FETPROACT-2-2014: Knowing, doing, being: cognition beyond problem solving

ACTION ACRONYM
TIMESTORM
ACTION FULL TITLE
"MIND AND TIME: INVESTIGATION OF THE TEMPORAL TRAITS OF HUMAN-MACHINE CONVERGENCE"
GRANT AGREEMENT NO:
641100

DELIVERABLE D5.1

Intermediate results on the temporal aspects of social self.

DUE DATE
MONTH 12

RESPONSIBLE PARTNER
UNIVERSITY OF SUSSEX

Contents

Intermediate results on the temporal aspects of social self.

1 Summary of D5.1

The first deliverable (D5.1) of work package 5 addresses the temporal aspects of the social self. The work package aims to uncover how the brain forms representations of time within and across both the self and in different contexts. To achieve this goal, our initial progress has focused on several key research directions. TIMESTORM members at University of Sussex have used innovative "substitutional reality" methods to explore how temporal and interoceptive perception interact with one another in naturalistic sensory settings, while members at Université Blaise Pascal have begun investigation of the effect of emotional context on humans' abilities to synchronize with others. Finally, TIMESTORM members at University of Sussex and Imperial College London have begun empirically investigating components of a putative predictive coding model of temporal perception.

Temporal and interoceptive perception in substitutional reality

Substitutional reality (SR) setups offer an exciting way of investigating interoceptive and temporal perceptions in naturalistic sensory environments. Using such a system (3.1), we report that coherent sensorimotor coupling (the coupling of an observer's head movements with the contents of their visual scene) has an impact on temporal judgments such that temporal intervals experienced in the presence of natural sensorimotor coupling are judged to be longer than those experienced in the absence of such coupling. Further, consistent with previous work not using SR, we find that the rate of presentation of a scene (framerate) modulates reported estimates of duration such that scenes presented at a faster rate are reported as persisting for longer than those presented at normal or slower speeds. We also find that individuals exhibiting low interoceptive awareness generally overestimate the duration of an interval, however this result needs further investigation to exclude alternative interpretations. This work provides a platform for further planned empirical work that explores the interactions between interoceptive and temporal processes, necessary to inform a model of fully embodied temporal cognition.

Empathic aspects of social time perception

A shared perception of time between multiple agents is crucial for fluent collaboration in social situations. Previously, there had been no systematic investigation of the influence of emotional context on humans' abilities to synchronize with one another. Here we report that the emotional content of a visual stimulus can affect the ability of human observers to maintain a rhythmic motor action. Specifically, when presented with images of angry or fearful human faces, observers' speed of rhythmic tapping increased, relative to when presented with other emotional faces (neutral, disgust sadness, joy). These results may indicate that the emotional content of others affects our internal motor rhythm, thus modifying timed behaviours such as rhythmic tapping. This work was carried out by TIMESTORM members at Université Blaise Pascal.

Towards a 'predictive coding' model of temporal and interoceptive perception

Bayesian models have previously been developed to describe how prior knowledge affects sensory perception, providing a minimal basis for predictive processing models of perception. In collaboration between TIMESTORM members at University of Sussex and Imperial College London, we aim to move towards a mechanistic model of predictive temporal perception, based on probabilistic population coding. This work will build upon work from Imperial to be contained in Deliverable 2.3. To this end, we have begun empirical work to uncover the basic required components of such a model. Consistent with previous Bayesian models, we report that recent sensory experience biases perception such that estimates of duration are shifted towards the mean of a range of intervals (5.1). Interestingly, we find evidence for a novel effect such that these biases can be formed distinctly for signals originating in different sensory modalities (vision and audition) - vision being biased toward short duration estimates and audition towards long, for example. These results provide key insights into the potential architecture of predictive processing models that will be developed jointly by TIMESTORM members at University of Sussex and Imperial College London.

A further investigation (5.2) used a classic electrophysiological effect, the mismatch negativity effect (MMN), to explore the possible neural correlates of prediction in the temporal context. Initial results support the idea that we can identify neural processes dedicated to the prediction of durations of sensory events, and further, that the identified processes might be distinguishable from previous demonstrations of simple rhythmic prediction - indicating processes explicitly related to duration prediction. Further investigations are required to determine the relationship between such neural correlates, the behavioural results discussed in 5.1, and the under-development population code models of predictive processing.

2 Introduction to deliverable D5.1

Deliverable D5.1 is the first deliverable dealing with the content of work package 5 which addresses temporal aspects of the social self. In particular, the tasks in work package 5 attempt to elucidate the development and maintenance of sensory processes that underlie and maintain sense of self both across time, and within short temporal contexts. In this document we outline our initial progress towards the goals outlined in this work package. This progress is broadly classified into several directions: initial development and testing of substitutional reality (SR) setups for the purpose of examining the interaction of human temporal and interoceptive perceptions in naturalistic sensory environments, initial investigations of the empathic aspects of social time perception, and empirical investigations of the basic components required for operation of a 'predictive coding' model of temporal perception generally, to be extended to the specific case of interoceptive perception. The intermediate products of these research directions are outlined separately below.

3 Temporal and interoceptive perception in substitutional reality

The recent surge in the quality and availability of virtual reality equipment provides us the opportunity to investigate human perception in more fully realized and manipulable sensory environments than ever before. Recent studies have demonstrated the utility of augmented, substitutional or virtual reality setups for the investigation of sense of self and presence (e.g. Suzuki et al., 2012; Suzuki et al., 2013). Of particular interest here is the influence of sensorimotor coupling in perception. In our daily lives we take for granted that the position and movements of our body, such as turning our head or shifting fixation of our eyes, are coupled with exteroceptive inputs such as vision. The coherent integration of information about bodily position and action with exteroceptive information appears to be critical to our sense of self and successful interaction with our environment as indicated by cases where such signals are in conflict (e.g. when visual and vestibular signals are in conflict it often results in severe nausea - motion sickness). A deep understanding of how interoceptive (internal bodily signals), proprioceptive (bodily position signals), and exteroceptive (e.g. vision, hearing, and touch) signals interact in a temporal context is necessary to inform a model of fully embodied temporal cognition. The work described in section 3.1 and 3.2 was carried out by Keisuke Suzuki, Warrick Roseboom and Anil Seth at the University of Sussex.

3.1 Time perception in substitutional reality: the influence of sensorimotor coupling

Using our substitutional reality setup we manipulated the coupling between visual and bodily signals such that temporal perception could be probed during a specific visual experience while observers experienced either coherent or incoherent visual and bodily movements (i.e. the visual world moved in accord with, or independently to, the observer's own head movements). We also investigated the influence of the presentation rate of the visual scene (video frame rate) on temporal judgements as the results of a recent virtual reality study suggest that perception of temporal intervals is strongly dependent on the speed at which visual information is presented (van Rijn, 2014).

3.1.1 Experiment 1

Methods

We pre-recorded a video of a market scene using an omni-directional (360^{o}) camera (see Figure 1.). This video was then presented using a head-mounted Oculus display with directional head tracking allowing for exploration of the pre-recorded environment using head movements that are nearly indistinguishable from real life. This method generates strong impressions of immersion and subjective reality (Suzuki et al., 2012).

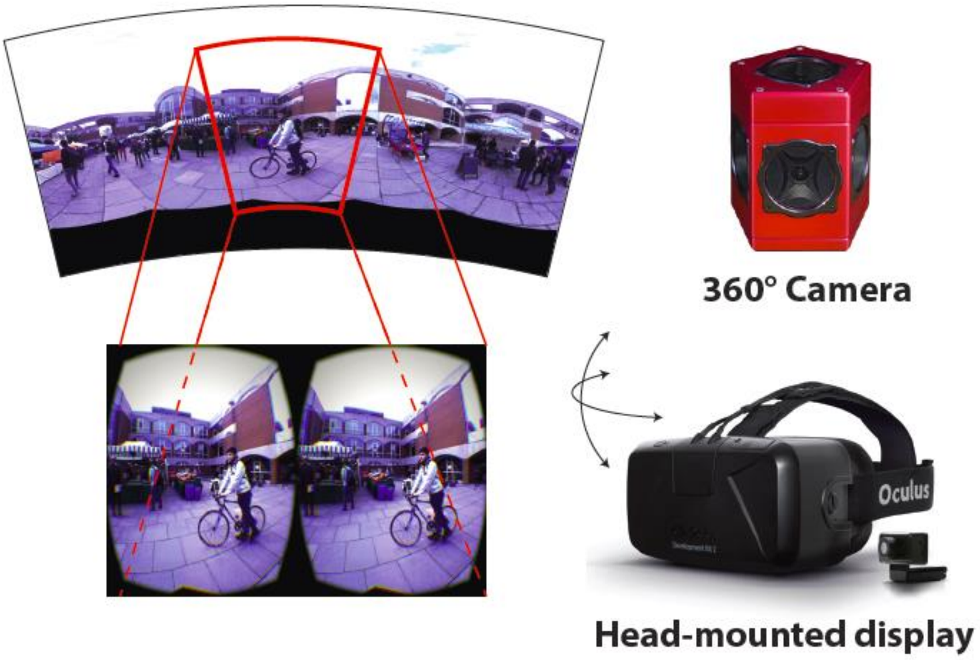

Figure 1. Market scene as viewed through the head-mounted Oculus display, shown bottom right beneath the omni-directional camera.

