The role of multisensory interplay in enabling temporal expectations

Felix Ball[1,2*], Lara E. Michels[1], Carsten Thiele[1], Toemme Noesselt[1,3]

[1] Department of Biological Psychology, Otto-von-Guericke University, Magdeburg, Germany
[2] Department of Neurology, Otto-von-Guericke University, Magdeburg, Germany
[3] Center of Behavioural Brain Sciences, Otto-von-Guericke University, Magdeburg, Germany

Abstract

Temporal regularities can guide our attention to focus on a particular moment in time and to be especially vigilant just then. Previous research provided evidence for the influence of temporal expectation on perceptual processing in unisensory auditory, visual, and tactile contexts. However, in real life we are often exposed to a complex and continuous stream of multisensory events. Here we tested – in a series of experiments – whether temporal expectations can enhance perception in multisensory contexts and whether this enhancement differs from enhancements in unisensory contexts. Our discrimination paradigm contained near-threshold targets (subject-specific 75% discrimination accuracy) embedded in a sequence of distractors. The likelihood of target occurrence (early or late) was manipulated block-wise. Furthermore, we tested whether spatial and modality-specific target uncertainty (i.e. predictable vs. unpredictable target position or modality) would affect temporal expectation (TE) measured with perceptual sensitivity (d') and response times (RT). In all our experiments, hidden temporal regularities improved performance for expected multisensory targets. Moreover, multisensory performance was unaffected by spatial and modality-specific uncertainty, whereas unisensory TE effects on d' but not RT were modulated by spatial and modality-specific uncertainty. Additionally, the size of the temporal expectation effect, i.e. the increase in perceptual sensitivity and decrease of RT, scaled linearly with the likelihood of expected targets. Finally, temporal expectation effects were unaffected by varying target position within the stream. Together, our results strongly suggest that participants quickly adapt to novel temporal contexts, that they benefit from multisensory (relative to unisensory) stimulation and that multisensory benefits are maximal if the stimulus-driven uncertainty is highest. We propose that enhanced informational content (i.e. multisensory stimulation) enables the robust extraction of temporal regularities which in turn boost (uni-)sensory representations.

Key Words: temporal expectation, temporal orienting, multisensory interplay, redundant target, spatial coincidence, auditory dominance

---

*Corresponding author:
Felix Ball
Otto-von-Guericke-University
Institute of Psychology
Department of Biological Psychology
Universitätsplatz 2, 39106 Magdeburg, Germany
felix.ball@ovgu.de

1 Introduction

The amount of information organisms are confronted with at any given moment is tremendous. It is therefore imperative to focus on particular aspects of the incoming information and to preferentially process the most relevant parts — as both information overflow and missing important bits of information can have severe consequences (e.g. in traffic). Spatial attention offers one solution to selectively increase the salience of particular information and has been the focus of numerous previous investigations (Ball et al., 2015; Ball and Busch, 2015; Ball et al., 2014; Kovalenko and Busch, 2016; Luck et al., 2004; Posner et al., 1980; Posner, 1980; Yeshurun and Carrasco, 1998). Another way to facilitate information processing is to anticipate when future objects and events may occur and what these events/objects might be (the *what*: Baylis and Driver (1993); Behrmann et al. (1998); Chen (2000); Duncan (1984); Kramer et al. (1997); Vecera and Farah (1994) and *when*: Correa et al. (2006, 2004); Coull and Nobre (2008); Doherty et al. (2005); Nobre (2001); Rohenkohl et al. (2011, 2014)). In this article we will differentiate between different aspects of temporal information influencing behaviour. The term temporal predictability will be used to denote exogenous factors, e.g. the manipulation of temporal regularities by experimental design. Endogenous factors derived from these objective temporal regularities – i.e. temporal expectations (TE) generated by the participant – will be referred to as temporal attention or temporal expectation (in accord with e.g. Bendixen et al., 2012).

