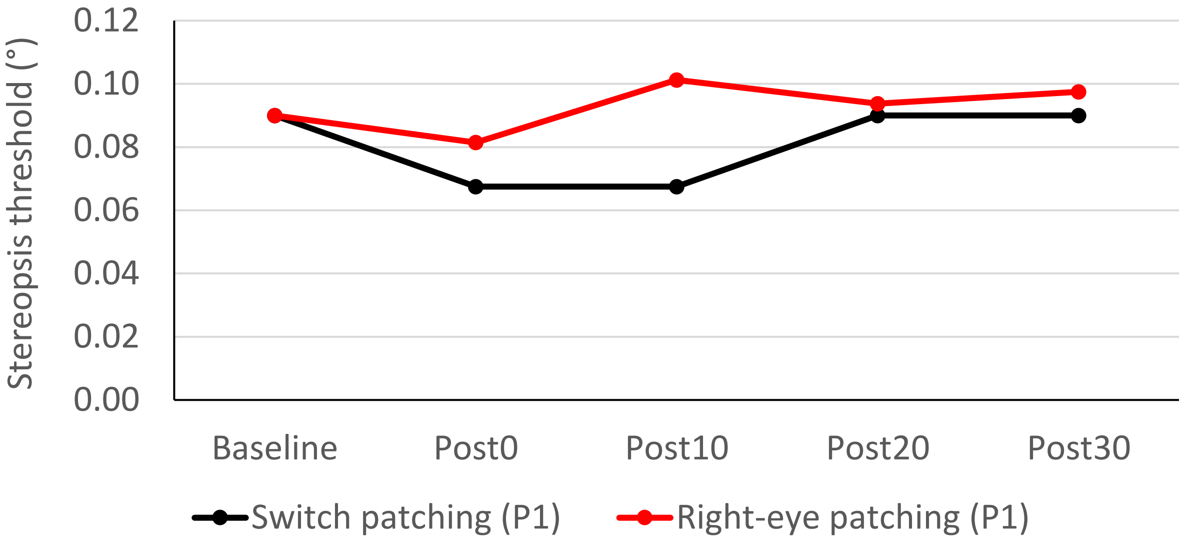


Figure 10.4.2. Stereopsis thresholds over the course of the experiment (apparent depth in front of the display plane), for 30 minutes of switch patching (black) and 30 minutes of right-eye patching (red). Both datasets originate from the same participant.

Visually inspecting the data (Figure 10.4.2, black), it appears that after 30 minutes of switch patching, there was a decrease in stereopsis threshold which lasted until 10 minutes after patching. This decrease in stereopsis threshold suggests a temporary improvement in sensitivity towards visual depth. Thereafter, the stereopsis threshold returned to baseline levels for the remainder of the experiment. In comparison, for the same participant, 30 minutes of right-eye patching (Figure 10.4.2, red) seems to have little effect on stereopsis, with the stereopsis thresholds at baseline levels throughout.

10.5 Chapter discussion

This chapter explored the effects of switch patching, where the left and right eyes were patched in alternation to see if this would benefit visual functioning (especially stereopsis), more so than with monocular patching, in an attempt to answer the third and final knowledge gap of this pilot study. Here, switch patching was implemented by first patching the right eye for 15 minutes, followed by the left eye for another 15 minutes, thus totalling 30 minutes of patching. There was only one participant (due to covid-related restrictions at the time), but this participant had also participated in Experiment 1, which patched the right eye exclusively for 30 minutes. Hence, a direct comparison between 30 minutes of switch patching against 30 minutes of right-eye patching was made. The inclusion of only one participant means the findings here are individual, with limited generalisability, but nonetheless offer a preliminary look into the effects of patching both eyes in alternation. On the whole, for this participant, there were two effects of switch patching.

The first effect, as measured in both tests of eye dominance (binocular rivalry, contrast letters), was that switch patching seemed to have a neutralising effect on the participant's right-eye dominance. The two visual tests do not fully agree on the specifics of this neutralising effect. In the binocular rivalry visual test, both eye dominance metrics (dominance ratio, dominance proportion) found switch patching to have reduced the participant's right-eye dominance, even turning into a slight left-eye dominance, but crucially, this neutralising effect was transient, because by 30 minutes after patching, the participant had returned to their baseline level of right-eye dominance. The contrast letters visual test also found switch patching to neutralise the participant's right-eye dominance, but according to this test, the neutralisation was durable, lasting for the remainder of the experiment session. Given the robustness of the binocular rivalry visual test in Experiment 1 (Chapter 8), perhaps more weight should be given to its account of switch patching producing a transient neutralising effect. Nonetheless, with two visual tests in agreement that the participant's initial right-eye dominance was reduced, switch patching may have some neutralising effects on binocular imbalance.

An argument could be made that the reduction in right-eye dominance was not a binocular neutralisation, but instead the imposition of left-eye dominance, on a participant who started as right-eye dominant at the time. The eye dominance metrics (dominance ratio, dominance proportion) found a slight left-eye dominance 10 minutes after switch patching. There was also a slight reduction in the duration of perceived fusion after switch patching, suggesting increased binocular suppression (Dieter et al., 2017), which for this participant, would be consistent with the onset of monocular dominance (Chapter 9; this participant perceived less fusion when becoming more monocularly dominant, i.e., a loose inverse relationship between duration of perceived fusion and dominance on the patched eye). If it was a left-eye dominance that was induced during switch patching, then it may have arisen from the patching sequence of right-eye first, then the left-eye, after which were the visual tests. If so, the patching effects were determined primarily by the latest episode of patching, the left eye in this case, not so much the patching sequence in its entirety. To test such a hypothesis, another experiment could be performed, patching the participant's left eye for 15 minutes, to see if the patching effects are similar to that found here with switch patching.

The second effect of switch patching, and perhaps a counterargument to suggestions that the switch patching here could in effect have been left-eye patching, was the reduction in stereopsis thresholds after switch patching. Reductions in stereopsis thresholds signify improved sensitivity to visual depth, and such reductions, albeit small in magnitude and duration, were found in both parts of the stereopsis visual test. Hence, for this participant, switch patching may have produced small improvements in sensitivity to visual depth. In Experiments 1 and 2, which patched the same participant monocularly and produced a dominance effect on the patched eye, stereopsis was found to have worsened slightly. Hence, at least in terms of stereopsis, and admittedly only for one participant, switch patching may not be entirely the same as monocular patching.

Overall, with preliminary data from only one participant, it is difficult to arrive at definitive statements to fully answer the knowledge gap presented for this chapter. The one participant did also participate in Experiments 1 and 2, so it was possible to see how the same person responds to different patching procedures. Ideally, the other participants in Experiment 1 would have participated in this experiment as well, such that there are more examples of how people respond to switch patching versus monocular patching, and from there, more generalisable findings may be discovered. With the caveat that the findings here are specific to the individual participant, as an answer to the knowledge gap, it seems as though switch patching may produce different effects to that produced from exclusively monocular patching. Most interestingly, for this participant, switch patching may have offered small and temporary improvements to stereopsis, while monocular patching was detrimental to stereopsis.

Chapter 11: Section discussion

This pilot study, in three experiments, offers a preliminary look at how short-term monocular patching might affect visual functioning. The development of this pilot study had its difficulties, and the experimentation occurred during covid; samples sizes were small (n=4 in Experiment 1, even n=1 in Experiments 2 and 3 – results entirely specific to that individual participant). Needing to collect as much data from few participants, and as a preliminary exploration into patching effects, the net was cast wide using several visual tests – perhaps not the best experiment design. Hence, this pilot study only offers some preliminary indications on how visual functioning changes after patching. Against this background, what was found in the three experiments of this pilot study?

Experiment 1 (Chapter 8) looked into the effects of patching the right eye for 30 minutes. Concurring with past research (e.g., Lunghi et al., 2011), after monocular patching, the patched eye became dominant, and this dominance shift was statistically significant despite having only four participants in the experiment. Hence, the shift in dominance towards the patched eye seems to be a robust effect of monocular patching. However, unlike past research (e.g., Lunghi et al., 2011), the dominance effect found here seemed not to last more than 10 minutes. The usefulness of such transient effects from patching seems doubtful. No other significant effects were found.

Experiment 2 (Chapter 9) explored the use of an even shorter patching duration (10 minutes on the right eye), as patching may be appealing if it can be performed quickly. One participant from Experiment 1 participated here, offering a comparison between 10 and 30 minutes of right-eye patching. For this participant, 10 minutes of patching produced a small, delayed and non-durable dominance effect on the patched eye, compared to the large and immediate dominance effect after 30 minutes of patching. Additionally, there seemed to be a slight worsening of stereopsis, for both durations of right-eye patching, but a larger sample is necessary to substantiate this individual's indications. Overall, for this participant, 10 minutes of monocular patching did not seem to produce strong patching effects.

Experiment 3 (Chapter 10) looked into switch patching, where the eyes were patched one after another, to check if this produced further effects (particularly with stereopsis) compared to monocular patching. There was one participant in this experiment, who had also participated in Experiment 1, thus offering a direct comparison between switch patching and right-eye patching. This participant started right-eye dominant, which was neutralised, or converted into a slight left-eye dominance, after switch patching. Unique among the three experiments, switch patching also seems to have produced small improvements in stereopsis for this participant. However, the effects from 30 minutes of switch patching, if there are indeed any, seem transient for this participant. To knowledge, switch patching is a novel procedure, and the possibility of stereopsis improvement could be studied further.

Altogether, the three experiments of this pilot study each have produced data and experience that could inform the direction of future studies. On a procedural front, for strong patching effects, patching durations likely need to be longer than 30 minutes, certainly not the 10 minutes as used in Experiment 2. Moreover, the visual test procedures can be streamlined: the contrast letters visual test was not as successful as anticipated and can be dropped, the stereopsis test could turn to a method of constant stimuli program to test both depth directions in one session, while the two eye dominance metrics (dominance ratio, dominance proportion) concur so only one is needed going forward. Amblyopia, one of the main motivations for this project, is defined by reduced visual acuity (e.g., Webber & Wood, 2005), so if short-term monocular patching were to be tested as a procedure for treating amblyopia, a measure of visual acuity should also be included as part of the visual test battery. In terms of research direction, the suggestion of slightly improved stereopsis after switch patching (Experiment 3, albeit n=1) is most interesting. First, the procedure of switch patching, and the possible finding of improved stereopsis are both novel. More importantly, a deficit associated with amblyopia, even after treatment, is the lack of stereovision (Webber & Wood, 2005), so there is a motivation to explore switch patching further.

However, before postulating further on modifications for the current study, an emphasis needs to be made about the transient patching effects found throughout the three experiments. Moreover, a key claim from Lunghi and colleagues (e.g., Lunghi et al., 2011; Lunghi et al., 2015b) is that short-term monocular patching calls upon visual cortex neuroplasticity, but the test paradigm here is perceptual and does not in itself offer direct evidence of neuroplasticity. Durability of effects, and genuine tapping of neuroplasticity, all for long-term improvements in visual functioning, are existential to the relevance of short-term monocular patching. Two directions for further studies on short-term monocular patching are proposed below, which may produce answers on the durability of patching effects and the extent of neuroplasticity involved. Additionally, an alternative to monocular patching is considered.

Animal study

As the start of a study on short-term monocular patching, when the effects of patching are unknown, animal experimentation may be more appropriate, laying the groundwork for further experimentation, which at a later stage when more is known, can human participants be involved. The relevance of short-term monocular patching hinges on effects that are productive (as in bringing about improvements in visual functioning) and durable. Durability possibly stems from changes in the brain structure after patching – neuroplasticity. There can be two questions: does short-term monocular patching tap into neuroplasticity, and if so, is this neuroplasticity productive on visual functioning? Animal experimentation could answer these two fundamental questions. Kittens have been used in past studies on abnormal visual functioning (e.g., Held & Hein, 1963; Mitchell & Gingras, 1998; Wiesel & Hubel, 1965) and the experimental procedures range from artificially induced abnormal visual experiences (e.g., eyelid suturing for monocular deprivation), to neural recordings, and dissection. For example, kittens which received eyelid suturing during the first three months of life later displayed behavioural deficits on visual tasks, reduced binocular neural activity, and reduced matter in visual brain areas (Wiesel & Hubel, 1965).

The range of invasive techniques possible in animal studies could offer more direct answers to the neuroplasticity and productivity surrounding short-term monocular patching, which in human studies, is only indirectly shown via measurements of neurotransmitters associated with neuroplasticity, and perceptual tests (e.g., Lunghi et al., 2011; Lunghi et al., 2015b). A possible animal study would involve kittens which received eyelid suturing to one eye for the first three months, which seems to be the critical period for cat visual development (see Wiesel & Hubel, 1965), thereby inducing an analogue of amblyopia in the kittens. Afterwards, variations of the short-term monocular patching procedure can be applied to the experimental group of kittens (e.g., differing in patching duration, switch versus purely monocular patching). Finally, the behavioural, neurological and anatomical development of the patched kittens could be compared to a control group of kittens which had eyelid suturing, but no patching. Thus, if short-term monocular patching is effective, then this might be observable in behavioural improvements on visual tasks, and in brain anatomy (e.g., enlargement of visual areas), which would suggest post-critical period neuroplasticity. Establishing such anatomical changes and behavioural improvements are perhaps necessary first steps, to determine that short-term monocular patching is both durable and productive, and therefore has promise for further experimentation.