Observers were presented with a continuous scene for 5 minutes at a time. During this time they were required to actively search for the appearance of a small mosaic patch in the scene (see Figure 2). The location of the patch was randomized to be different on each trial. Upon detecting the patch, observers were required to press a key. The mosaic patch would then persist for a test interval of ~1500 ms (a normal distribution centered on 1500 ms with a range of +/- 1000 ms) at which point observers again needed to press the key, before finally pressing the key a third time when the same period of time between the first two keypresses had again elapsed - a temporal reproduction task similar to the classic ready-set-go task (e.g. Jazayeri & Shadlen, 2010; see Figure 2).

24 subjects naïve as to experimental purpose participated in two 1.5 hour experimental sessions during which they completed four different experimental blocks: Sensorimotor input coupled (Live) or uncoupled (Replay), and normal or fast (150%) speed playback. (57.6 trials, standard deviation = 5.2, on average for each condition). In the Live condition, observers were required to use their coherent sensorimotor coupling to search for the appearance of the target. In the Replay condition, observers also searched for the target, but the movement of the visual input was based on the replay of the head movements from a previous observer. Movements of their own head did not contribute to which part of the visual scene was visible. During the experiments we also measured observers' heart rates and speed of head movement as indicators of their bodily states.

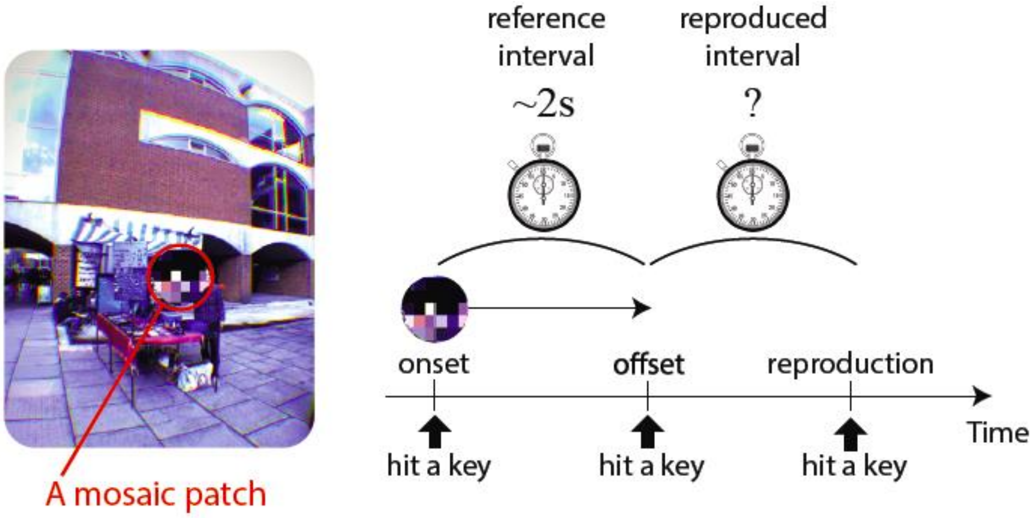

Figure 2. Example mosaic stimulus and task from Experiment 1.

Results and Discussion

Only 14 of the 24 observers' data contributed to the final analysis. Five subjects were unable to complete the experiment due to motion sickness (common in virtual reality), while a further five subjects were excluded as their reproduced intervals did not show a proportional relationship to the physically presented interval.

As shown in Figure 3, observers reproduced intervals as significantly longer in the Live (normal speed = 52 ms, fast = -1 ms) condition than in the Replay (normal speed = -11 ms, fast = -42 ms) condition ($F(1,14) = 12.05$, $p < 0.01$). At least for the normal speed Live condition, the interval appears to be reproduced as longer than veridical. Intervals were also reproduced as longer in the normal speed presentation than the fast ($F(1,14) = 8.00$, $p < 0.05$), though this contrast was not significant for a simple contrast of only the Replay data. We found no significant difference in heart rates between the different conditions.

These results clearly suggest that coherent sensorimotor coupling has a significant effect on temporal judgements. However, why subjective perception of a temporal interval should be *longer* when sensorimotor coupling is intact is not immediately clear. One possible explanation for the effect in this case may be an artifact introduced by the testing method. In the Live condition, observers actively moved their head in order to search for the target. Having found the target, observers would then generally stop head movement in order to complete the reproduction task. Having completed the task, the observer would then begin to move their head in search of the appearance of the next target. However, in the Replay condition observers passively viewed the replayed movement until completing the task. If, for the given trial, the Replayed observer reproduced the interval as short, there would be an onset of visual motion as that observer proceeded to search for the next target, possibly biasing the observer in the current Replay condition to complete their reproduction early. However, in the opposite case, when the observer in the Replay condition reproduced the interval as shorter than the previous observer whose replay they were viewing, the onset of visual motion in the replay would not affect reproduction as the observer would have already completed their reproduction. Over many trials, this asymmetry may have produced our result.

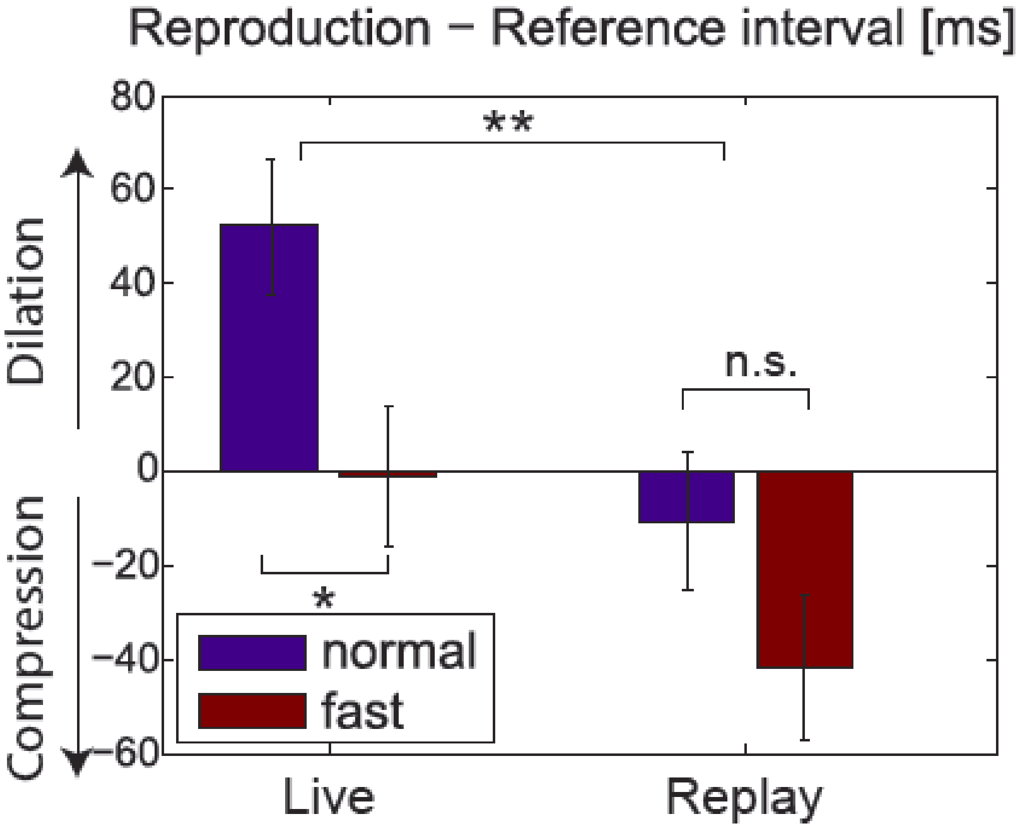

Figure 3. Results of mean reproduced interval in Experiment 1 by condition. Reproduction-Reference of 0 ms indicates that the interval was reproduced veridically, while negative values indicate that the interval was reproduced as shorter and positive values as longer than the physical presentation. Error bars indicate +/- standard error of the mean.

3.1.2 Experiment 2

To address the possible concerns with Experiment 1 outlined above, in Experiment 2 we used a different task. Rather than physically reproducing the interval, we had observers complete a magnitude estimation task. This task required that observers report the actual value of the presentation that they experienced (e.g. observers report the interval was 1500 ms long). The video was the same omni-directional market scene as used in Experiment 1 and again the task was to search for the appearance of a mosaic patch and press the button upon finding it. However, there were two critical differences from Experiment 1. Upon pressing the button to indicate that the patch had been found, the patch would change to a red hue. The patch would then persist for 1500, 1750, or 2000 ms (pseudo-randomly selected on a trial by trial basis according to a method of constant stimuli) before disappearing. A horizontal scale would then appear on the screen with a minimum of 0 ms on the left and a maximum of 5 s on the right side of the scale. Observers used the mouse to indicate on this scale the duration for which the previously presented interval seemed to persist.