Previous research on temporal attention preferentially used three main paradigms (see Nobre and Rohenkohl, 2014, for a recent review) which have been based on rhythmic variations, temporal cueing, and foreperiod duration. In studies using rhythmic variations, temporal expectations are automatically generated by presenting an isochronous stimulus sequence (Cravo et al., 2013; Jones et al., 2002; Mathewson et al., 2010; Rohenkohl et al., 2012; Sanabria et al., 2011). Target stimuli are either shown at the end of or are embedded within the rhythmic sequence. Only targets presented in phase with the rhythm are temporally predictable, while arrhythmically presented targets are unpredictable. In temporal cueing experiments (Correa et al., 2004; Coull and Nobre, 1998; Griffin et al., 2001, 2002; Jepma et al., 2012; Miniussi et al., 1999) a signal predicts the delay between cue and target (e.g. 200 ms vs. 800 ms) with

---

List of abbreviations:

TE	– temporal expectation/ temporal expectancy
IE	– inverse efficiency
MSI	– multisensory interplay
EEG	– electroencephalography
A	– audio/ auditory
V	– visual
AV	– audiovisual
RT	– response time

a certain probability (e.g. 75%), in close resemblance to spatial cueing paradigms (Posner et al., 1980; Posner, 1980). Here, TE can be manipulated on a trial-by-trial basis, whereas belief about cue validity builds up over time. There is corroborating evidence from both rhythm and cueing studies indicating that temporal predictability of events enables us to create temporal expectations which in turn improve performance: they enhance detectability of targets, increase accuracy in discrimination tasks (e.g. frequency judgement), and decrease response times (Nobre and Rohenkohl, 2014). The third approach investigating TE utilizes foreperiod paradigms (Lange and Röder, 2006; Lange et al., 2003; Niemi and Näätänen, 1981; Rolke and Hofmann, 2007; Westheimer and Ley, 1996) in which hazard rates – the conditional probability of the occurrence of a target given that it has not yet been presented – are manipulated (Nobre and Rohenkohl, 2014). In particular, the cue-target delay (i.e. the foreperiod) is varied between blocks (e.g. short or long foreperiod); temporal regularities are not explicitly cued, thus temporal expectation builds up over trials. In these studies performance consistently decreases with increasing foreperiod duration, and it has been suggested that this might be due to participant's decreased temporal precision or participant's higher temporal uncertainty with increasing cue-target intervals (Klemmer, 1956; Näätänen and Merisalo, 1977; Näätänen et al., 1974; Niemi and Näätänen, 1981).

The paradigms above all have in common that the effects of temporal attention were tested implicitly – i.e. knowledge about time-of-target-occurrence was not explicitly assessed – but nevertheless, the temporal predictable context improved performance. Another line of research directly investigated the representation of time using temporal bisection tasks and switch paradigms (Akdoğan and Balcı, 2016; Balci et al., 2009; Balcı et al., 2011; Bogacz et al., 2006; Çavdaroğlu et al., 2014; Çoşkun et al., 2015; Freestone et al., 2015) among other tasks. Results from both human and animal studies revealed that participants were able to base their temporal decisions on – sometimes noisy – time estimates. The noise intrinsic in these time estimates can be due to exogenous factors (variability of external sources) and additionally due to the endogenous properties of the temporal representations. Concordantly, several computational models have been put forward to account for the observed effects including pacemaker accumulator and drift diffusion models (see e.g. for a recent review Balcı and Simen, 2016). Given several similarities between explicit and implicit timing results, intrinsic temporal estimators such as pacemaker accumulators might be used for both, the explicit and implicit use of temporal regularities.

Another similarity of the paradigms mentioned above is that they investigate temporal attention explicitly or implicitly but in the absence of additional — potentially distracting — information. Indeed, in most of these studies, the target is presented in isolation and can easily be perceived as target (e.g.

targets are colour coded, presented at the end of sequences, or presented in isolation after the cue, and thus are quite obvious). In the last years, novel paradigms have been designed to create more ecologically valid contexts with distracting information and with targets which are less obvious (e.g. Jaramillo and Zador, 2011; Shen and Alain, 2011). Among them are attentional blink studies (stimulus sequences with an embedded target and probe; e.g. Shen and Alain, 2011, 2012) and studies combining foreperiod with rhythmic designs in which the hazard rate of targets – which themselves are hidden in a sequence of distracting stimuli – varies (Jaramillo and Zador, 2011).