Activity during patching

If the patching effects found in this pilot study seem small, is this attributable to the observers' inactivity during patching? Naturalistically, during childhood when one's visual system is in development, one moves about to explore, reach out and manipulate objects in the environment – a synergy of vision and motor, which seems essential to developing the visual system (James, 2010). In previous experiments, manual copying of letters (i.e., vision and motor) was associated with greater activity in the visual cortices than with passive letter viewing (James, 2010), and participants who actively controlled their exploration of objects recognised the objects faster than those who passively viewed a video of the exploration (Harman, Humphrey, & Goodale, 1999; James, Humphrey, & Goodale, 2001). Additionally, kittens which actively explored their environment developed normal visual functioning, while those which only passively viewed did not (Held & Hein, 1963). Altogether, the literature suggests that the visuo-motor synergy is necessary for visual learning and development. Hence, in this pilot study, with participants only passively watching TV during patching, there could have been quite limited visual processing under patched conditions, perhaps then resulting in limited patching effects.

A potential study on short-term monocular patching could explore the contribution of activity on the patching effect. To begin with, the study could experiment with animal subjects (e.g., kittens), as an extension of the previous point proposing animal studies. There would be a sample of monocularly-deprived animals (via eyelid-suturing during their critical period for visual development). Then, the animals would be paired in a yoked apparatus (see setup in Held & Hein, 1963): one animal is patched and moves about, and through the yoke and carrier, the other animal is moved along and only passively views. After patching, a behavioural, neurological and anatomical comparison could be made between the active-passive animal pair. As above, the importance of activity during patching would be indicated by greater behavioural improvements, and neurological and anatomical recovery in the visual areas, for the active group of animals. For human research, the participants will be assessed on visual tests (a modification of this pilot study). For the active condition, activities calling on vision and motor (whilst sat down, for safety) could be drawing tasks or building with toy blocks. For the inactive controls, the participants would only watch videos of drawing or block building. A comparison of patching effects between the active and inactive groups can then be made. If the visuo-motor synergy is important in visual learning, then patching effects should be larger for the active group than with the inactive control group.

Dichoptic training versus monocular patching

It has been noted that there are some theoretical objections to monocular patching as a treatment for amblyopia: it trains one eye, hoping for eventual binocularity, when instead, amblyopia may fundamentally be a binocular imbalance, with monocular deficits a consequence (Hess & Thompson, 2013). Hence, a new line of amblyopia research explores training binocularity through dichoptic viewing: presented between left and right eyes are contrast-adjusted complementary images, such that the full view is visible only when there is binocular combination (Hess & Thompson, 2015). In studies with both dichoptic training and monocular patching groups, dichoptic training seems to have the edge, producing similar improvements in visual acuity, stereoacuity and reading speed but in less time than patching (Gambacorta et al., 2018), or altogether larger and broader improvements than patching (Vedamurthy et al., 2015), or produced further visual acuity and stereoacuity gains after the initial monocular patching effects have saturated (Liu & Zhang, 2018; Zhang, Cong, Klein, Levi, & Yu, 2014). Should further studies on short-term monocular patching be fruitless, dichoptic training could be another research direction.

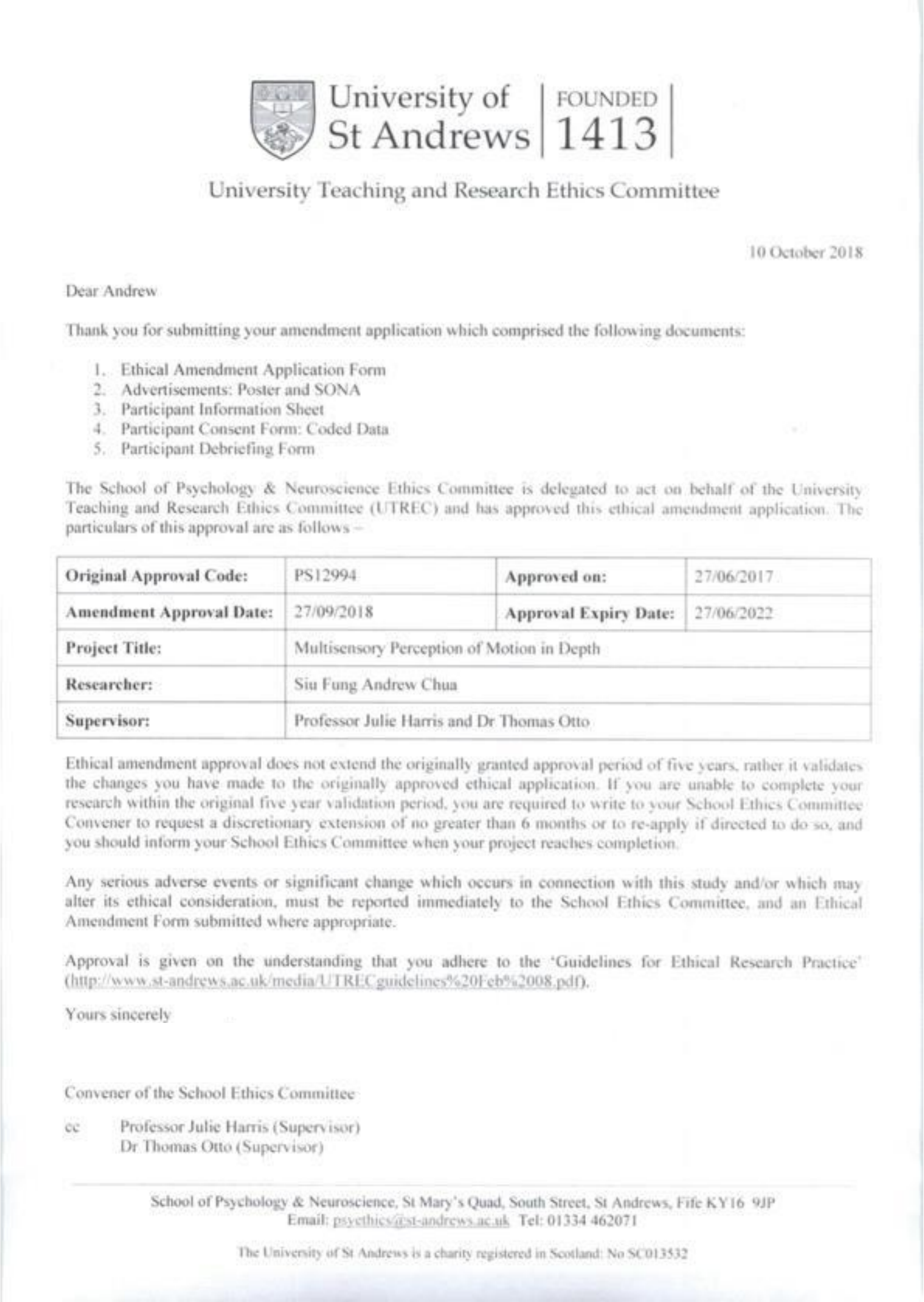
Chapter 12: Thesis conclusion

In two projects, this thesis examined two directions in which visual perception and functioning might be enhanced. The first project sought to combine visual signals with auditory signals, examining if the audio-visual conjunction would benefit decision-making in urgent real-life scenarios – looming. Looming – the motion towards oneself, implying impending danger – can be signalled visually and auditorily, and there is a survival advantage in responding quickly to such signals (e.g., Franconeri & Simons, 2003; Neuhoff, 2001). In a series of four behavioural experiments, covering basic versus multi-cue 'realism' stimuli and refining the stimulus timing, no evidence was found for speedy multisensory processing specially towards audio-visual looming signals – the RSEs towards the ALVL condition were not particularly large, and seems to counter the suggestions of 'selective integration' (Cappe et al., 2009). Rather, the comparative approach and computational modelling analyses showed that a simple probability summation rule with interactions allowed (a single negative correlation parameter on RTs, two additional-noise parameters) could explain the behavioural performance towards looming signals. The implication of probability summation is that the audio-visual RTs are an inheritance from the unisensory RTs: RTs in the audio-visual looming condition are fast because RTs in the unisensory looming components are also fast. There seems not to be a special processing mechanism just for processing audio-visual looming signals.

The second research direction of this thesis was to test if a novel eye patching technique would be practical and useful as a procedure for improving visual functioning, healthy or impaired. The eye patching technique, termed 'short-term monocular patching', was previously found to increase dominance on the patched eye, purportedly demonstrating latent neuroplasticity in adult visual systems (Lunghi et al., 2015a; Lunghi et al., 2011; Lunghi et al., 2015b), that might open possibilities in treating amblyopia. A pilot study consisting of three experiments was conducted, trialling short patching durations, patching both eyes in alternation, and measuring the changes in eye dominance and stereopsis. From 30 minutes of monocular patching, with only a sample of four participants, a statistically significant dominance effect on the patched eye was found, concurring with past research (Lunghi et al., 2011). Yet the sample size was small (even n=1); the importance of the pilot study lies not so much in its data per se, but in the experience and preliminary indications in the data, which informs the modifications and possible directions for future studies.

Bringing this thesis to a close, my experimentation brings some insights about vision. First, the conjunction of vision and audition is beneficial in responding to looming motion, but apparently not via an intuitively appealing special mechanism; only a simple interactive probability summation rule is needed. Second, adult visual functioning can briefly be manipulated using quick periods of eye patching, but the potential of patching will require further experimentation.

Appendix A – Audio-visual looming project ethical approval letter



University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

10 October 2018

Dear Andrew

Thank you for submitting your amendment application which comprised the following documents:

1. Ethical Amendment Application Form
2. Advertisements: Poster and SONA
3. Participant Information Sheet
4. Participant Consent Form: Coded Data
5. Participant Debriefing Form

The School of Psychology & Neuroscience Ethics Committee is delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has approved this ethical amendment application. The particulars of this approval are as follows –

Original Approval Code:	PS12994	**Approved on:**	27/06/2017
Amendment Approval Date:	27/09/2018	**Approval Expiry Date:**	27/06/2022
Project Title:	Multisensory Perception of Motion in Depth		
Researcher:	Siu Fung Andrew Chua		
Supervisor:	Professor Julie Harris and Dr Thomas Otto		

Ethical amendment approval does not extend the originally granted approval period of five years, rather it validates the changes you have made to the originally approved ethical application. If you are unable to complete your research within the original five year validation period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and an Ethical Amendment Form submitted where appropriate.

Approval is given on the understanding that you adhere to the 'Guidelines for Ethical Research Practice' (http://www.st-andrews.ac.uk/media/UTRECguidelines%20Feb%2008.pdf).

Yours sincerely

Convener of the School Ethics Committee

cc Professor Julie Harris (Supervisor)
Dr Thomas Otto (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