As in Experiment 1, there were two factors that we manipulated: sensorimotor coupling and speed. However, as motion sickness was a serious problem in Experiment 1 (5 observers could not complete the task), in Experiment 2 the Replay condition was presented on a standard computer monitor rather than through the Oculus. We also expanded the speed condition to include a slow level. Consequently, there were two experimental factors, sensorimotor coupling with two levels (Live and Replay) and presentation speed with three: slow, normal, and fast (50%, 100%, 150% speed). The number of trials for each condition across subjects is 34.65 (standard deviation = 2.36).

Prior to participating in the main experiment, observers completed a heartbeat counting task (Schandry, 1981, Garfinkel et al., 2015). In which they were required to sit still and passively attend to the rhythm of their own heart beat for given durations (20s, 30s, and 40s, twice each). At the conclusion of each monitoring period, observers had to report how many heart beats had occurred during this period. The physical heart rate was monitored and interoceptive accuracy was estimated as the correspondence between physical and reported heart beats.

Results and Discussion

24 observers naïve as to experimental purpose participated in the experiment. Three observers were excluded from the analysis because their reports did not show a proportional relation to the physically presented intervals.

As indicated by Figure 4, and replicating Experiment 1, there was a main effect of presentation condition such that reported durations in the Live (intervals overestimated by 28 ms relative to veridical) condition were significantly longer than those in the Replay (intervals overestimated by 14 ms relative to veridical) condition ($F(1,20) = 7.24$, $p<0.01$). There was also a main effect of the speed of scene such that reported durations in scenes in which the presentation was slow were shorter than normal speed and fast scenes (intervals overestimated by 9, 19 and 36 ms relative to veridical, respectively, $F(2,20) = 4.73$, $p<0.01$).

These results confirm those from Experiment 1 in showing that intervals experienced in the presence of appropriate sensorimotor coupling are generally overestimated relative to those experienced passively where visual motion is not related to self-motion. Moreover, the significant effect of speed of scene is consistent with previous results (van Rijn, 2014), though here we were able to find the effect in the absence of vective motion.

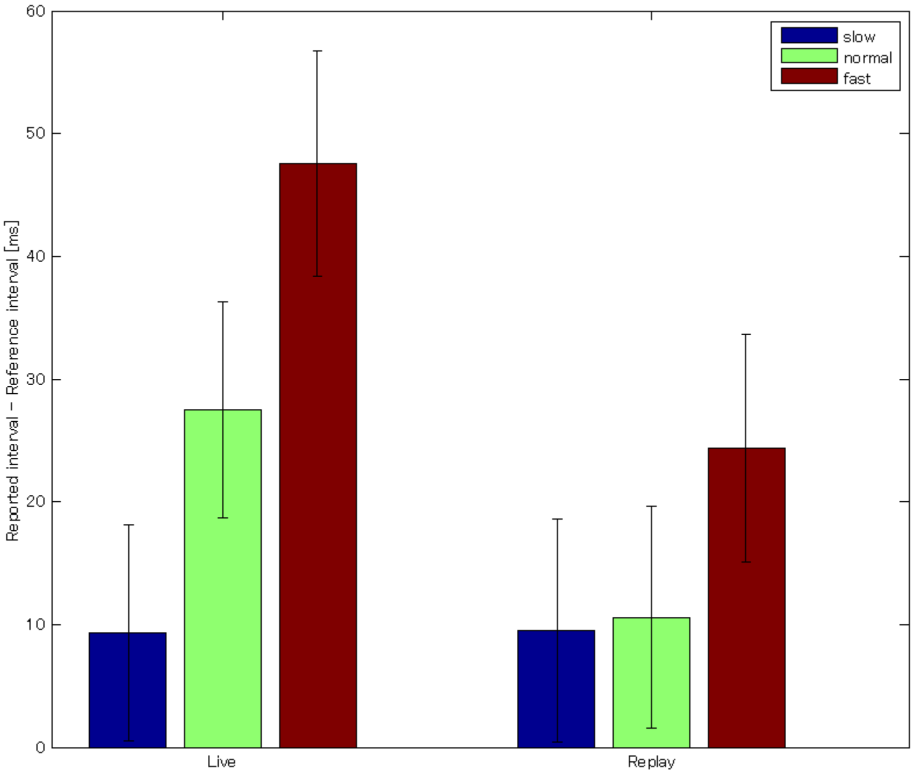

Figure 4. Results of mean reported interval in Experiment 2 by condition. Reported-Reference of 0 ms indicates that the interval was reported veridically, while negative values indicate that the interval was reported as shorter and positive values as longer than the physical presentation. Error bars indicate +/- standard error of the mean.

To investigate the possible interaction of interoception and temporal perception we compared the results after splitting the data into high and low interoceptive accuracy observers, based on median performance in the heart beat counting task. This is a common convention in dealing with interoceptive accuracy (e.g. Garfinkel et al., 2015). As shown in Figure 5, for observers who demonstrated low interoceptive accuracy, reported durations were overestimated relative to the presented duration by more than observers with high interoceptive accuracy (Low interoceptive accuracy: 96 ms for Live, 93 ms; for Replay, High interoceptive accuracy: 11 ms for Live, -4 ms for Replay; $F(1,10) = 142.39$, $p < 0.01$). This result may indicate that interoceptive and temporal accuracy are related. However, as the task used in this study to indicate interoceptive accuracy was a fundamentally temporal task (counting or otherwise estimating the magnitude of an interval containing a series of rhythmic stimuli, i.e. the heart beat), it is possible that this effect only indicates that individuals with higher temporal precision tend to be more accurate for at least these two different temporal tasks – an unsurprising result. To distinguish between these possibilities, in future studies we will also examine interoceptive accuracy or awareness using non-temporal tasks.

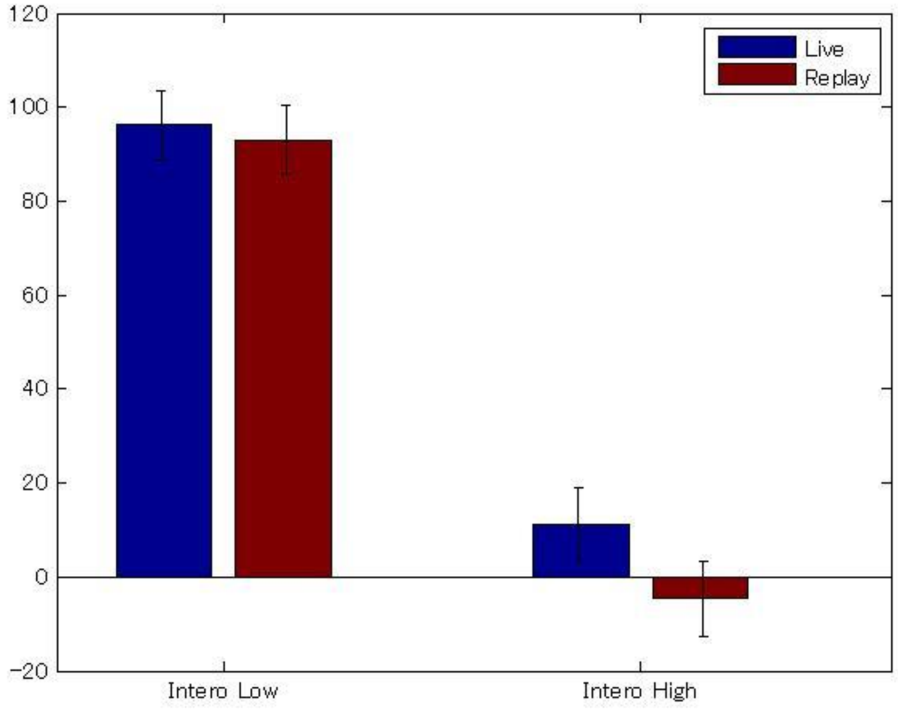

Figure 5. Results of mean reported interval in Experiment 2 by median-split by performance on a heartbeat counting task. Reported-Reference of 0 ms indicates that the interval was reported veridically, while negative values indicate that the interval was reported as shorter and positive values as longer than the physical presentation. Error bars indicate +/- standard error of the mean.

3.2 Time perception in substitutional reality: which "reality" are you in?

An interesting observation from our initial investigations into temporal perception in substitutional reality (Section 3.1) was the confirmation that the speed of a scene strongly affects reported duration. This supports previous (van Rijn, 2014) findings in a virtual reality driving setting wherein the motion in the scene is vective, that is, the person was moving forward through the scene. In our case, the motion reflects self-generated bodily movements (more relevant to embodied cognition scenarios). The fact that speed of visual scene affects subjective temporal perception regardless of type of scene led us to speculate about whether scene-speed driven differences in temporal judgements might be a reasonable index of the degree to which an observer feels present in a substitutional reality experience.