A different promising approach to investigate temporal expectation in more ecologically valid context could include the use of multisensory stimuli, as many real-life events stimulate more than one sense. Concordantly, there is evidence that seeing lip movements can enhance speech perception (Grant and Greenberg, 2001; Reisberg et al., 1987; Risberg and Lubker, 1978; Sumby and Pollack, 1954) and that multisensory perception also improves later memory retrieval (Luria, 1968; Shams and Seitz, 2008). Moreover, several psychophysical studies indicate that redundant multisensory stimulation can improve performance relative to unisensory stimulation (Alais and Burr, 2004; Driver and Noesselt, 2008; Forster et al., 2002; Gondan et al., 2005; Jaekl and Harris, 2009; Noesselt et al., 2010; Parise et al., 2012; Sinnett et al., 2008; Stevenson et al., 2014; Talsma et al., 2007; Van der Burg et al., 2008) and some have pointed at enhanced MSI with less reliable sensory input (Beauchamp et al., 2010; Meredith and Stein, 1983, 1986b; Werner and Noppeney, 2010) and with increasing uncertainty (Körding et al., 2007). Hence a manipulation of uncertainty or stimulus reliability should affect the strength of MSI. Concordantly, studies on visual perception modulated by sound revealed that visual sensitivity for less reliable visual stimuli is improved by simultaneously presenting an irrelevant, uninformative sound (e.g. Jaekl and Harris, 2009; Noesselt et al., 2010; Van der Burg et al., 2008), and that performance increases non-linearly when target information is doubled (presenting an audiovisual target instead of just auditory or visual target; e.g. Gondan et al., 2005). Therefore it is at least conceivable that multisensory stimulation – potentially by means of its higher informational content – can aid the statistical learning mechanisms (Barakat et al., 2013) underlying the built-up of temporal expectation. However, to our knowledge there is to date little experimental support for this hypothesis.

Several studies have looked into the relationship how spatial and modality-specific attention interacts with multisensory integration but with mixed results (e.g. Alsius et al., 2005; Bertelson et al., 2000; Mozolic et al., 2008; Shore and Simic, 2005; Vroomen et al., 2001; Werkhoven et al., 2009). Only few studies investigated the interplay of cross-modal effects and temporal expectations (Bolger et al., 2013;

Jones, 2015; Lange and Röder, 2006; Menceloglu et al., 2016; Miller et al., 2012; Mühlberg et al., 2014) but they focused on other aspects than the influence of multisensory stimulation on temporal expectation in their studies. For instance, Lange and Röder (2006) used a temporal attention paradigm and tested whether knowledge about temporal regularities in one modality can be transferred to another modality (though note that no combined multisensory signals were presented). In each block, participants were instructed to attend to either short or long cue-target delays and to either auditory or tactile stimuli. Lange and Röder (2006) observed shortened response times (RT) for temporally expected targets. Remarkably, they also observed that RTs were faster for stimuli in the unattended modality when presented at expected time points — supporting the notion that knowledge about temporal regularities is stored as a supramodal representation (for similar findings see Bolger et al., 2013; Jones, 2015; Miller et al., 2012). Mühlberg et al. (2014) used a similar crossmodal transfer paradigm as Lange and Röder (2006) and tested visual-tactile stimulus combinations. Instead of attending certain foreperiod-modality combinations, participants received block-wise information about target interval and modality probabilities. More importantly, the likelihoods of occurrence (early, late) of the primary, most likely target (e.g visual) and the secondary target (e.g tactile) were manipulated (early primary target implies late secondary target and vice versa). The authors hypothesized that performance of the secondary target should be boosted at the expected time point – i.e. time when the primary target is expected – if temporal attention operates supramodally. In clear contradistinction to Lange and Röder (2006), Mühlberg et al. (2014) observed temporal expectation effects only for the primary but not for the secondary modality when presented early and RT effects for late targets suggested modality-specific mechanisms. This difference between studies with regard to modality-specific vs. supramodal temporal expectations might be due to different modality combinations, task instructions and paradigms used in the two studies.