The University of St Andrews is a charity registered in Scotland: No SC013532

Appendix B – Individual AIC-selected models

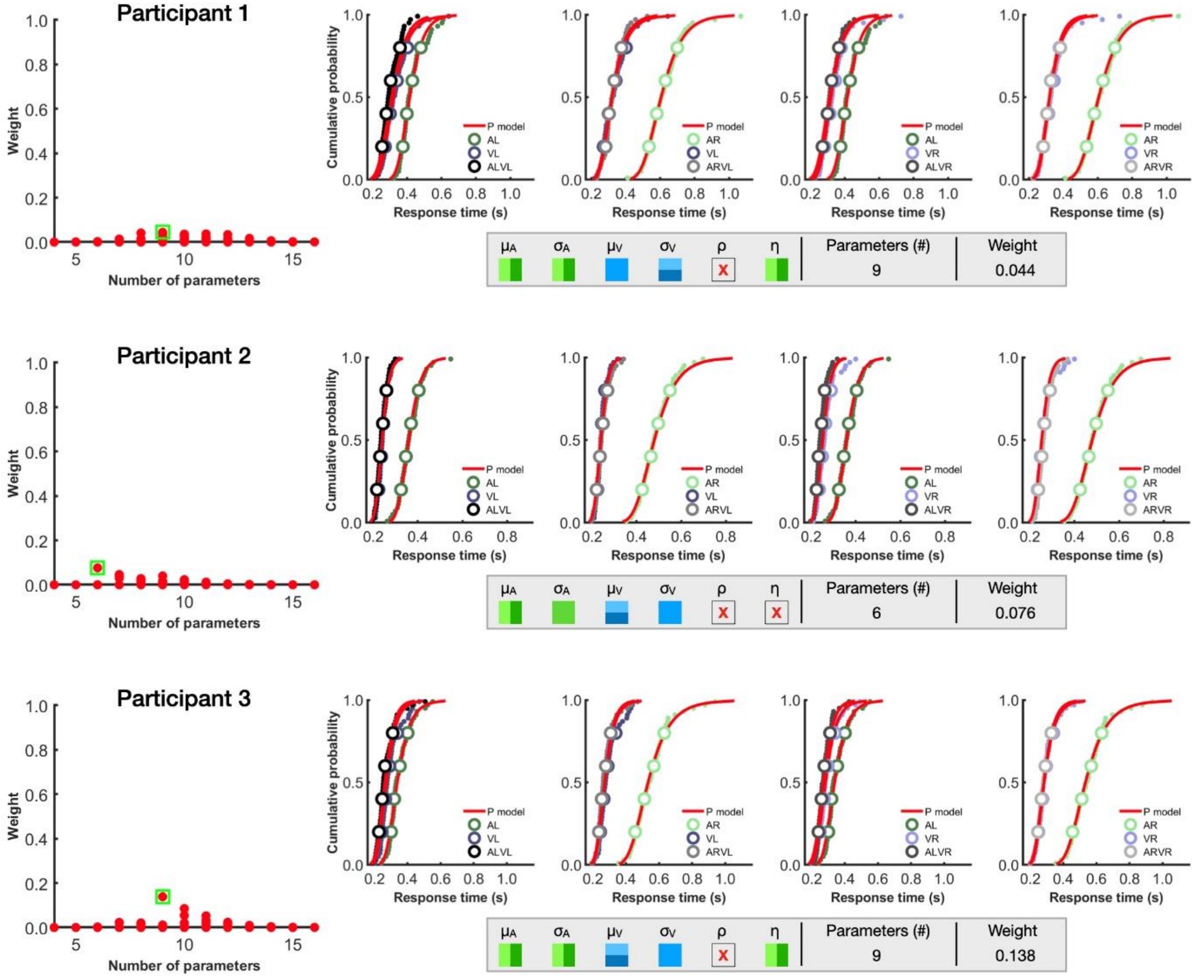
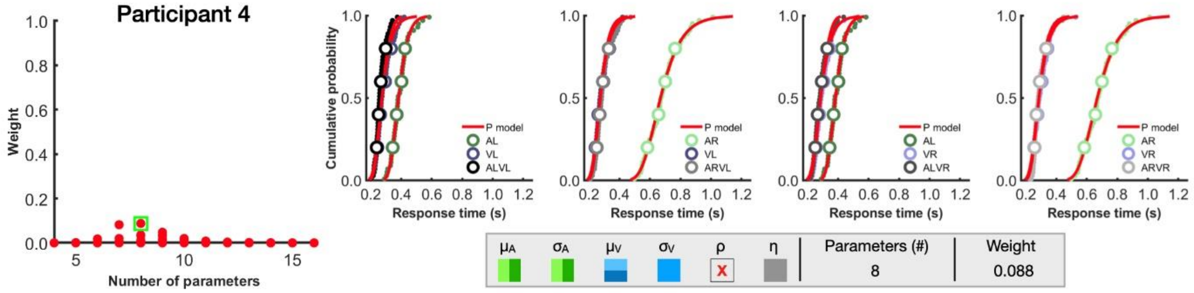
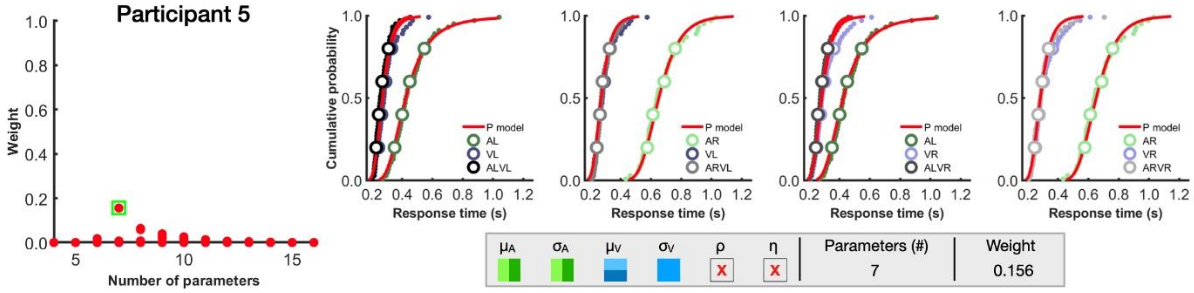


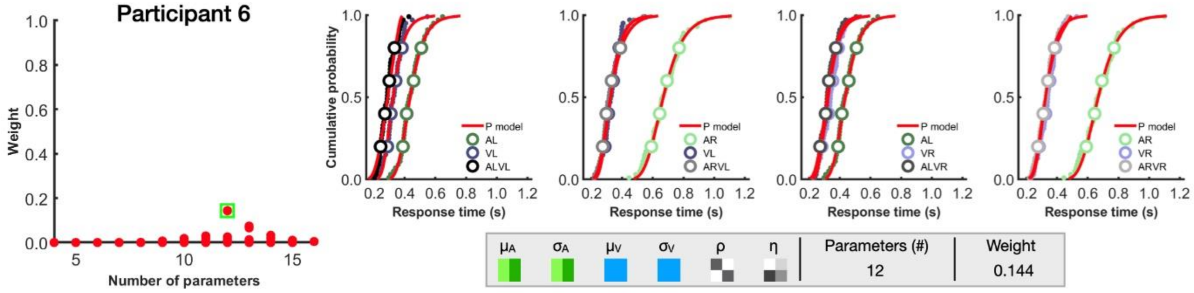
Figure B. Individual model selection using the AIC (Akaike, 1974; Wagenmakers & Farrell, 2004). In each panel, from left to right, there are the AIC weights for each of the 576 candidate models (model with highest AIC weight selected, highlighted in a green box), the selected model's (red line) fit to the participant's data (dots), below which is the parameterisation and AIC weight for the participant's selected model (see Figure 5.1.1 for the definitions of parameter symbols). This figure continues below, showing the individual model selections for all 20 participants.



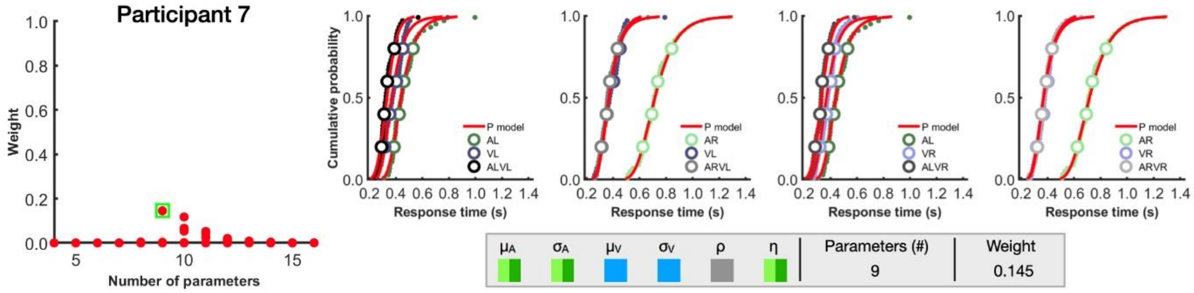
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	☒	▨	8	0.088



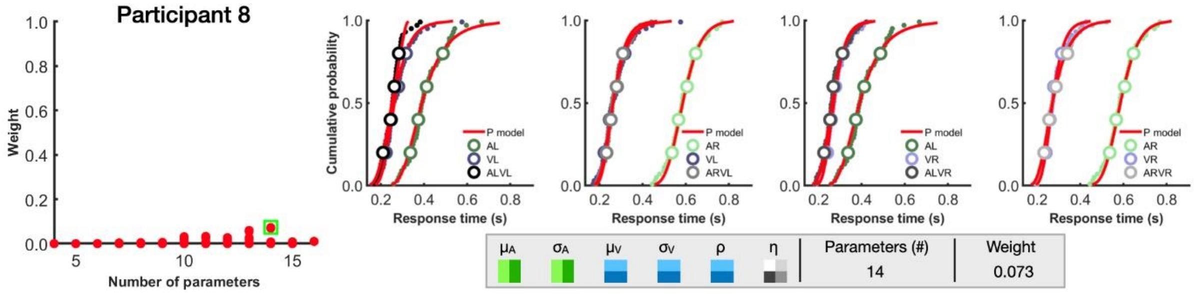
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	☒	☒	7	0.156



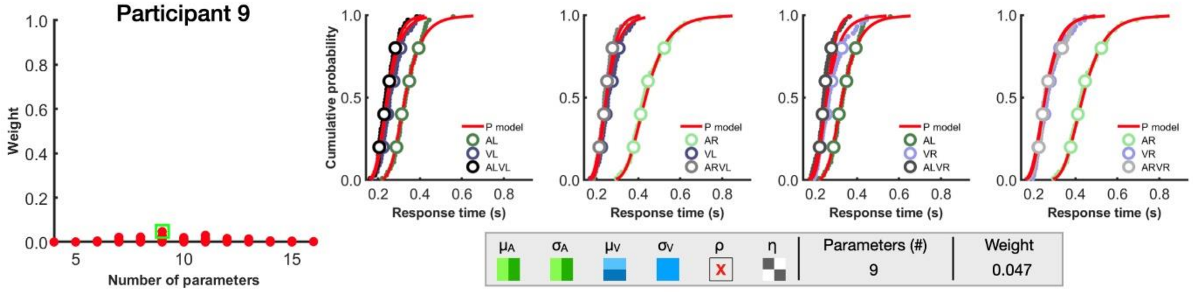
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	▨	▨	12	0.144



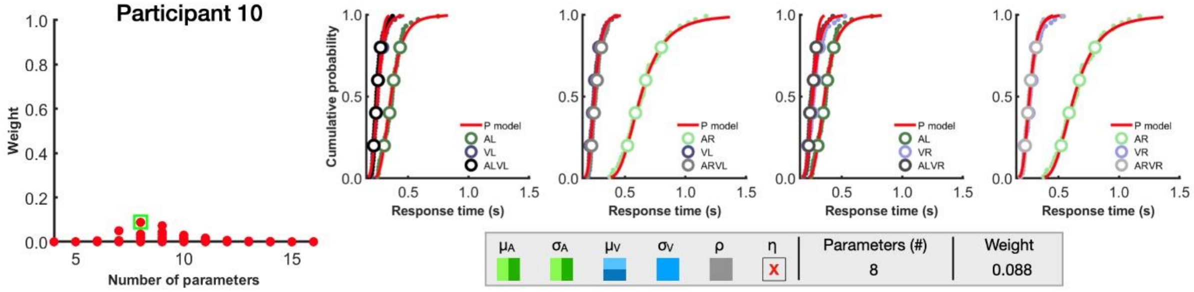
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	▨	🟩	9	0.145



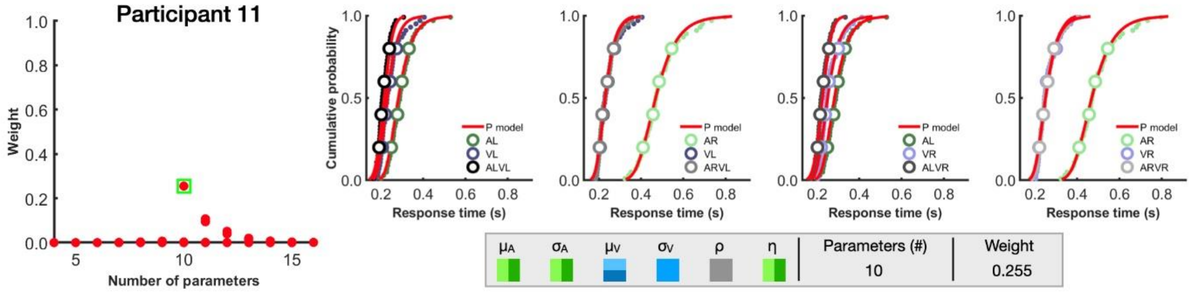
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	🟦	▨	14	0.073


Participant 9

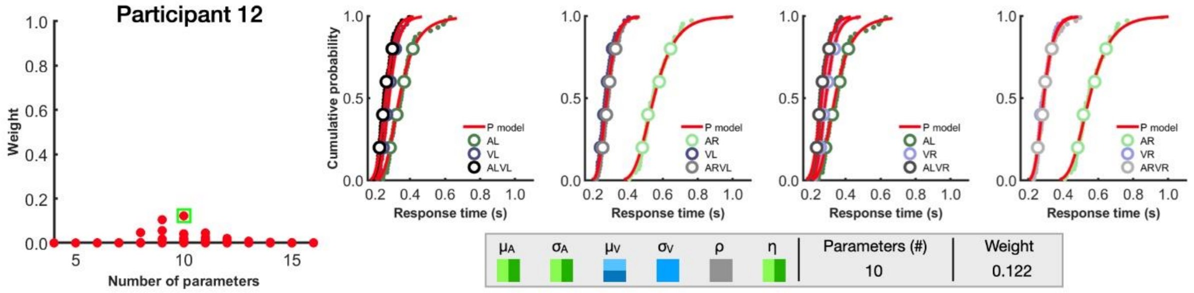
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	☒	▨	9	0.047