To test this idea we used the same basic experimental setup as in Experiment 2 from Section 3.1, but rather than presenting a full field of view through the Oculus we constrained the range of view to a rectangular screen centered on fixation (70% of display size; see Figure 6). There were two presentation conditions. In one condition (frame), outside of the central presentation area where the market scene was presented, we presented the 'real world' projected from cameras mounted on the front of the Oculus headset. This view is what the observer would have been able to see in this visual range if they were not wearing the headset. In the other condition (black) the area outside of the central market presentation was blacked out. Our hypothesis was that in the frame scene, in which observers had information from both the 'real world' and the substitutional world, the effect of the different scene speeds (slow, normal and fast) would be reduced or abolished as the observer would still feel as though they were existing in the 'real world'. As the speed of the real world scene was not changed, changes in the speed of the substitutional reality scene would not affect temporal judgements. In the black scene, the range of view of the market scene was limited but this substitutional reality was the only information on which the observer had to base their world. Consequently, we expected to see a strong influence of scene speed on subjective duration, as reported in Section 2.

As in previous experiments, observers again also participated in a heartbeat counting task as an indicator of their interoceptive accuracy.

Figure 6. Example of the different stimulus presentation conditions from the Experiment described in 3.2. The left panel contains a depiction of the framed surround presentation with the outside of the frame containing video of the 'real world' around the observer. The right panel contains a depiction of the black surround presentation with the outside of the frame filled with black. In both cases the centrally presented scene was the same SR market scene.

3.2.1 Results and Discussion

25 observers naïve as to experimental purpose participated in the experiment. Five observers were excluded from the analysis because their reports did not show a proportional relation to the physically presented intervals.

As indicated by Figure 7, there was a main effect of presentation condition such that reported durations in the Framed (intervals overestimated by 128 ms relative to veridical) condition were significantly longer than those in the Black (intervals overestimated by 140 ms relative to veridical) condition (F(1,19) = 4.21, $p<0.05$). There was also a main effect of the speed of scene such that reported durations in scenes in which the presentation was slow were shorter than normal speed and fast scenes (intervals overestimated by 129, 131 and 146 ms relative to veridical, respectively, F(2,19) = 3.79, $p<0.05$). There was no significant interaction of presentation condition and speed. As there was no evidence of an interaction between scene speed and presentation factors, these results do not support the hypothesis that presentations containing a compromise of substitutional reality and 'real world' information would result in the observer being immune to changes in speed of the centrally presented visual scene.

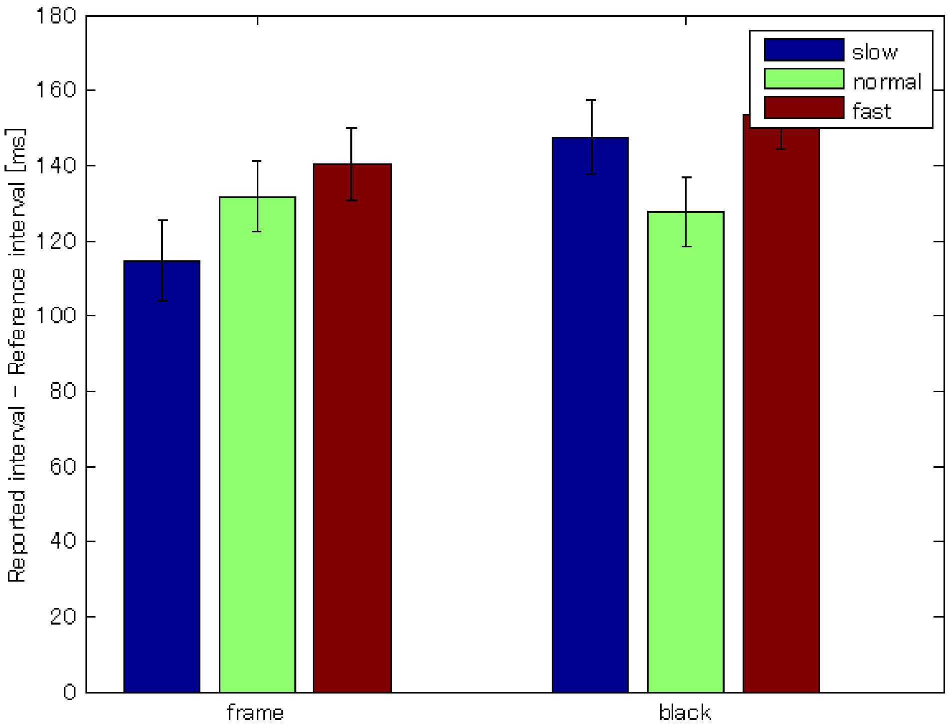

Figure 7. Results of mean reported interval in the framed reality experiment by condition. Reported-Reference of 0 ms indicates that the interval was reported veridically, while negative values indicate that the interval was reported as shorter and positive values as longer than the physical presentation. Error bars indicate +/- standard error of the mean.

To again investigate the possible interaction of interoception and temporal perception split the data into high and low interoceptive accuracy observers, based on median performance in the heart beat counting task. As shown in Figure 8, and as reported previously (3.1.2), for observers who demonstrated low interoceptive accuracy, reported durations were overestimated relative to the presented duration by more than observers with high interoceptive accuracy (Low interoceptive accuracy: 249 ms for Framed, 259 ms for Black; High interoceptive accuracy: -20 ms for Framed, -27 ms for Black; F(1,10) = 771.65, p<0.01). Again however, it is not possible at this stage to determine whether this difference indicates a true difference in temporal performance related to interoceptive accuracy, or whether it simply indicates a difference in performance in general temporal perception. Further studies are required to address this issue.

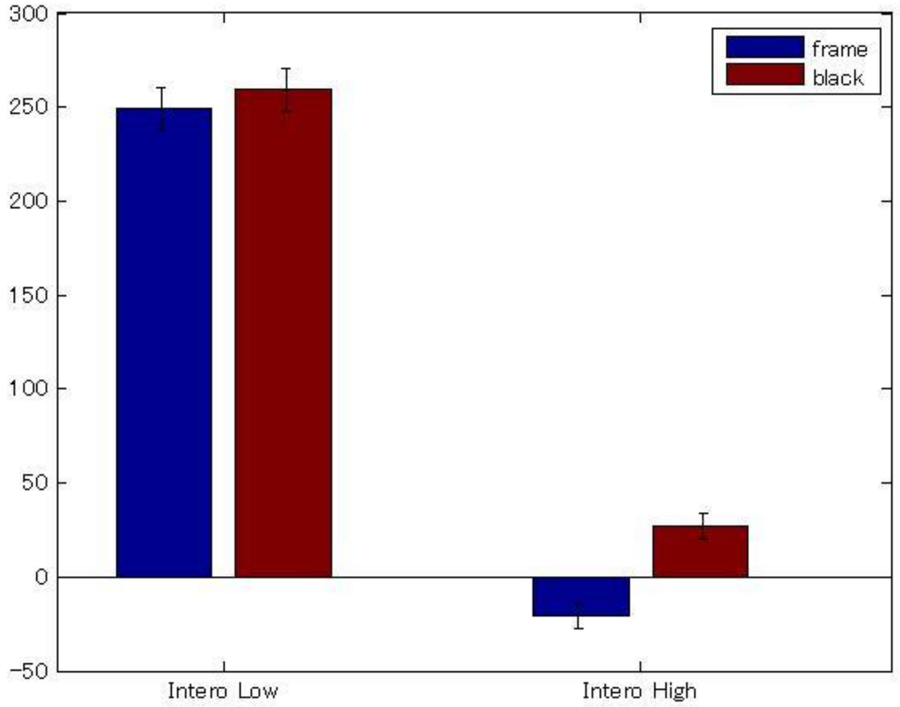

Figure 8. Results of mean reported interval in framed reality experiment by median-split by performance on a heartbeat counting task. Reported-Reference of 0 ms indicates that the interval was reported veridically, while negative values indicate that the interval was reported as shorter and positive values as longer than the physical presentation. Error bars indicate +/- standard error of the mean.

4 Empathic aspects of social time perception

Shared time perception is crucial for accomplishing fluent collaboration in multi-agent setups. Partner Université Blaise Pascal brings in TIMESTORM important expertise in the field, which plans to expand further in the second year of the project.

Until recently, there was no systematic investigation of the effect of emotional context on humans' abilities to synchronize with others and how these abilities change in the course of development. To address this knowledge gap, TIMESTORM members at Université Blaise Pascal recently explored the impact of the emotional states of the self and others on time judgments. The first published result (Droit-Volet & Provasi, 2015) investigated the influence of emotional images on a synchronized tapping task. Observers were presented with a rhythmic sequence of faces containing different expressions (disgust, neutrality, sadness, joy, anger, and fear) and were required to synchronise their tapping with presentation of the images. Following the synchronization phase, the face images would stop appearing, but observers were required to continue tapping at the same rhythm. As indicated in Figure 9, tapping continuation following exposure to fearful or angry faces caused a speeding up of the tapping rhythm compared with presentations of other emotions (neutral, disgust sadness, joy). These results may indicate a speeding up of the internal motor rhythm associated with particular emotional stimuli.