Another recent study Menceloglu et al. (2016) investigated the interplay of temporal predictability, modality-specific attention and the congruency of visual and spoken syllables. In particular, Menceloglu and colleagues tested which of two modalities (auditory or visual) was more likely to be affected by co-stimulation in a second, unattended modality when the onset of the semantic stimuli (i.e. syllables) were temporally predictable. To this end, the authors presented auditory targets with congruent or incongruent visual stimuli and vice versa, with a short or long delay after a warning cue. When targets were temporally expected, RT slowing due to incongruent stimulation in the second modality was more pronounced for visual distractors than for auditory distractors. The authors concluded that temporal expectation are affected by (in)congruent audiovisual semantic stimuli and that the transfer between visual and auditory information is asymmetrical with increased weight of unattended visual signals during temporal expectation. Although the authors included redundant multisensory stimulation in their experiment, it remains unclear whether redundant stimulation affects temporal expectations differently than unisensory stimulation (as unisensory stimuli were not presented), and whether any interplay can be observed with non-semantic stimuli as there is some evidence that audiovisual speech stimuli favour visual inputs since lip movements precede the spoken syllable by up to 100 ms (Schroeder et al., 2008) and are thus different from simple audiovisual events. Finally, the study by Menceloglu et al. (2016) observed no interaction effects on accuracy measures. Hence, it remains unresolved whether multisensory temporal expectation effects are limited to differential response preparation, or whether multisensory temporal expectations can in fact enhance sensory representations and improve discrimination sensitivity.

We therefore aimed at investigating the interplay of temporal predictability and multisensory stimulation under varying levels of uncertainty in humans, focused on discrimination sensitivity and modified an established unisensory paradigm (Jaramillo and Zador, 2011) to this end. [1] Jaramillo and Zador (2011) had investigated auditory temporal expectation effects in rodents. In their paradigm, a sequence of random pure tones was presented during each trial. A target tone (wobble of either a low or high frequency sound) was embedded in each sequence. Rodents had to discriminate the target sound frequency. To induce temporal expectation, Jaramillo and Zador (2011) manipulated the frequency of target positions within the stimulus sequence and within blocks. In “expect early blocks”, targets were presented at early positions in the majority (85%) of trials and at late positions in remaining trials. In “expect late blocks”, the likelihood of early and late target occurrence was reversed. Comparing early targets in “early blocks” (expected targets) with early targets in “late blocks” (unexpected targets), the authors reported that rodents showed improved performance and RTs in expected (relative to unexpected) early target trials. While the authors used an ecologically valid experimental design, it remains unclear whether it can easily be applied to untrained humans, and – most importantly – how multisensory stimulation would affect temporal expectation. To test for an effect of multisensory stimulation on TE in humans, we presented sequences of auditory, visual, and audiovisual stimuli (synchronous auditory and visual sequences) in this study. As in Jaramillo and Zador (2011), temporal expectation was manipulated across blocks: in “expect early blocks”, targets were more likely to appear early within the stimulus sequence and in “expect late blocks”, targets were more likely to appear late within the stimulus sequence. Auditory and visual targets were defined by deviating frequencies (either low or high) relative to distractor stimuli

---

[1] One possibility to increase the ecological validity of experimental designs when investigating TE is to include distracting information. Here, we favoured Jaramillo & Zador’s paradigm as other ecologically valid paradigms (which also present target stimuli among distracting stimuli) have often investigated the effects of temporal expectation on the attentional blink (perception of a target following a primary target). However, investigating the attentional blink was not an aim of our study.

frequencies. We hypothesized that temporal expectation should lead to an increase in perceptual sensitivity in expected relative to unexpected trials. Furthermore, RTs should be shortened in expected trials. Finally, the effect of temporal expectation should be most pronounced for multisensory targets.