Participant 10

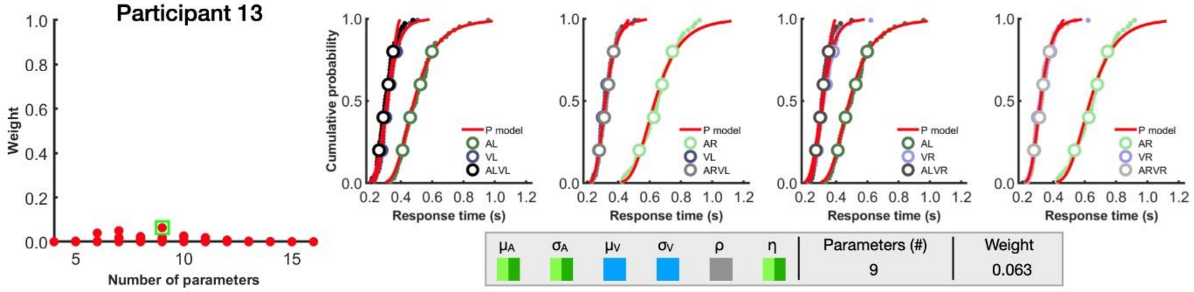
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	◼	☒	8	0.088


Participant 11

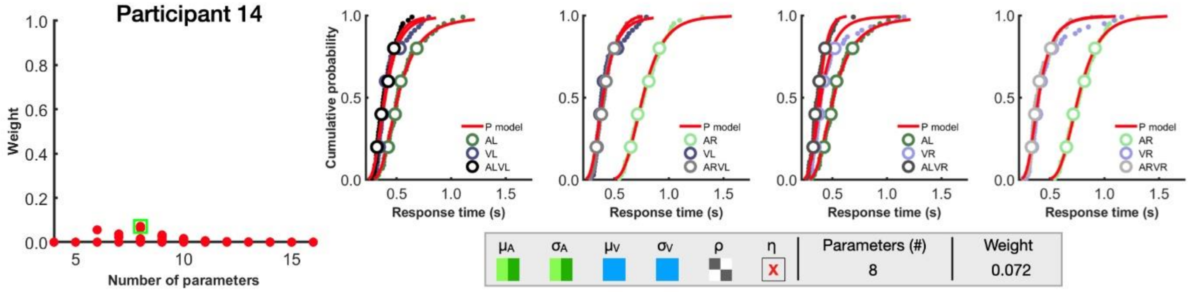
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	◼	🟩	10	0.255

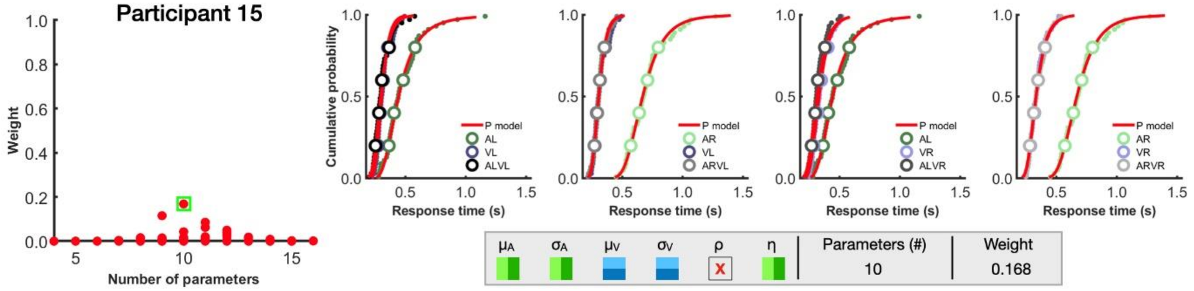

Participant 12

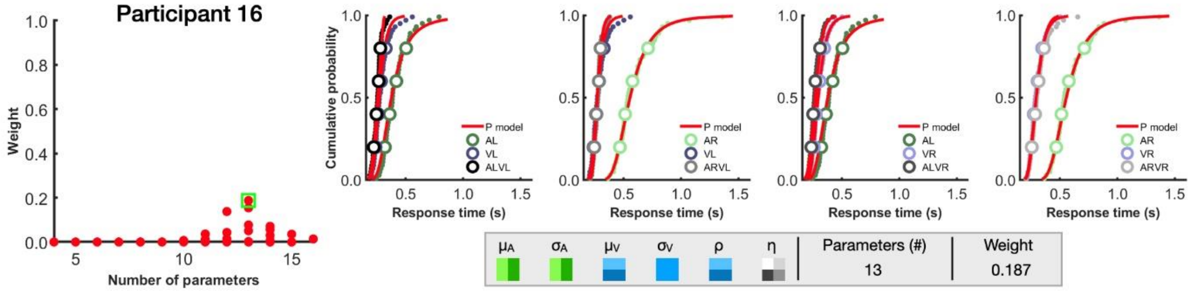
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	◼	🟩	10	0.122

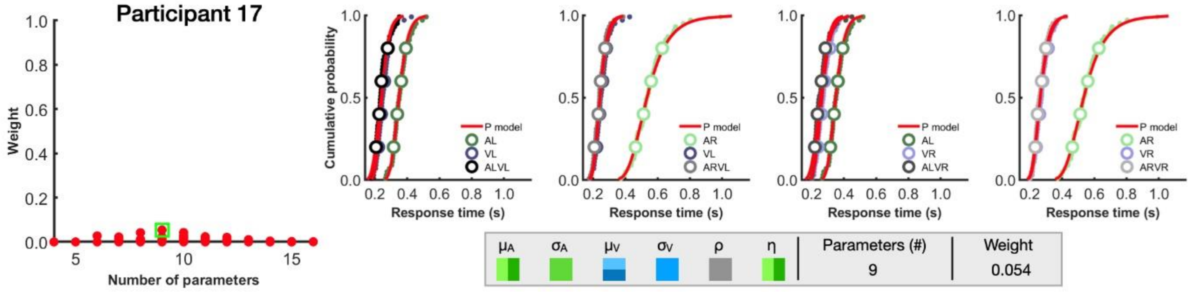

Participant 13

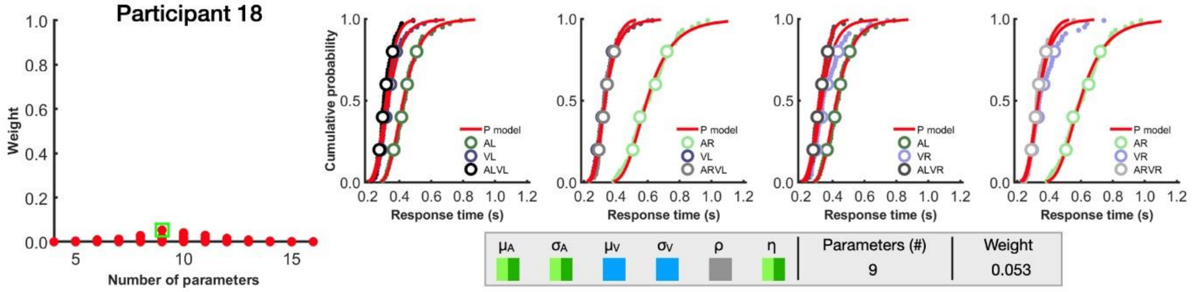
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	◼	🟩	9	0.063

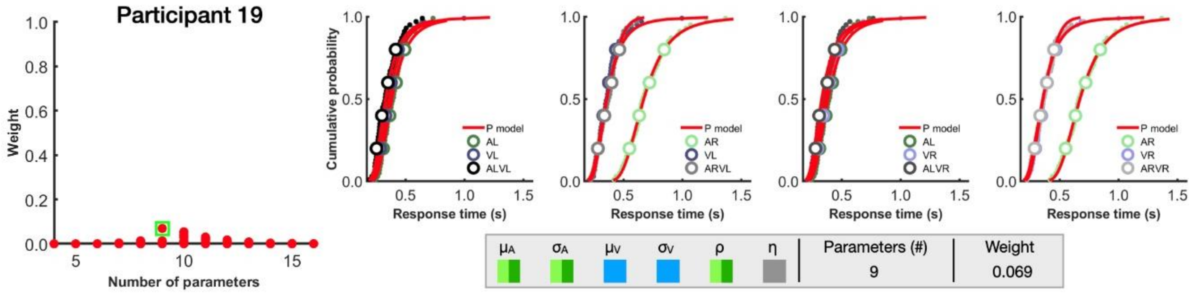

Participant 14
Weight
Number of parameters
Cumulative probability
Response time (s)
P model
AL
VL
ALVL
AR
ARVL
VR
ALVR
ARVR
μA σA μV σV ρ η
Parameters (#) 8
Weight 0.072


Participant 15
Weight
Number of parameters
Cumulative probability
Response time (s)
P model
AL
VL
ALVL
AR
ARVL
VR
ALVR
ARVR
μA σA μV σV ρ η
Parameters (#) 10
Weight 0.168

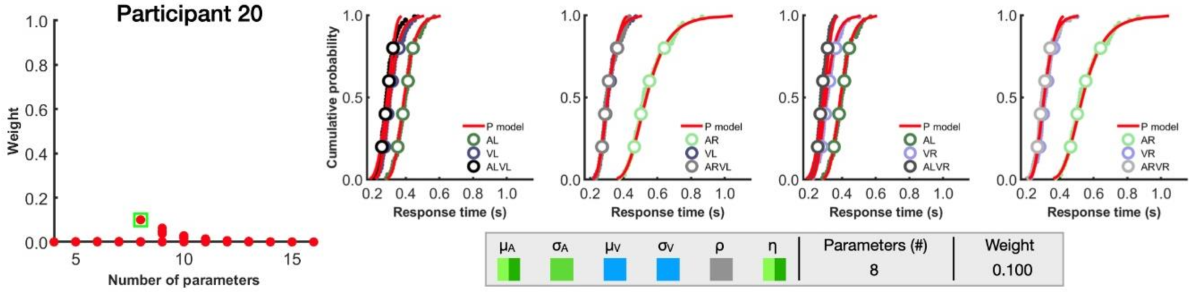

Participant 16
Weight
Number of parameters
Cumulative probability
Response time (s)
P model
AL
VL
ALVL
AR
ARVL
VR
ALVR
ARVR
μA σA μV σV ρ η
Parameters (#) 13
Weight 0.187


Participant 17
Weight
Number of parameters
Cumulative probability
Response time (s)
P model
AL
VL
ALVL
AR
ARVL
VR
ALVR
ARVR
μA σA μV σV ρ η
Parameters (#) 9
Weight 0.054


Participant 18
Weight
Number of parameters
Cumulative probability
Response time (s)
P model
AL
VL
ALVL
AR
ARVL
VR
ALVR
ARVR
μA σA μV σV ρ η
Parameters (#) 9
Weight 0.053