A second study is scheduled for the following months by UBP, which will build on the existing results and will further explore the developmental aspects of shared time perception with children of different ages (also associated with Deliverable 4.2).

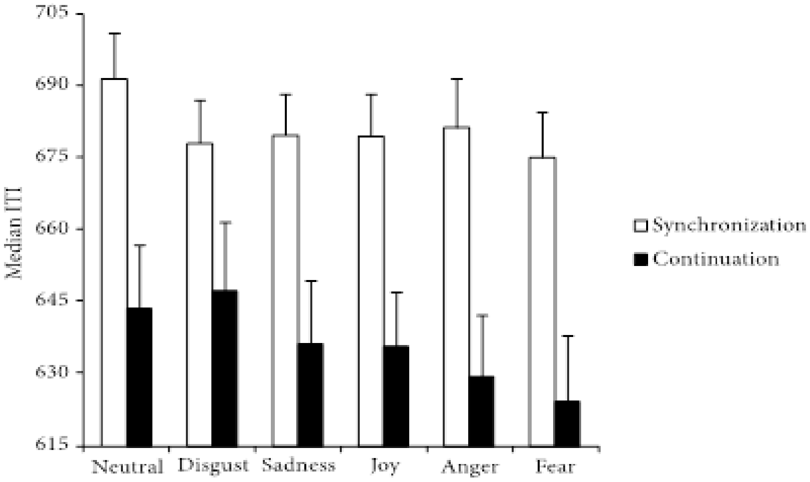

Figure 9. Median inter-tap intervals in the synchronization and continuation phases as a function of the emotional of faces presented in the rhythmic sequence.

5 Towards a 'predictive coding' model of temporal and interoceptive perception

As part of the collaborative work between University of Sussex and Imperial College London, we aim to develop predictive processing model(s) of temporal and interoceptive perception. These models will utilise Bayesian Inference through probabilistic population codes (e.g. Ma et al., 2006; Figure 10). To this end, Timestorm members at Imperial College London, David Bhowmik, Zafeirios Fountas, Pedro Mediano and Murray Shanahan, have begun basic implementation of these models. In population code models, the receptive fields of populations of neurons can be altered by recently exposed stimulus values. Selective changes centered on the mean of the recently experienced stimulus values can result in altered output from the population code (putatively, altered perception). The work conducted thus far demonstrates that such changes in the gain of neurons in the population code results in a shifted Gaussian output. Changes in the variances or means in isolation, or a combination of changes in the gain, variances and means produce skewed Gaussian outputs (e.g. Figure 3 Schwartz et al., 2007).

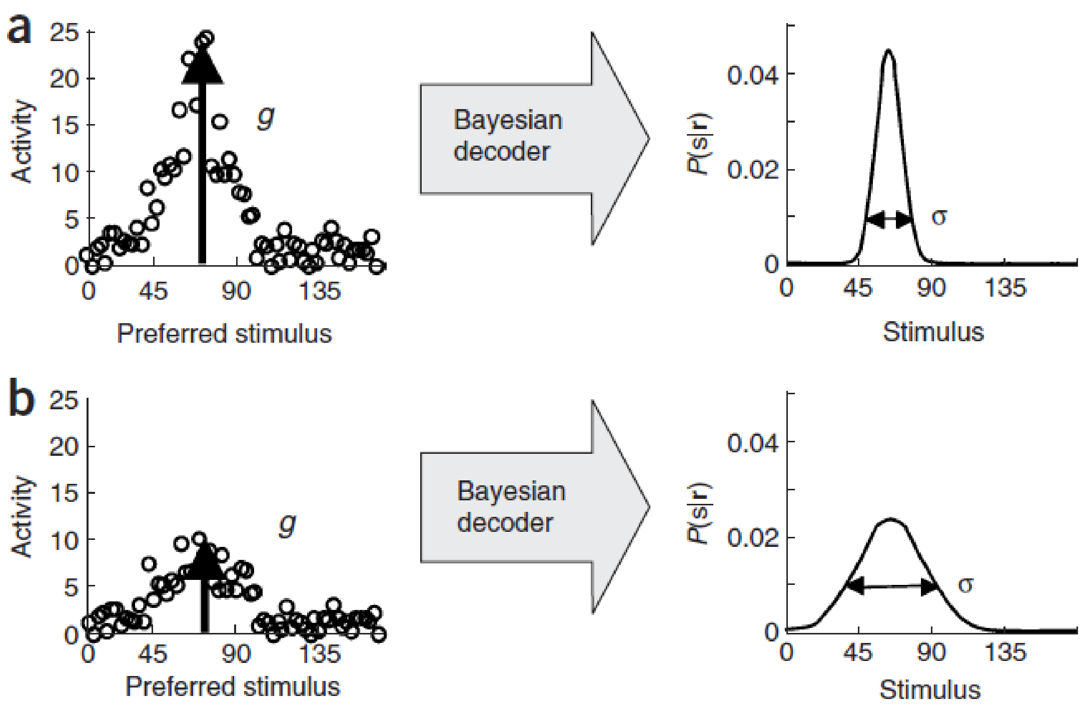

Figure 10. The population activity of a single trial response to a stimulus value (left) along with the stimulus estimate based on the decoded response (right). Neurons in this population have a Gaussian tuning curve. The mean of the distribution is close to the stimulus value at which population activity peaks. Gain of the population code is inversely proportional to the variance of the distribution (demonstrated as the difference between low variance (a) and high variance response (b). Figure taken from Ma et al., 2006.

At this stage, however, little is known about the precise architecture required to accurately replicate human-like perception. To further understand this aspect, we have initially focused on trying to elucidate the role of recent sensory history on subsequent perception in the temporal domain. Below we outline the basic problem and our recent progress. The work described in section 5.1 was carried out by Warrick Roseboom, Darren Rhodes and Anil Seth at the University of Sussex. The work in section 5.2 was carried out by Warrick Roseboom and Acer Chang, also University of Sussex.

5.1 The effect of recent sensory experience on time perception

The physical world provides a constant stream of sensory information that the human brain has to process, analyze and respond to. Estimates of perceptual properties are not veridical, that is, for example, we do not perceive the size of an object accurately – rather, we may perceive the object as being slightly larger or smaller than reality. Perceptual biases persist in a wide range of experimental tasks such as angle, distance and length estimation (Petzschner, Glasauer, & Stephan, 2015). The modulation of perception is also evident when an organism learns the average length of an interval after being repeatedly exposed to a distribution of interval duration (Jazayeri & Shadlen, 2010; 2015). Behavioral responses are affected by the mean and distribution of previously exposed durations. Such tendencies suggest that the brain uses a priori information in the estimation of perceptual properties.

Bayesian Decision Theory (BDT) provides a descriptive account of how such perceptual biases may occur. To explain how an interval may be perceived as being closer to the mean of a distribution of previously exposed intervals, the Bayesian framework posits that you combine prior expectations with current sensory estimates. Prior expectations are acquired by learning the distribution of intervals an individual has been exposed to and are represented as a probability distribution (Figure 11). The current sensory estimate (likelihood) of the duration of an interval is also represented. The reported interval is then the point-by-point multiplication of the prior and likelihood, resulting in the posterior distribution from which an estimate is obtained (Shi et al. 2013). Recently, it has further been proposed that the timing of single events also obeys the Bayesian framework (Rhodes & Di Luca, in review). In this account, the reported timing of a signal is the combination of the time you expected to sense a stimulus with the time you sensed it. Here, we begin our investigation into how these expectations are developed and maintained by examining the boundaries of what kind of information the priors may contain. Specifically, we wanted to know whether it is possible for human perception to maintain distinct priors, constrained by the type of signal presented – sensory modality contingent priors.