We tested these hypotheses in a series of 6 experiments. As the strength of multisensory interplay can be affected by stimulus uncertainty, we manipulated two sources of uncertainty, spatial congruency of audiovisual stimuli and target modality to investigate whether this has any further effect on TE. In particular we tested the effect of uni- vs. multisensory stimulation on temporal expectation under different levels of noise (low and high spatial and modality-specific target uncertainty) in Experiments 1-4. Spatial uncertainty was manipulated by presenting auditory and visual stimuli in close proximity (low uncertainty) versus presenting auditory stimuli via headphones (high uncertainty). Modality-specific target uncertainty was manipulated by presenting either multisensory and unisensory sequences (with the respective audiovisual or unisensory visual or auditory targets; low uncertainty) or only multisensory sequences with audiovisual or unisensory visual or auditory targets — the latter together with a non-target in the second modality (high uncertainty). [2] In the first four experiments, hazard rates were held constant and we always used the identical early and late target position out of eleven possible positions. In control experiments 5-6, we tested for the effect of different hazard rates (Exp.5) and multiple potential target positions (Exp.6) on temporal expectation. To anticipate, we observed consistent TE effects on perceptual sensitivity only in multisensory contexts with redundant audiovisual targets.

2 General Methods

The General Methods section is based on the design of Experiment 1. As all other experiments are variations of Experiment 1, only deviations from its methods are stated in the following experiment-specific methods sections below.

---

[2] Note that we use the term 'uncertainty', commonly used in the decision theory literature, to indicate that participants had to make a decision about target frequency when they couldn't predict its upcoming spatial position or modality (dependent on the experiment). Especially, in the case of spatial uncertainty other terms such as spatial coincidence or congruence could have been used. However, these terms relate more closely to a physical property of the stimulus (namely its spatial position) rather than to participants' uncertainty. Furthermore, the term uncertainty allows us to refer to both, uncertainty in space and about target modality alike.

2.1 Participants

In all experiments, participants were tested after giving signed informed consent. Volunteers who reported any neurological or psychiatric disorders or reduced and uncorrected visual acuity were excluded from the study. Participants were also excluded if they expressed a severe response bias (one response option used in more than 65% of all trials) and/or performance well below chance level in one or more conditions (accuracy below 25%). Testing of participants in each experiment continued until a total of 30 participants – given the exclusion criteria – was reached. This study was approved by the local ethics committee of the Otto-von-Guericke University, Magdeburg.

2.2 Apparatus and stimuli

The experiment was programmed using the Psychophysics Toolbox (Version 3; Brainard, 1997) and Matlab 2012b (Mathworks Inc.). Stimuli were presented on a LCD screen (22″, 120 Hz, SAMSUNG 2233RZ) with optimal timing and luminance accuracy for vision researches (Wang and Nikolić, 2011). Resolution was set to 1650x1080 pixels and the refresh rate to 60 Hz. Participants were seated in front of the monitor at a distance of 102 cm (eyes to fixation point). Responses were collected with a wireless mouse (Logitech M325). Accurate timing of stimuli (<= 1 ms) and the mouse (<= 10 ms) was confirmed with a BioSemi Active-Two EEG amplifier system connected with a microphone and photodiode. Mouse's timing precision was confirmed by analysing the jitter between the recorded onset of the click sound of the mouse button and the onset of an EEG trigger which was sent immediately after the mouse click was recognized by the OS.

Uni- or multisensory stimulus sequences (pure tones, circles filled with chequerboards, or a combination of both) were presented for each trial. Chequerboards subtended 3.07° visual angle, and were presented above the fixation cross (centre to centre distance of 2.31°). Sounds were presented from one speaker placed on top of the screen (Experiments 1, 3, and 5) at a distance of 7.06° from fixation, 4.76° from chequerboard's centre, and 3.22° from chequerboard's edge (note that this is below the minimal vertical audible angle; Strybel and Fujimoto, 2000) or via headphones (Sennheiser HD 650; Experiments 2, 4, and 6). The speaker was vertically aligned with the centre of the chequerboard stimulus. Chequerboards were presented on a dark grey background (RGB: 25.5). The fixation cross (white) was presented 2.9° above the screen's centre.