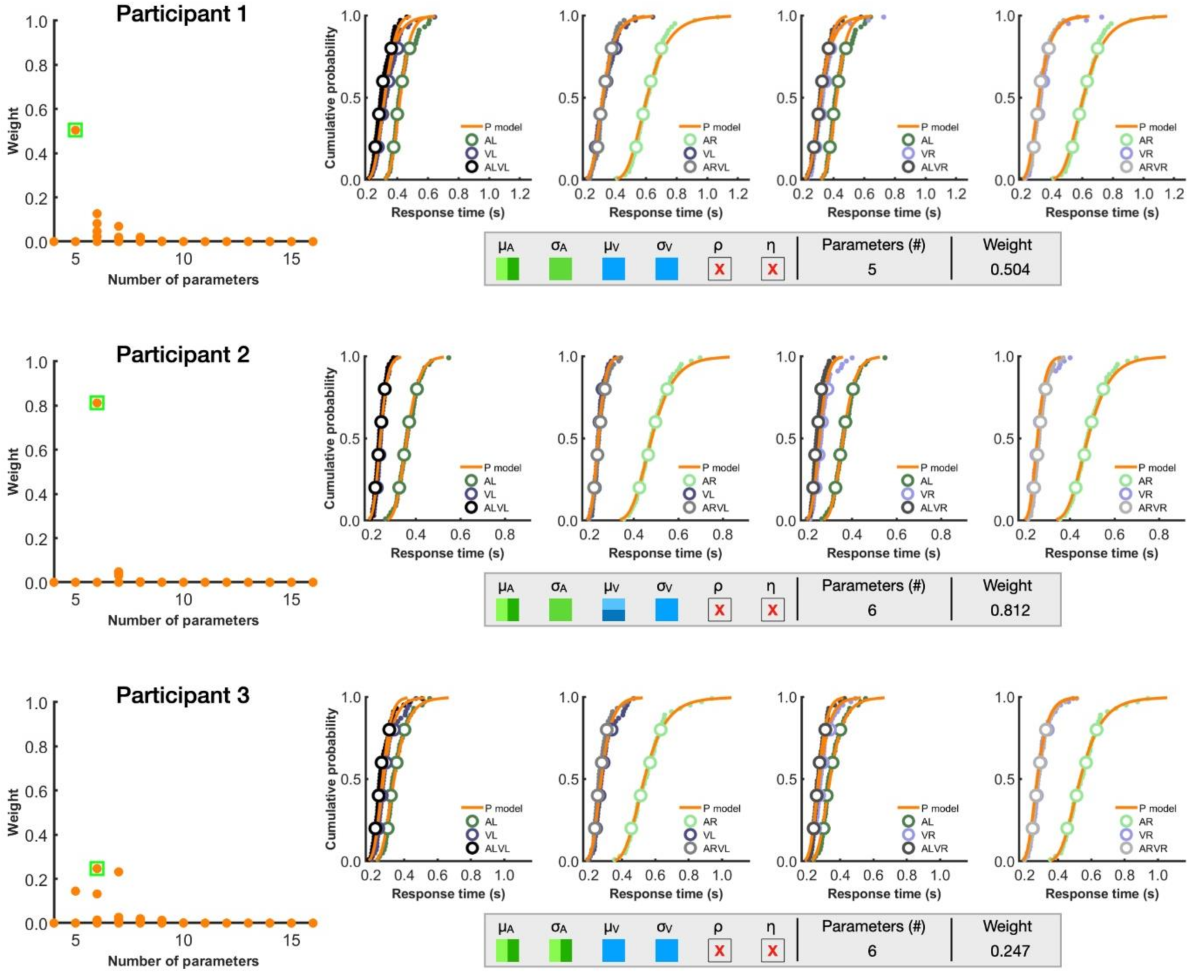


μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩🟩	🟦	🟦	🟩🟩	◼	9	0.069



μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩	🟦	🟦	◼	🟩🟩	8	0.100

Appendix C – Individual BIC-selected models



Participant 1

μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩	🟦	🟦	☒	☒	5	0.504

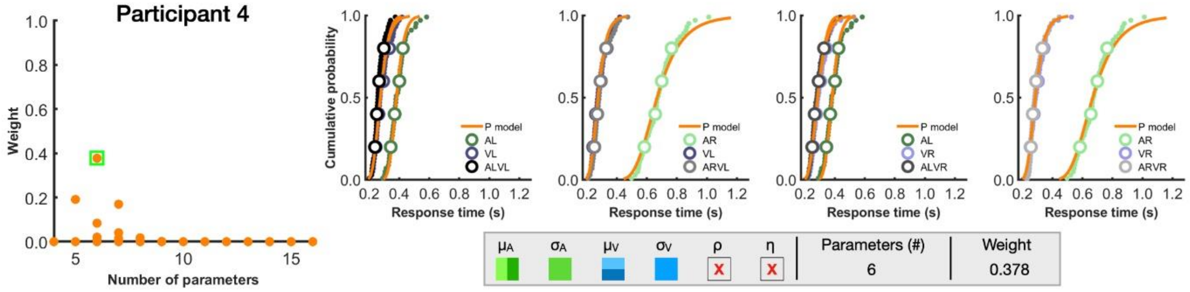
Participant 2

μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩	🟦🟦	🟦	☒	☒	6	0.812

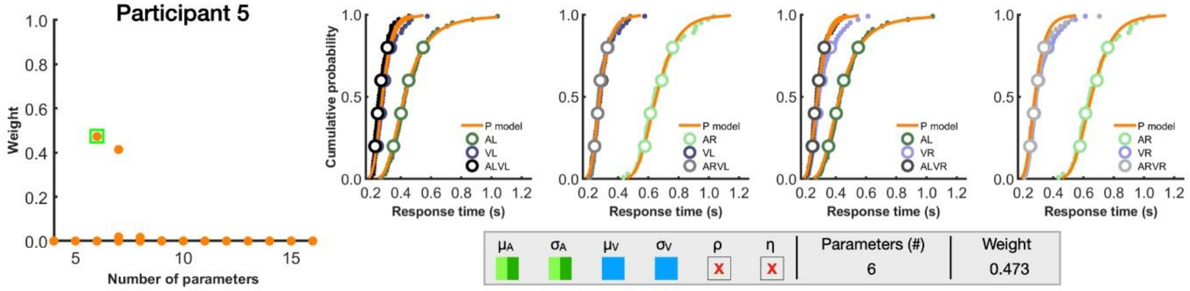
Participant 3

μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩🟩	🟦	🟦	☒	☒	6	0.247

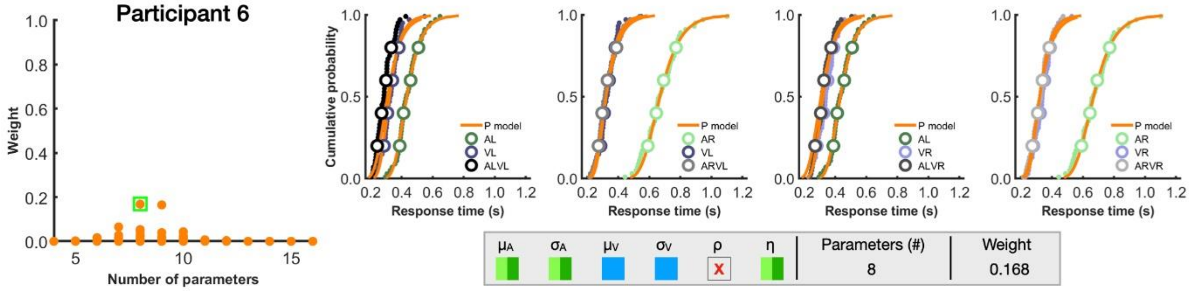
Figure C. Individual model selection using the BIC (Schwarz, 1978). In each panel, from left to right, there is the participant's BIC weights for each of the 576 candidate models (model with highest weight chosen, highlighted in green box), the selected model's fit (orange line) to the participant's behavioural data (dots), below which is the parameterisation of the selected model (see Figure 5.1.1 for the parameter symbols). This figure continues below, showing model selection for all 20 participants.



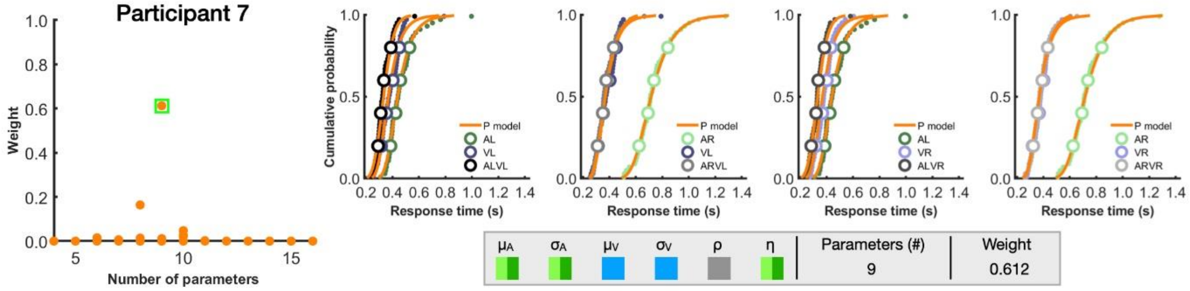
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	☒	☒	6	0.378



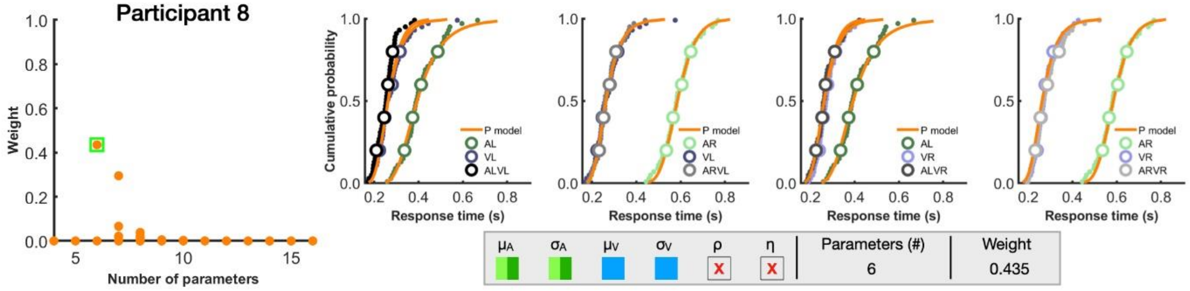
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	☒	☒	6	0.473



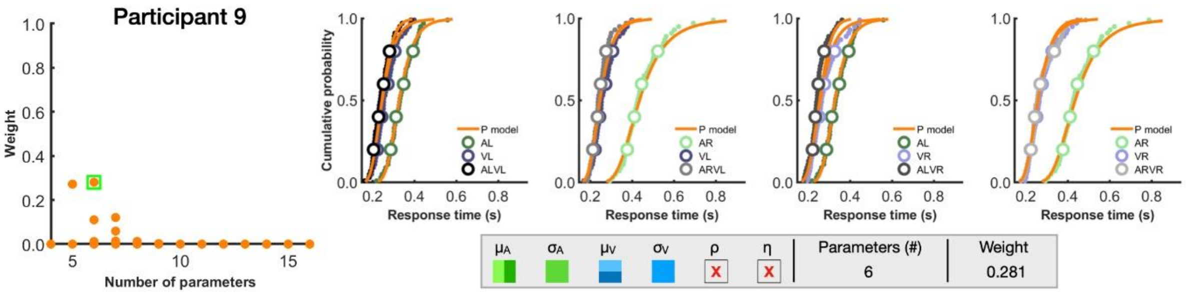
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	☒	🟩	8	0.168



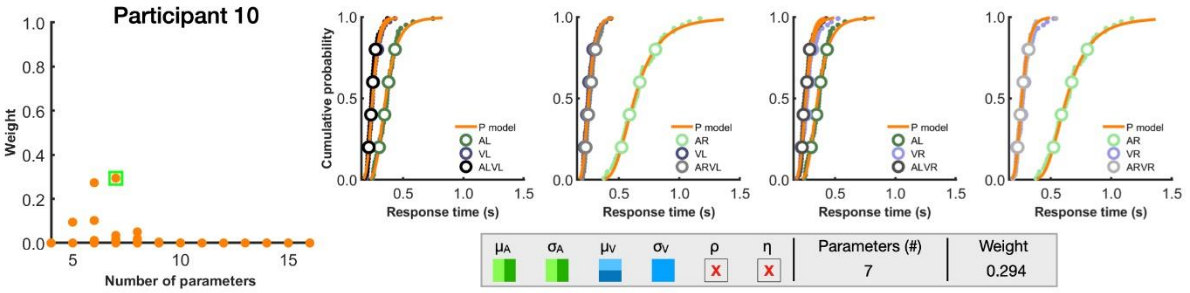
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	▨	🟩	9	0.612



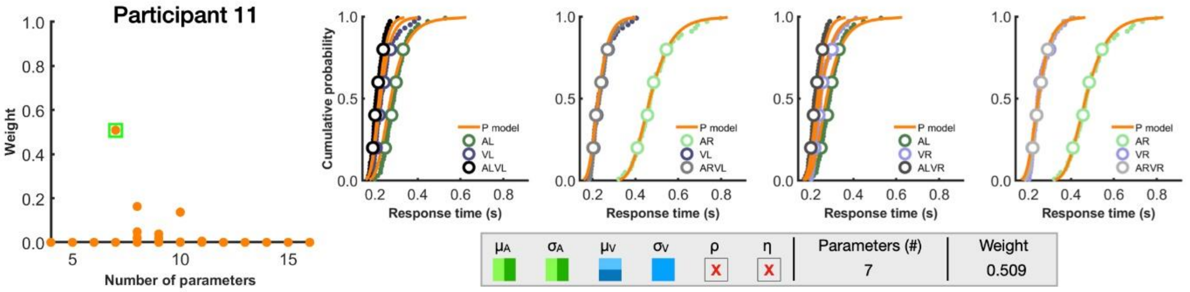
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩	🟩	🟦	🟦	☒	☒	6	0.435



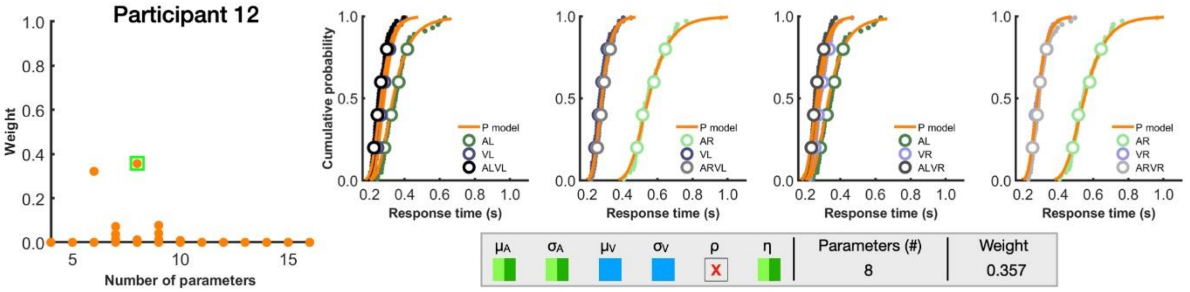
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩	🟦🟦	🟦	X	X	6	0.281



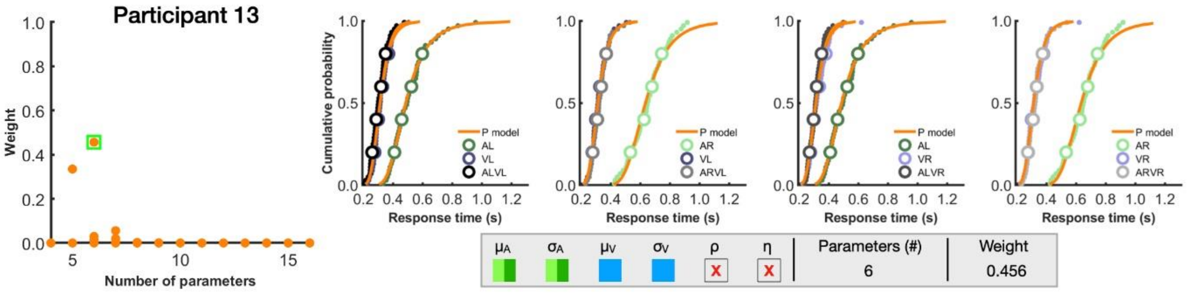
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩🟩	🟦🟦	🟦	X	X	7	0.294