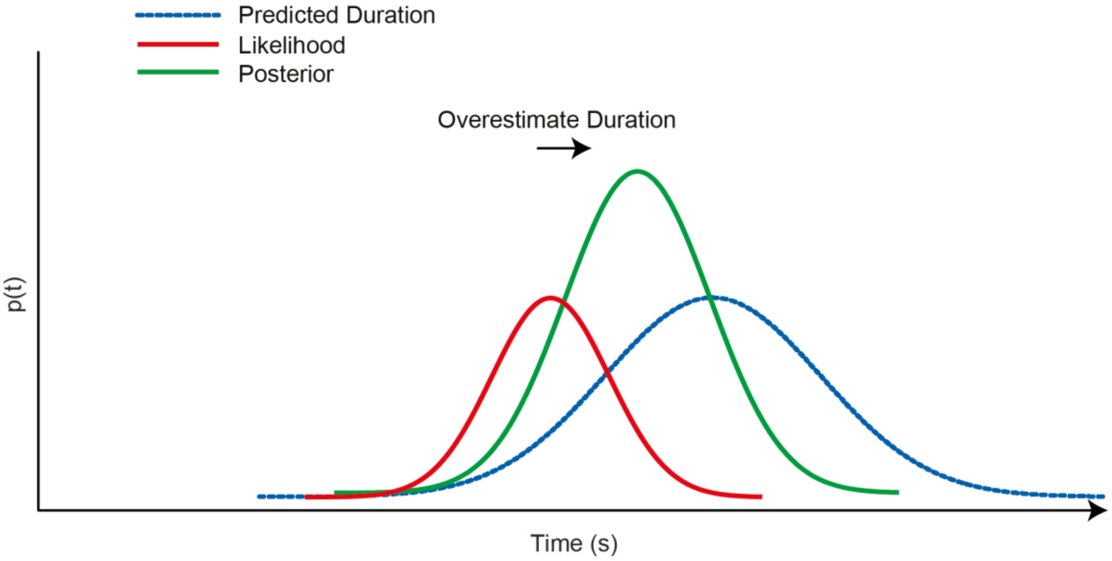

Figure 11. Illustration of the 'regularization' effect in interval duration estimation. The reported duration of an interval (posterior; green line), is the result of the combination of the predicted duration of an interval (prior; blue line), with the current sensory estimate of an interval (likelihood; red line). In this example, an interval that was shorter than the mean of previously exposed intervals is reported as being longer than it was in reality.

5.1.1 Methods

To investigate the effect of contingent priors in duration perception, we presented observers with sequences of four successive events. The first three events in the sequence were presented isochronously, that is, the temporal interval between them (the inter-stimulus-interval; ISI) was physically regular. The fourth signal in the sequence was presented at some variable timing across a broad range of anisochrony, depending on the ISI. Observers were required to report whether the final event in the sequence appeared to occur early or late relative to the rhythm established by the first three events (Li, Rhodes & Di Luca, 2016; Figure 12). The event sequences could be comprised of quiet auditory beeps (1000 Hz, ~50 db SPL) or dim visual flashes (luminance defined Gaussian blob). To look for the modality contingent effect, in a single experimental session participants viewed 50% of trials contained auditory sequences and 50% visual, randomly interspersed on a trial-by-trial basis. Observers completed two one-hour experimental sessions over two days. In one experimental session, the sequences of visual events were normally distributed around a mean ISI of 650 ms (min 50 ms; max 1250 ms) while the auditory event sequences were normally distributed around a mean ISI of 1550 ms (min 950 ms; max 2150 ms). In the other session, the modality means were switched (i.e. visual 1550 ms, auditory 650 ms). The order of completion of these sessions was counter-balanced across participants.

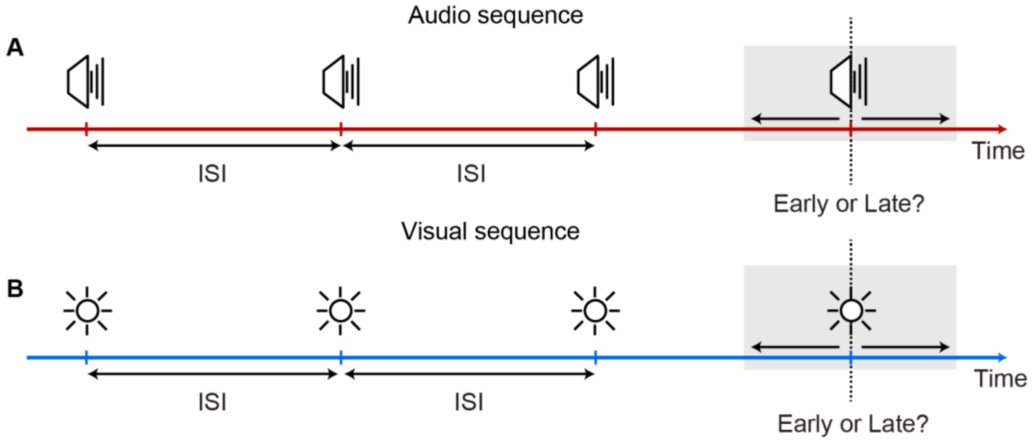

Figure 12. Example stimulus sequence from experiment described in 4.1. In each trial, a sequence of four events, either visual or auditory was presented. The first three events were separated from one another by the same temporal interval, with the fourth and final event presented at some temporal offset from physical isochrony. The observer's task was to report whether the final event was presented early or late relative to the expected rhythm.

5.1.2 Results and Discussion

19 naïve observers and WR participated in the experiment. In order to examine the possible distortions from veridical duration perception, we fitted cumulative Gaussians to the early/late judgments for each ISI, for each of the four conditions (650/1550, visual/auditory) for the data obtained from all 20 observers. The mean of this fitted cumulative Gaussian - the point at which observers report with equal likelihood that that final stimulus was earlier or later than expected - was taken as the point of subjective isochrony (PSI) The PSI is taken as an estimate of the timing at which the observers subjectively felt the same time had elapsed between the third and last events as had elapsed between each of the previous repetitions.

As Figure 13 demonstrates, there were consistent biases in the point of subjective isochrony (PSI) as a function of ISI. The PSI biases are consistent with a temporal regularization effect such that observers reported relatively shorter intervals to be longer than experienced and vice versa. Importantly, these results suggest that the regularization effect is constrained by the modality of presentation – evidence for modality specific temporal predictions/priors. If this were not the case, the lines of PSIs depicted in Figure 13 would be continuous and the PSIs for ISIs that are physically identical (950 - 1250 ms) would be the same. Moreover, the effect appears to be identical for each modality in each experimental block (i.e. the lines overlap for both audio and visual short and audio and visual long) suggesting that the influence of the prior on temporal perception was, in this case, similar for auditory and visual events.

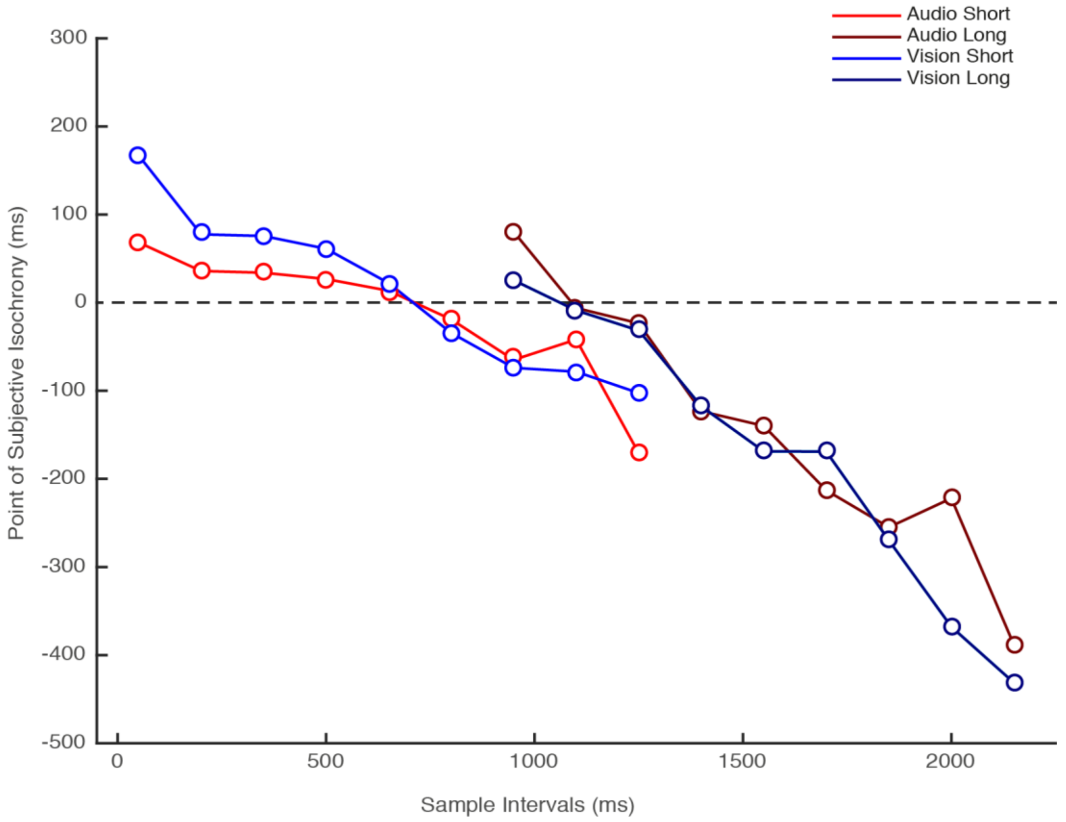

Figure 13. Point of subjective isochrony (PSI) as a function of the interval presented. PSI is here estimated based on the data of all 20 observers. Each line represents the PSI for each of the four conditions. A PSI of 0 would mean that a sample interval was perceived veridically, but here the data shows biases consistent with a temporal regularization effect.

Shown in Figure 14 are the averaged results for the PSI of 20 observers, having fitted a cumulative Gaussian to early/late judgements for each of the four conditions for only the overlapping ISIs (950-1250 ms), for each observer. A two factor analysis of variance (mean ISI, 650 or 1550 ms; modality of presentation, visual or auditory) reveals a significant main effect (F(1,19) = 15.19, p = 0.001) of ISI-mean (650 or 1550 ms) such that the PSI is significantly shorter for judgements of these overlapping intervals in the context of a 650 mean (mean PSI = -77 ms) than in the context of a 1550 ms mean (mean PSI = -3 ms). No other effects were significant (*ps* > 0.57). The absence of a main effect of modality again suggests that the influence of the prior on temporal perception was similar for auditory and visual events.

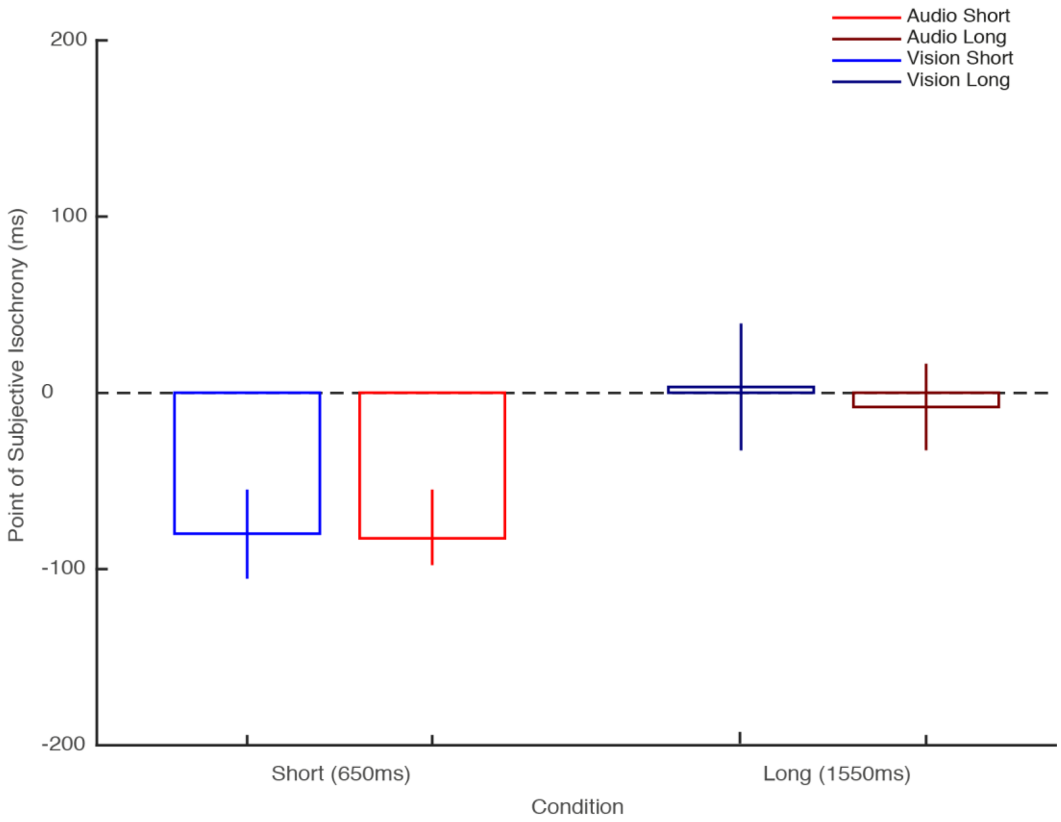

Figure 14. Mean point of subjective isochrony (PSI) for 20 observers obtained from cumulative Gaussians fitted to early/late judgements for each of the four conditions for the overlapping ISIs (950-1250 ms).

These results clearly suggest that it is possible for human observers to concurrently acquire and maintain distinct priors for temporal perception, constrained by the sensory modality of presentation. The added complexity of requiring the ability to acquire and maintain multiple distinct priors suggests that the effect of sensory expectation or prediction (the prior) in cases such as this (e.g. Jazayeri and Shadlen, 2010; Li et al., 2016; Rhodes & Di Luca, in press) is unlikely to be driven by very low level sensory processes, but is rather driven by mid or high level perceptual processes. Further experiments will be required to confirm this.

5.2 Neural correlates of temporal interval prediction: temporal mismatch negativity

An important aspect of elaborating computational models of human temporal processing is to determine the possible neural underpinnings of such processes. Asserting the importance of prediction within determining subsequent perception, many classical studies have demonstrated the action of simple predictive networks in the human brain using electrophysiological methods (EEG). In particular, the classic mismatch negativity (MMN) effect has been extensively studied over the past several decades (see Näätänen et al., 2007, for review). The MMN refers to a difference in the event related potentials (ERP) associated with sensory events that have been frequently experienced in the recent past, the standard, relative to ERPs elicited following rare, or deviant sensory events (see Figure 15).

The classic example of the MMN is found for auditory pitch. In a regular sequence of events at some temporal frequency (e.g. 2 Hz) a brief auditory tone is presented. The tone can be either the standard tone, so-called because it makes up ~85% of presentations in the sequence, or the deviant tone, which appears on the remaining ~15%. In the example in Figure 15 the standard is a 1000 Hz puretone, lower in pitch than the deviant 1127 Hz tone. The presence of the deviant is randomly interspersed within the sequence such that the observer cannot be certain of when it will appear.

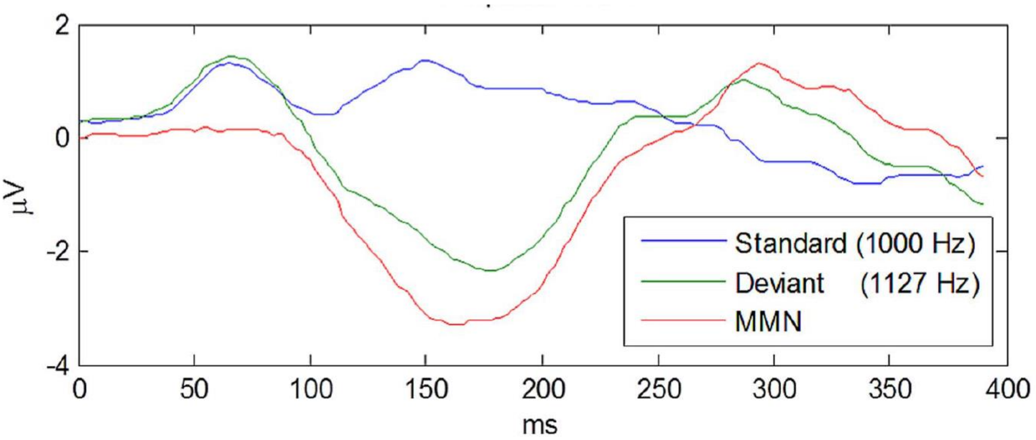

Figure 15. Example ERPs elicited following standard (blue) or deviant (green) auditory event in a series of auditory events. The red line indicates the difference between these ERPs – the MMN effect. Figure taken from Leider et al., 2013.

A temporal version of the MMN has also been reported (e.g. Campbell & Davalos, 2015) wherein the deviant is defined by a temporal irregularity in the sequence of events (e.g. successive events in a stream are typically separated by 500 ms but the deviant presentation occurs after only 250 ms). However, we are interested perception of temporal intervals themselves, rather than irregularities in rhythm. To our knowledge, no previous studies have addressed whether it is possible to find a MMN effect for temporal intervals as distinct from rhythm. As MMN has long been treated as a neural correlate of sensory prediction, and indeed elaborated predictive processing models have been developed to describe the MMN in other contexts (e.g. from Leider et al., 2013; Phillips et al., 2015), extending this approach to temporal intervals is a potentially highly productive research direction. Additionally, as we are interested in prediction of temporal intervals generally, we will also investigate whether it is possible to find evidence of MMN for different combinations of sensory signals. If we can find equivalent neural processes accompanying different signal combinations for specific temporal intervals, this would be evidence for a generalized predictive process of temporal interval, unconstrained by signal modality of origin. Given that behavioural studies examining sensory prediction are often conducted under similar experimental conditions to that we examine here (e.g. 5.1), finding duration specific neural correlates of prediction would be a major step forward in our understanding of how the predictive processes that we find evidence for behaviourally (5.1) are implemented in the human brain.

5.2.1 Methods

There were three stimulus presentation conditions: visual isochronous, visual anisochronous, cross-modal. In all cases, individual visual stimuli were 10 ms flashes of luminance defined Gaussian blobs against a gray background. Auditory stimuli were 10 ms 1500 Hz puretones. Intervals were defined by the onset and offset of some combination of these events. For the two visual conditions (Isochronous and Anisochronous), the onset and offset were both visual flashes. For the cross-modal condition the onset and offset were presented in different sensory modalities. There were two interval conditions: standard interval of 150 ms and deviant of 600 ms, and vice versa. All conditions were tested in individual blocks with the order of completion counterbalanced within an observer for the standard/deviant interval duration, and across observers for the stimulus types.