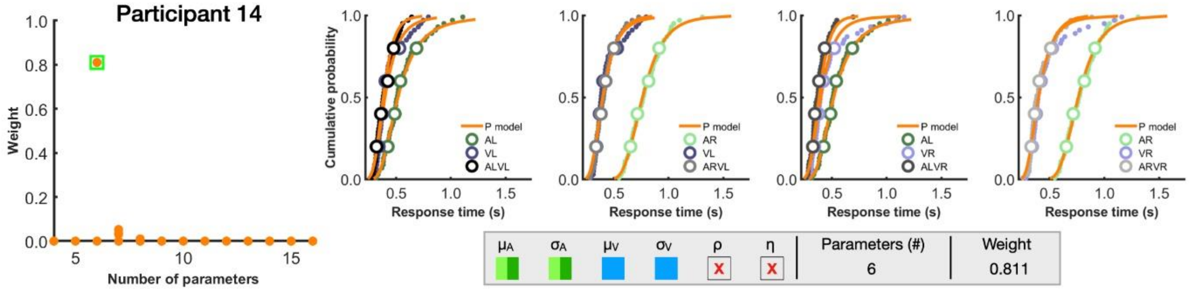
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩🟩	🟦🟦	🟦	X	X	7	0.509

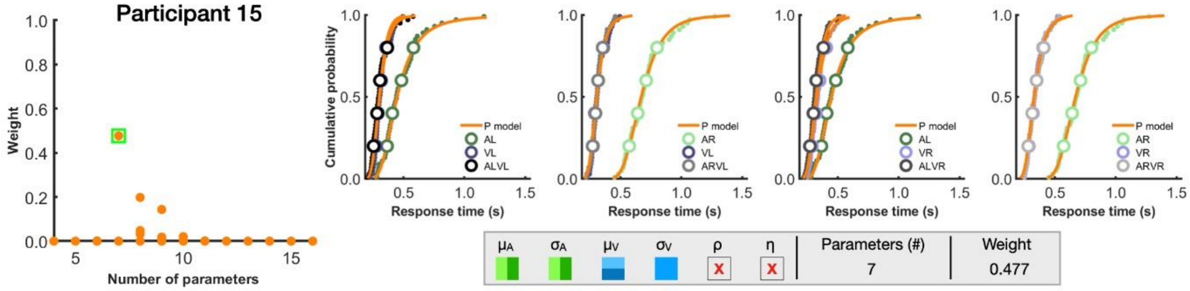


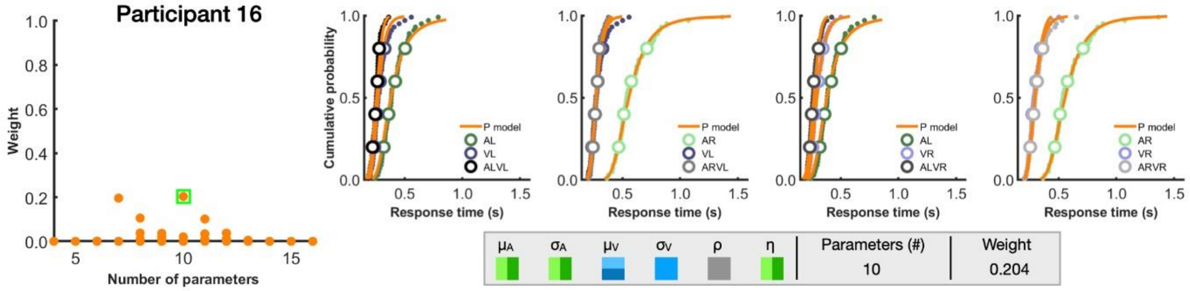
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩🟩	🟦	🟦	X	🟩🟩	8	0.357

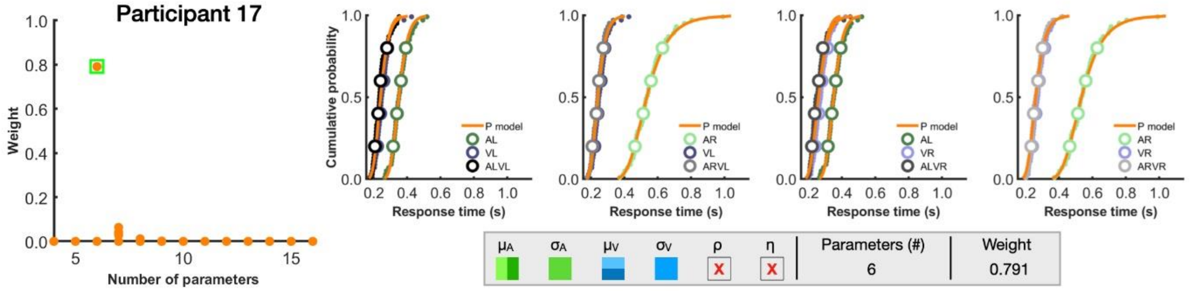


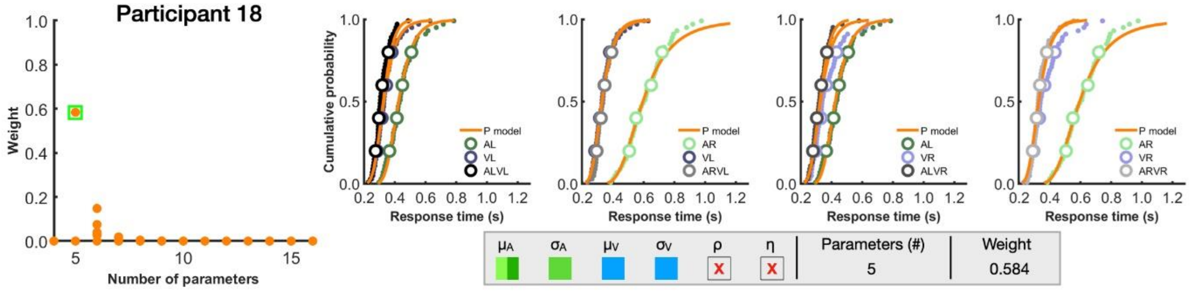
μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩🟩	🟦	🟦	X	X	6	0.456

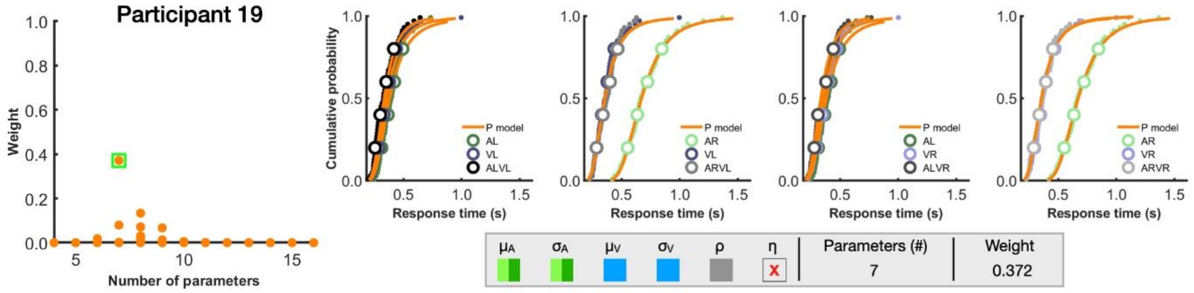

Participant 14
Weight
Number of parameters
Cumulative probability
Response time (s)
P model
AL
VL
ALVL
AR
ARVL
VR
ALVR
ARVR
μA
σA
μV
σV
ρ
η
X
X
Parameters (#)
6
Weight
0.811


Participant 15
Weight
Number of parameters
Cumulative probability
Response time (s)
P model
AL
VL
ALVL
AR
ARVL
VR
ALVR
ARVR
μA
σA
μV
σV
ρ
η
X
X
Parameters (#)
7
Weight
0.477

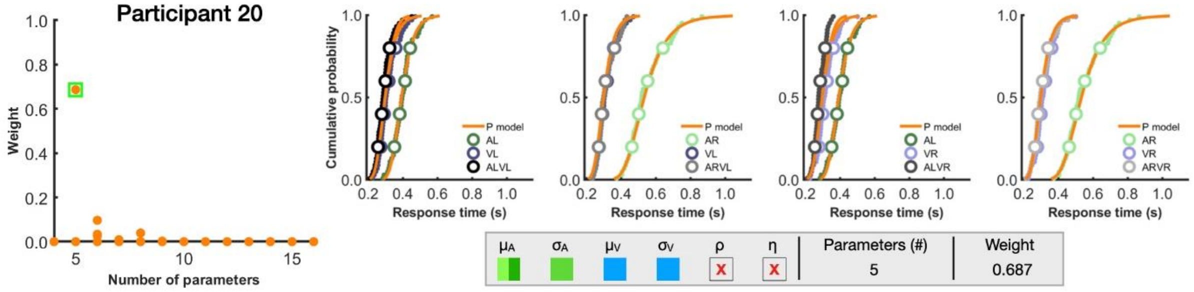

Participant 16
Weight
Number of parameters
Cumulative probability
Response time (s)
P model
AL
VL
ALVL
AR
ARVL
VR
ALVR
ARVR
μA
σA
μV
σV
ρ
η
Parameters (#)
10
Weight
0.204


Participant 17
Weight
Number of parameters
Cumulative probability
Response time (s)
P model
AL
VL
ALVL
AR
ARVL
VR
ALVR
ARVR
μA
σA
μV
σV
ρ
η
X
X
Parameters (#)
6
Weight
0.791


Participant 18
Weight
Number of parameters
Cumulative probability
Response time (s)
P model
AL
VL
ALVL
AR
ARVL
VR
ALVR
ARVR
μA
σA
μV
σV
ρ
η
X
X
Parameters (#)
5
Weight
0.584

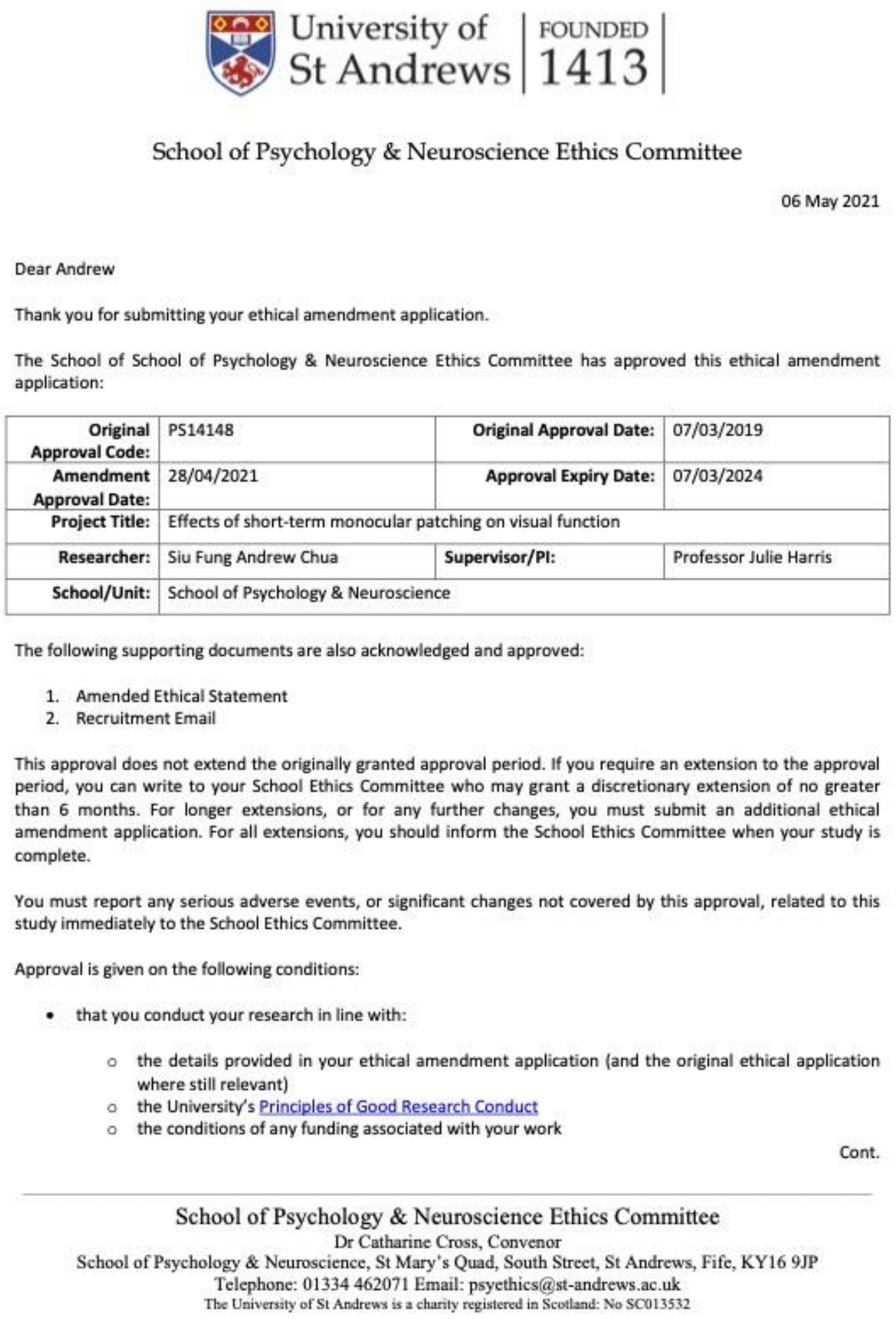


μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩🟩	🟦	🟦	■	X	7	0.372



μA	σA	μV	σV	ρ	η	Parameters (#)	Weight
🟩🟩	🟩	🟦	🟦	X	X	5	0.687

Appendix D – Short-term monocular patching project ethical approval



University of St Andrews | FOUNDED 1413

School of Psychology & Neuroscience Ethics Committee

06 May 2021

Dear Andrew

Thank you for submitting your ethical amendment application.