Each block of trials for each condition consisted of 250 individual presentations, presented at either exactly 1.5 Hz (Isochronous), or with an average presentation rate of 1.5 Hz and a given inter-pair-interval drawn from a uniform random distribution with a range of 1-2 Hz (Anisochronous and Cross-modal). Within these 250 trials 80% (200) were standard presentations and 20% (50) were deviants. The position of the deviants within the sequence was pseudo-randomised but constrained to be presented at least once in every 10 trials. There was no explicit task – observers were required only to passively pay attention to the sequence of events. Each block of 250 trials took approximately 10 minutes to complete – the entire experiment taking approximately 1 hour.

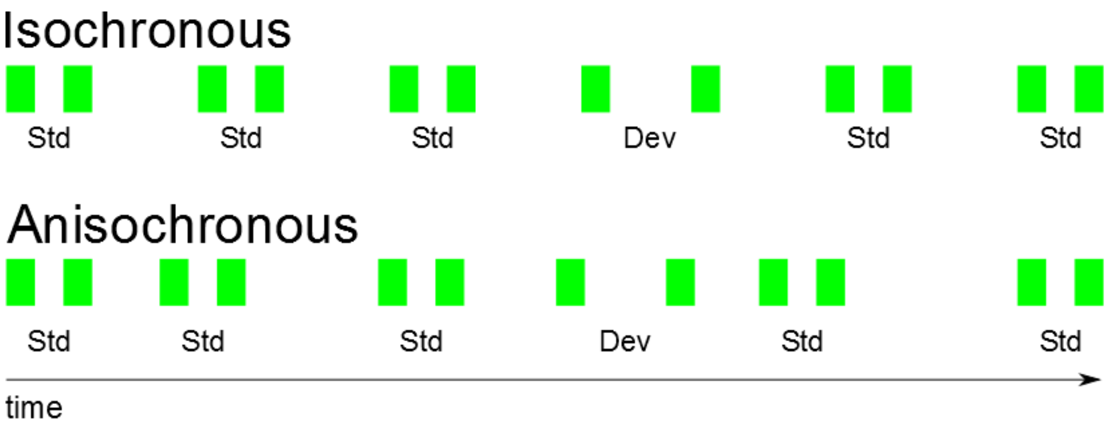

Figure 16. Schematic depiction of example stimulus sequences from Experiment 4.2. In this example, the standard interval is defined by two events separated by a shorter temporal interval (e.g. 150 ms) than the deviant events (e.g. 600 ms). In the Isochronous presentation, the pairs of events repeat regularly and the deviant deviates not only in interval duration, but also rhythm. In the Anisochronous presentation the interval between successive stimulus pairs is pseudo-random and therefore, the deviant differed from the standard only in interval duration. This allows us to differentiate the neural processes associated with prediction of rhythm from those associated with prediction a temporal interval itself.

EEG signals were recorded using a 64 channel cap at a sampling rate of 2048 Hz. The data were down-sampled to 512 Hz, a band-pass (0.5~30 Hz) filter applied, and epoched (-100-1000 ms) time locked to the timing of the first stimulus in each pair. Baseline potential was determined based on an average of recordings from -100~0 ms, relative to the first stimulus. Eye movement artefacts were removed by independent components analysis.

5.2.2 Preliminary Results and Discussion

As can be seen from the preliminary data shown in Figure 17, there is evidence for a MMN effect in all three conditions: the ERPs associated with the standard and deviant presentation strongly diverge around 100 ms after the presentation of the second, interval defining stimulus (250 ms after the initial stimulus of the pair). Figure 17 represents the data from a single observer, but a similar pattern of results is evident in the data of three further observers. Although the effect in the cross-modal condition appears qualitatively different, it is important to consider two things. First, while the duration was physically the same for all conditions, it was not in any way subjectively equated. Therefore the degree of deviance may not be the same in the cross-modal condition as in the visual only conditions. Indeed, this might be expected given that the precision of audio-visual temporal judgements is known to be worse than that for within-modality combinations (Burr et al., 2009). In other sensory contexts the magnitude of MMN is known to be related to the degree of apparent deviance (Näätänen et al., 2007), and so that may contribute to any possible differences here. The second point to consider is that the data presented in Figure 17 is from the single Cz electrode. While it is reasonable that ERPs resulting from the two visual only conditions might be similar despite their difference in temporal structure, the cross-modal condition presents two different sensory modalities, and so the source of predictive processing may be spatially dissimilar from these other cases.

Despite these possible problems, the most critical aspect of this data is that the MMN effect can be obtained in the Anisochronous visual condition and that this effect is qualitatively similar to that obtained in the rhythmically presented Isochronous condition. As the Anisochronous visual condition lacks any predictable rhythmic information, the MMN here must be indexing predictions made explicitly about the duration of a presented interval. Further study is needed to constrain precisely what the MMN represents in this case, and what exactly the underlying neural source is, but these early results show promise in our ability to find the neural correlates of temporal interval prediction in the human brain.

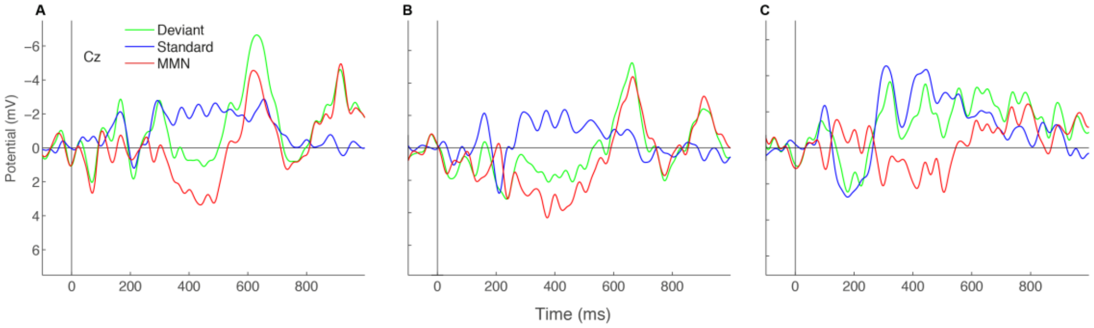

Figure 17. Data from an example observer. (A) ERPs recorded from electrode Cz during presentations of 150 ms duration standard (blue) and 150 ms deviant (green) trials, along with their difference (MMN; red) for the Isochronously presented visual stimulus condition. (B) and (C) as above, but for the Anisochronously presented visual stimulus condition, and the cross-modally presented condition respectively.

6 References

Burr, D. et al. (2009). Temporal mechanisms of multimodal binding. Proc. R. Soc. B. 276, 1761-69.

Campbell, A.M. & Davalos, D.B. (2015). Levels of attention and task difficulty in the modulation of interval duration mismatch negativity. Front. Psychol. 6:1619.

Droit-Volet, S. & Provasi, J. (2015). Synchrony and emotion. Timing and Time perception.

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biol Psychol, 104*, 65-74.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. Nat Neurosci, 13(8), 1020-1026. doi: 10.1038/nn.2590

Jazayeri M., & Shadlen M. N. (2015). A Neural Mechanism for Sensing and Reproducing a Time Interval. Curr Biol: 1–12.

Li, M., Rhodes, D. & Di Luca, M. (2016). For the Last Time: Temporal Sensitivity and Perceived Timing of the Final Stimulus in an Isochronous Sequence. Timing & Time Perception.

Lieder, F., Stephan, K.E., Daunizeau, J., Garrido, M.I., Friston, K.J. (2013). A Neurocomputational Model of the Mismatch Negativity. PLoS Comput Biol 9(11): e1003288.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clinical Neurophysiology 118: 2544–2590.

Petzschner FH, Glasauer S, Stephan KE (2015) A Bayesian perspective on magnitude estimation. Trends Cogn Sci 19: 285–293. doi:10.1016/j.tics.2015.03.002.

Rhodes, D. & Di Luca, M. (in review). Individual Stimulus Timing is Perceived by Integrating Sensory Information and Dynamically-Updated Expectations.

Schandry, R. (1981). Heart beat perception and emotional experience. Psychophysiology, 18(4), 483–488.

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. Nature Reviews Neuroscience, 8(7), 522–535.

Shi, Z., Church, R. M. & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends. Cogn. Sci.* 17**,** 556–564

Suzuki K., Wakisaka S., Fujii N. (2012). Substitutional reality system: a novel experimental platform for experiencing alternative reality. Scientific Reports 2:459 10.1038/srep00459

Suzuki, K. et al. (2013) Multisensory integration across interoceptive and exteroceptive domains modulates self-experience in the rubber-hand illusion. Neuropsychologia, 51, 2909–2917

van Rijn, H. (2014). It's time to take the psychology of biological time into account: speed of driving affects a trip's subjective duration. *Front. Psychology* **5**, 1028.