The School of School of Psychology & Neuroscience Ethics Committee has approved this ethical amendment application:

Original Approval Code:	PS14148	**Original Approval Date:**	07/03/2019
Amendment Approval Date:	28/04/2021	**Approval Expiry Date:**	07/03/2024
Project Title:	Effects of short-term monocular patching on visual function		
Researcher:	Siu Fung Andrew Chua	**Supervisor/PI:**	Professor Julie Harris
School/Unit:	School of Psychology & Neuroscience		

The following supporting documents are also acknowledged and approved:

1. Amended Ethical Statement
2. Recruitment Email

This approval does not extend the originally granted approval period. If you require an extension to the approval period, you can write to your School Ethics Committee who may grant a discretionary extension of no greater than 6 months. For longer extensions, or for any further changes, you must submit an additional ethical amendment application. For all extensions, you should inform the School Ethics Committee when your study is complete.

You must report any serious adverse events, or significant changes not covered by this approval, related to this study immediately to the School Ethics Committee.

Approval is given on the following conditions:

- that you conduct your research in line with:
    - the details provided in your ethical amendment application (and the original ethical application where still relevant)
    - the University's Principles of Good Research Conduct
    - the conditions of any funding associated with your work

Cont.

School of Psychology & Neuroscience Ethics Committee
Dr Catharine Cross, Convenor
School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife, KY16 9JP
Telephone: 01334 462071 Email: psyethics@st-andrews.ac.uk
The University of St Andrews is a charity registered in Scotland: No SC013532

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. doi:10.1109/TAC.1974.1100705

Alais, D., Newell, F., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing and Perceiving, 23*(1), 3-38. doi:10.1163/187847510X488603

Aleshin, S., Ziman, G., Kovács, I., & Braun, J. (2019). Perceptual reversals in binocular rivalry: improved detection from OKN. *Journal of Vision, 19*(3), 5-5. doi:10.1167/19.3.5

American Optometric Association. (n.d.). Visual acuity. Retrieved from https://www.aoa.org/healthy-eyes/vision-and-vision-correction/visual-acuity?sso=y

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*(2), 154-179. doi:10.1037/0033-295X.93.2.154

Bach, D. R., Neuhoff, J. G., Perrig, W., & Seifritz, E. (2009). Looming sounds as warning signals: the function of motion cues. *International Journal of Psychophysiology, 74*(1), 28-33. doi:10.1016/j.ijpsycho.2009.06.004

Bach, D. R., Schächinger, H., Neuhoff, J. G., Esposito, F., Salle, F. D., Lehmann, C., . . . Seifritz, E. (2008). Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cerebral Cortex, 18*(1), 145-150. doi:10.1093/cercor/bhm040

Baker, D. H., Kaestner, M., & Gouws, A. D. (2016). Measurement of crosstalk in stereoscopic display systems used for vision research. *Journal of Vision, 16*(15), 14-14. doi:10.1167/16.15.14

Ball, W., & Tronick, E. (1971). Infant responses to impending collision: optical and real. *Science, 171*(3973), 818. doi:10.1126/science.171.3973.818

Baumgartner, R., Reed, D. K., Tóth, B., Best, V., Majdak, P., Colburn, H. S., & Shinn-Cunningham, B. (2017). Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias. *Proceedings of the National Academy of Sciences, 114*(36), 9743. doi:10.1073/pnas.1703247114

Blake, R., & Boothroyd, K. (1985). The precedence of binocular fusion over binocular rivalry. *Perception & Psychophysics, 37*(2), 114-124. doi:10.3758/BF03202845

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience, 3*(1), 13-21. doi:10.1038/nrn701

Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences, 11*(3), 118-125. doi:10.1016/j.tics.2006.12.006

Bossi, M., Hamm, L. M., Dahlmann-Noor, A., & Dakin, S. C. (2018). A comparison of tests for quantifying sensory eye dominance. *Vision Research, 153*, 60-69. doi:10.1016/j.visres.2018.09.006

Bossi, M., Tailor, V. K., Anderson, E. J., Bex, P. J., Greenwood, J. A., Dahlmann-Noor, A., & Dakin, S. C. (2017). Binocular therapy for childhood amblyopia improves vision without breaking interocular suppression. *Investigative Ophthalmology & Visual Science, 58*(7), 3031-3043. doi:10.1167/iovs.16-20913

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436. doi:10.1163/156856897X00357

Bronkhorst, A. W., & Houtgast, T. (1999). Auditory distance perception in rooms. *Nature, 397*(6719), 517-520. doi:10.1038/17374

Burr, D., & Alais, D. (2006). Combining visual and auditory information. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso, & P. U. Tse (Eds.), *Progress in Brain Research* (Vol. 155, pp. 243-258): Elsevier.

Calcagno, E. R., Abregú, E. L., Eguía, M. C., & Vergara, R. (2012). The role of vision in auditory distance perception. *Perception, 41*(2), 175-192. doi:10.1068/p7153

Camponogara, I., Komeilipoor, N., & Cesari, P. (2015). When distance matters: perceptual bias and behavioral response for approaching sounds in peripersonal and extrapersonal space. *Neuroscience, 304*, 101-108. doi:10.1016/j.neuroscience.2015.07.054

Canzoneri, E., Magosso, E., & Serino, A. (2012). Dynamic sounds capture the boundaries of peripersonal space representation in humans. *PLOS ONE, 7*(9), e44306. doi:10.1371/journal.pone.0044306

Cappe, C., Thelen, A., Romei, V., Thut, G., & Murray, M. M. (2012). Looming signals reveal synergistic principles of multisensory integration. *The Journal of Neuroscience, 32*(4), 1171. doi:10.1523/JNEUROSCI.5517-11.2012

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia, 47*(4), 1045-1052. doi:10.1016/j.neuropsychologia.2008.11.003

Carmel, D., Arcaro, M., Kastner, S., & Hasson, U. (2010). How to create and use binocular rivalry. *Journal of Visualized Experiments*(45), 2030. doi:10.3791/2030

Carpenter, R. H. S., Reddi, B. A. J., & Anderson, A. J. (2009). A simple two-stage model predicts response time distributions. *The Journal of physiology, 587*(Pt 16), 4051-4062. doi:10.1113/jphysiol.2009.173955

Carpenter, R. H. S., & Williams, M. L. L. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature, 377*(6544), 59-62. doi:10.1038/377059a0

Chadnova, E., Reynaud, A., Clavagnier, S., & Hess, R. F. (2017). Latent binocular function in amblyopia. *Vision Research, 140*, 73-80. doi:10.1016/j.visres.2017.07.014

Chakrabarti, A., & Ghosh, J. K. (2011). AIC, BIC and recent advances in model selection. In P. S. Bandyopadhyay & M. R. Forster (Eds.), *Philosophy of Statistics* (Vol. 7, pp. 583-605). Amsterdam: North-Holland.

Chua, S. F. A., Liu, Y., Harris, J. M., & Otto, T. U. (2022). No selective integration required: a race model explains responses to audiovisual motion-in-depth. *Cognition, 227*, 105204. doi:10.1016/j.cognition.2022.105204

Cleveland Clinic. (2022). 20/20 Vision. Retrieved from https://my.clevelandclinic.org/health/articles/8561-2020-vision

Colonius, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology, 34*(3), 253-275. doi:10.1016/0022-2496(90)90032-5

Crawford, M. L. J., Harwerth, R. S., Smith, E. L., & von Noorden, G. K. (1996). Loss of stereopsis in monkeys following prismatic binocular dissociation during infancy. *Behavioural Brain Research, 79*(1), 207-218. doi:10.1016/0166-4328(95)00247-2

Daw, N. W. (1998). Critical periods and amblyopia. *Archives of Ophthalmology, 116*(4), 502-505. doi:10.1001/archopht.116.4.502

Dieter, K. C., Sy, J. L., & Blake, R. (2017). Individual differences in sensory eye dominance reflected in the dynamics of binocular rivalry. *Vision Research, 141*, 40-50. doi:10.1016/j.visres.2016.09.014

Ellemberg, D., Lewis, T. L., Maurer, D., & Brent, H. P. (2000). Influence of monocular deprivation during infancy on the later development of spatial and temporal vision. *Vision Research, 40*(23), 3283-3295. doi:10.1016/S0042-6989(00)00165-6

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162-169. doi:10.1016/j.tics.2004.02.002

Farrell, S., & Lewandowsky, S. (2015). An introduction to cognitive modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An Introduction to Model-Based Cognitive Neuroscience* (pp. 3-24). New York, NY: Springer New York.

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65*(7), 999-1010. doi:10.3758/BF03194829

Freiberg, K., Tually, K., & Crassini, B. (2001). Use of an auditory looming task to test infants' sensitivity to sound pressure level as an auditory distance cue. *British Journal of Developmental Psychology, 19*(1), 1-10. doi:10.1348/026151001165903

Gambacorta, C., Nahum, M., Vedamurthy, I., Bayliss, J., Jordan, J., Bavelier, D., & Levi, D. M. (2018). An action video game for the treatment of amblyopia in children: a feasibility study. *Vision Research, 148*, 1-14. doi:10.1016/j.visres.2018.04.005

Gau, R., Bazin, P.-L., Trampel, R., Turner, R., & Noppeney, U. (2020). Resolving multisensory and attentional influences across cortical depth in sensory cortices. *eLife, 9*, e46856. doi:10.7554/eLife.46856

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473-490. doi:10.1162/089892999563544

Gondan, M., Lange, K., Rösler, F., & Röder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin & Review, 11*(2), 307-313. doi:10.3758/BF03196575

Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception, & Psychophysics, 78*(3), 723-735. doi:10.3758/s13414-015-1018-y

Graziano, M. S., Yap, G. S., & Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science, 266*(5187), 1054. doi:10.1126/science.7973661

Graziano, M. S. A., Reiss, L. A. J., & Gross, C. G. (1999). A neuronal representation of the location of nearby sounds. *Nature, 397*(6718), 428-430. doi:10.1038/17115

Grice, G. R., Canham, L., & Gwynne, J. W. (1984). Absence of a redundant-signals effect in a reaction time task with divided attention. *Perception & Psychophysics, 36*(6), 565-570. doi:10.3758/BF03207517

Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology, 9*(22), 1315-1318. doi:10.1016/S0960-9822(00)80053-6

Heathcote, A., Brown, S., & Cousineau, D. (2004). QMPE: estimating Lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound. *Behavior Research Methods, Instruments, & Computers, 36*(2), 277-290. doi:10.3758/BF03195574

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review, 9*(2), 394-401. doi:10.3758/BF03196299

Heathcote, A., Brown, S. D., & Wagenmakers, E.-J. (2015). An Introduction to Good Practices in Cognitive Modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An Introduction to Model-Based Cognitive Neuroscience* (pp. 25-48). New York, NY: Springer New York.

Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology, 56*, 872-876. doi:10.1037/h0040546

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology, 63*(3), 289-293. doi:10.1037/h0039516

Hess, R. F., & Thompson, B. (2013). New insights into amblyopia: binocular therapy and noninvasive brain stimulation. *Journal of American Association for Pediatric Ophthalmology and Strabismus, 17*(1), 89-93. doi:10.1016/j.jaapos.2012.10.018

Hess, R. F., & Thompson, B. (2015). Amblyopia and the binocular approach to its therapy. *Vision Research, 114*, 4-16. doi:10.1016/j.visres.2015.02.009

Holmes, J. M., & Clarke, M. P. (2006). Amblyopia. *The Lancet, 367*(9519), 1343-1351. doi:10.1016/S0140-6736(06)68581-4

Huygelier, H., van Ee, R., Lanssens, A., Wagemans, J., & Gillebert, C. R. (2021). Audiovisual looming signals are not always prioritised: evidence from exogenous, endogenous and sustained attention. *Journal of Cognitive Psychology, 33*(3), 282-303. doi:10.1080/20445911.2021.1896528

Innes, B. R., & Otto, T. U. (2019). A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent. *Scientific Reports, 9*(1), 2921. doi:10.1038/s41598-019-39924-6

James, K. H. (2010). Sensori-motor experience leads to changes in visual processing in the developing brain. *Dev Sci, 13*(2), 279-288. doi:10.1111/j.1467-7687.2009.00883.x

James, K. H., Humphrey, G. K., & Goodale, M. A. (2001). Manipulating and recognizing virtual objects: where the action is. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 55*, 111-120. doi:10.1037/h0087358

Kim, H.-W., Kim, C.-Y., & Blake, R. (2017). Monocular perceptual deprivation from interocular suppression temporarily imbalances ocular dominance. *Current Biology, 27*(6), 884-889. doi:10.1016/j.cub.2017.01.063

Kinchla, R. A. (1974). Detecting target elements in multielement arrays: a confusability model. *Perception & Psychophysics, 15*(1), 149-158. doi:10.3758/BF03205843

Kingdom, F. A. A., & Prins, N. (2016a). Adaptive methods. In F. A. A. Kingdom & N. Prins (Eds.), *Psychophysics (Second Edition)* (pp. 119-148). San Diego: Academic Press.

Kingdom, F. A. A., & Prins, N. (2016b). Psychometric Functions. In F. A. A. Kingdom & N. Prins (Eds.), *Psychophysics (Second Edition)* (pp. 55-117). San Diego: Academic Press.

Kleiner, M., Brainard, D., & Pelli, D. (2007). *What's new in Psychtoolbox-3?* Paper presented at the ECVP.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience, 16*(8), 1412-1425. doi:10.1162/0898929042304796

Lack, L. C. (1974). Selective attention and the control of binocular rivalry. *Perception & Psychophysics, 15*(1), 193-200. doi:10.3758/BF03205846

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: principles and practice*. In. Retrieved from https://sk.sagepub.com/books/computational-modeling-in-cognition doi:10.4135/9781483349428

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764-766. doi:10.1016/j.jesp.2013.03.013

Li, R. W., Ngo, C., Nguyen, J., & Levi, D. M. (2011). Video-game play induces plasticity in the visual system of adults with amblyopia. *PLOS Biology, 9*(8), e1001135. doi:10.1371/journal.pbio.1001135

Li, X., Liang, Z., Kleiner, M., & Lu, Z.-L. (2010). RTbox: a device for highly accurate response time measurements. *Behavior Research Methods, 42*(1), 212-225. doi:10.3758/BRM.42.1.212

Lin, J. Y., Franconeri, S., & Enns, J. T. (2008). Objects on a collision path with the observer demand attention. *Psychological Science, 19*(7), 686-692. doi:10.1111/j.1467-9280.2008.02143.x

Liu, X.-Y., & Zhang, J.-Y. (2018). Dichoptic training in adults with amblyopia: additional stereoacuity gains over monocular training. *Vision Research, 152*, 84-90. doi:10.1016/j.visres.2017.07.002

Liu, Y., & Otto, T. U. (2020). The role of context in experiments and models of multisensory decision making. *Journal of Mathematical Psychology, 96*, 102352. doi:10.1016/j.jmp.2020.102352

Luce, R. D. (1986). *Response times: their role in inferring elementary mental organization*. New York: Oxford University Press.

Lunghi, C., Berchicci, M., Morrone, M. C., & Di Russo, F. (2015a). Short-term monocular deprivation alters early components of visual evoked potentials. *The Journal of physiology, 593*(19), 4361-4372. doi:10.1113/JP270950

Lunghi, C., Burr, D. C., & Morrone, C. (2011). Brief periods of monocular deprivation disrupt ocular balance in human adult visual cortex. *Current Biology, 21*(14), R538-R539. doi:10.1016/j.cub.2011.06.004

Lunghi, C., Burr, D. C., & Morrone, M. C. (2013). Long-term effects of monocular deprivation revealed with binocular rivalry gratings modulated in luminance and in color. *Journal of Vision, 13*(6), 1-1. doi:10.1167/13.6.1

Lunghi, C., Emir, Uzay E., Morrone, Maria C., & Bridge, H. (2015b). Short-term monocular deprivation alters GABA in the adult human visual cortex. *Current Biology, 25*(11), 1496-1501. doi:10.1016/j.cub.2015.04.021

Lunghi, C., & Sale, A. (2015). A cycling lane for brain rewiring. *Current Biology, 25*(23), R1122-R1123. doi:10.1016/j.cub.2015.10.026

Maconachie, G. D. E., & Gottlob, I. (2015). The challenges of amblyopia treatment. *Biomedical Journal, 38*(6), 510-516. doi:10.1016/j.bj.2015.06.001

Maier, J. X., & Ghazanfar, A. A. (2007). Looming biases in monkey auditory cortex. *The Journal of Neuroscience, 27*(15), 4093. doi:10.1523/JNEUROSCI.0330-07.2007

Mercier, M. R., & Cappe, C. (2020). The interplay between multisensory integration and perceptual decision making. *NeuroImage, 222*, 116970. doi:10.1016/j.neuroimage.2020.116970

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology, 14*(2), 247-279. doi:10.1016/0010-0285(82)90010-X

Mitchell, D. E., & Gingras, G. (1998). Visual recovery after monocular deprivation is driven by absolute, rather than relative, visually evoked activity levels. *Current Biology, 8*(21), 1179-1182. doi:10.1016/S0960-9822(07)00489-7

Moher, J., Sit, J., & Song, J.-H. (2015). Goal-directed action is automatically biased towards looming motion. *Vision Research, 113*, 188-197. doi:10.1016/j.visres.2014.08.005

Morishita, H., & Hensch, T. K. (2008). Critical period revisited: impact on vision. *Current Opinion in Neurobiology, 18*(1), 101-107. doi:10.1016/j.conb.2008.05.009

Nadarajah, S., & Kotz, S. (2008). Estimation methods for the multivariate t distribution. *Acta Applicandae Mathematicae, 102*(1), 99-118. doi:10.1007/s10440-008-9212-8

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature, 395*(6698), 123-124. doi:10.1038/25862

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology, 13*(2), 87-110. doi:10.1207/S15326969ECO1302_2

Noorani, I., & Carpenter, R. H. S. (2011). Full reaction time distributions reveal the complexity of neural decision-making. *European Journal of Neuroscience, 33*(11), 1948-1951. doi:10.1111/j.1460-9568.2011.07727.x

Noorani, I., & Carpenter, R. H. S. (2016). The LATER model of reaction time and decision. *Neuroscience & Biobehavioral Reviews, 64*, 229-251. doi:10.1016/j.neubiorev.2016.02.018

Orioli, G., Bremner, A. J., & Farroni, T. (2018). Multisensory perception of looming and receding objects in human newborns. *Current Biology, 28*(22), R1294-R1295. doi:10.1016/j.cub.2018.10.004

Otto, T. U. (2019). RSE-box: an analysis and modelling package to study response times to multiple signals. *The Quantitative Methods for Psychology, 15*(2), 112-133. doi:10.20982/tqmp.15.2.p112

Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *The Journal of Neuroscience, 33*(17), 7463. doi:10.1523/JNEUROSCI.4678-12.2013

Otto, Thomas U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology, 22*(15), 1391-1396. doi:10.1016/j.cub.2012.05.031

Otto, T. U., & Mamassian, P. (2017). Multisensory decisions: the test of a race model, its logic, and power. *Multisensory Research, 30*(1), 1-24. doi:10.1163/22134808-00002541

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of experimental psychology. Human perception and performance, 37*(1), 58-71. doi:10.1037/a0020747

Paquier, M., Côté, N., Devillers, F., & Koehl, V. (2016). Interaction between auditory and visual perceptions on distance estimations in a virtual environment. *Applied Acoustics, 105*, 186-199. doi:10.1016/j.apacoust.2015.12.014

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*(4), 437-442. doi:10.1163/156856897X00366

Prins, N., & Kingdom, F. A. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Frontiers in Psychology, 9*. doi:10.3389/fpsyg.2018.01250

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences, 24*(5), 574-590. doi:10.1111/j.2164-0947.1962.tb01433.x

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1226-1243. doi:10.1037/a0036801

Raftery, A. E. (1999). Bayes factors and BIC: comment on “a critique of the Bayesian information criterion for model selection”. *Sociological Methods & Research, 27*(3), 411-427. doi:10.1177/0049124199027003005

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychol Bull, 86*(3), 446-461. doi:10.1037/0033-2909.86.3.446

Rosenblum, L. D., Carello, C., & Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception, 16*(2), 175-186. doi:10.1068/p160175

Rosenblum, L. D., Wuestefeld, A. P., & Saldaña, H. M. (1993). Auditory looming perception: influences on anticipatory judgments. *Perception, 22*(12), 1467-1482. doi:10.1068/p221467

Schiff, W., Caviness, J. A., & Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of "looming". *Science, 136*(3520), 982-983. doi:10.1126/science.136.3520.982

Schiff, W., & Oldak, R. (1990). Accuracy of judging time to arrival: effects of modality, trajectory, and gender. *Journal of Experimental Psychology: Human Perception and Performance, 16*(2), 303-316. doi:10.1037/0096-1523.16.2.303

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, K., Mustovic, H., Schächinger, H., . . . Di Salle, F. (2002). Neural processing of auditory looming in the human brain. *Current Biology, 12*(24), 2147-2151. doi:10.1016/S0960-9822(02)01356-8

Serino, A., Annella, L., & Avenanti, A. (2009). Motor properties of peripersonal space in humans. *PLOS ONE, 4*(8), e6582. doi:10.1371/journal.pone.0006582

Shaw, L. H., Freedman, E. G., Crosse, M. J., Nicholas, E., Chen, A. M., Braiman, M. S., . . . Foxe, J. J. (2020). Operating in a multisensory context: assessing the interplay between multisensory reaction time facilitation and inter-sensory task-switching effects. *Neuroscience, 436*, 122-135. doi:10.1016/j.neuroscience.2020.04.013

The British Psychological Society. (2014). *BPS code of human research ethics*: The British Psychological Society.

Todd, J. W. (1912). *Reaction to multiple stimuli*: The Science Press.

Townsend, J. T., Liu, Y., Zhang, R., & Wenger, M. J. (2020). Interactive parallel models: no Virginia, violation of Miller's race inequality does not imply coactivation and yes Virginia, context invariance is testable. *The Quantitative Methods for Psychology, 16*(2), 192-212. doi:10.20982/tqmp.16.2.p192

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: new capacity measures and predictions for a response time inequality series. *Psychological Review, 111*(4), 1003-1035. doi:10.1037/0033-295X.111.4.1003

Tyll, S., Bonath, B., Schoenfeld, M. A., Heinze, H.-J., Ohl, F. W., & Noesselt, T. (2013). Neural basis of multisensory looming signals. *NeuroImage, 65,* 13-22. doi:10.1016/j.neuroimage.2012.09.056

Vedamurthy, I., Nahum, M., Huang, S. J., Zheng, F., Bayliss, J., Bavelier, D., & Levi, D. M. (2015). A dichoptic custom-made action video game as a treatment for adult amblyopia. *Vision Research, 114,* 173-187. doi:10.1016/j.visres.2015.04.008

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192-196. doi:10.3758/BF03206482

Weakliem, D. L. (1999). A critique of the Bayesian information criterion for model selection. *Sociological Methods & Research, 27*(3), 359-397. doi:10.1177/0049124199027003002

Webber, A. L., & Wood, J. (2005). Amblyopia: prevalence, natural history, functional effects and treatment. *Clinical and Experimental Optometry, 88*(6), 365-375. doi:10.1111/j.1444-0938.2005.tb05102.x

Wiesel, T. N., & Hubel, D. H. (1965). Extent of recovery from the effects of visual deprivation in kittens. *J Neurophysiol, 28*(6), 1060-1072. doi:10.1152/jn.1965.28.6.1060

Woods, A. J. (2012). Crosstalk in stereoscopic displays: a review. *Journal of Electronic Imaging, 21*(4), 1-22. doi:10.1117/1.JEI.21.4.040902

Yang, C.-T., Altieri, N., & Little, D. R. (2018). An examination of parallel versus coactive processing accounts of redundant-target audiovisual signal processing. *Journal of Mathematical Psychology, 82,* 138-158. doi:10.1016/j.jmp.2017.09.003

Zahorik, P., Brungart, D., & Bronkhorst, A. (2005). Auditory distance perception in humans: a summary of past and present research. *Acta Acustica united with Acustica, 91,* 409-420.

Zhang, J.-Y., Cong, L.-J., Klein, S. A., Levi, D. M., & Yu, C. (2014). Perceptual learning improves adult amblyopic vision through rule-based cognitive compensation. *Investigative Ophthalmology & Visual Science, 55*(4), 2020-2030. doi:10.1167/iovs.13-13739

Zhou, J., Baker, D. H., Simard, M., Saint-Amour, D., & Hess, R. F. (2015). Short-term monocular patching boosts the patched eye's response in visual cortex. *Restorative Neurology and Neuroscience*, *33*, 381-387. doi:10.3233/RNN-140472