Chequerboards and sounds could serve as targets or distractors. Visual and auditory target frequencies were individually adjusted to a 75% accuracy level at the beginning of the experiment (see below Procedure; average target frequency values of all experiments are listed in Table 1). The distractor frequencies were jittered randomly between 4.6, 4.9, and 5.2 cycles per degree for chequerboards and between 2975, 3000, and 3025 Hz for sounds. Furthermore, the intensities for both target and distractor chequerboards and sounds were varied randomly throughout the stimulus sequences. The non-white checkers were jittered between 63.75, 76.5, and 89.25 RGB (average grey value of 76.5 RGB). The sound intensities were jittered between 20%, 25%, and 30% of the maximum sound intensity (average of 25% = 52 dB[A]). The sound intensity in the experiments with headphones was adjusted to match the sound intensity used for speaker experiments. The mean frequencies used are virtually identical across experiments (see Table 1; all Bayes factors - $BF_{01} >= 21.35$, indicating an approximate ratio of 25:1 in favour of the null hypothesis).

[Table 1 about here.]

2.3 Procedure

Participants were seated in a dark, sound-attenuated chamber. For each trial, a sequence consisting of 11 stimuli was presented. Stimulus duration was 100 ms and stimuli were separated by a 100 ms gap. All stimuli within a sequence were either auditory, visual, or combined auditory and visual stimuli (synchronous presentation). On multisensory trials, targets were always redundant audiovisual stimulus pairs, i.e. the stimulus frequency of both modalities was either lower or higher than distractors' frequencies. For each trial, we presented one target stimulus or target stimulus pair (audiovisual sequences) at the 3rd (onset at 400 ms, early target) or 9th position (onset at 1600 ms, late target) of the sequence (see below control Exp. 6 for a test of stimulus position on TE). Participants were instructed to maintain fixation throughout the experiment and were told that a target was present in each trial. They were required to discriminate the frequency (low or high) of the target as quickly and accurately as possible using a 2-alternative forced-choice procedure. One thumb for each response option was used (key bindings were counterbalanced across participants), and the response recording started with the onset of the first stimulus of the sequence and up to 1500 ms after sequence's offset (see Analysis section below for the definition of the response window for valid responses). Each trial ended either after the participant's response or else after 1500 ms if no response was registered, and was followed by a 200 - 400 ms inter-trial-interval (see Fig. 1 for design).

The experiment contained three sessions: an initial training session to familiarise participants with the task, a threshold determination session, and the main experiment. During training (24 trials) and threshold determination blocks (144 trials), we presented unisensory sequences only (auditory or visual). Low and high frequency, early and late occurring, and auditory and visual targets were balanced in these blocks. There were always 2 threshold determination blocks. After threshold acquisition, visual and auditory stimuli were individually adjusted to 75% accuracy for all of the aforementioned conditions. In the main experiment, separated into 6 blocks (168 trials per block, i.e. 1008 trials total), we presented all stimulus types (unisensory auditory and visual and multisensory stimuli) and modulated temporal expectation by presenting different numbers of early and late targets within blocks. A 86% likelihood of early target occurrence (always at the 3rd position) and a 14% likelihood of late targets (9th position) within the stimulus sequence was used for “expect early” blocks. In “expect late” blocks, early target occurrence was reduced to 43%. We chose this procedure instead of a complete reversal of probabilities in order to obtain a robust estimate of the performance in unexpected early trials (thereby modifying Jaramillo and Zador’s original paradigm). Expected and unexpected blocks (3 blocks each) alternated throughout the experiment, and the type of the first block was counterbalanced across participants. Importantly, participants were naive with regard to the changing likelihoods of target position across blocks.

Within each block, the number of trials was balanced with regard to sequence types and target frequencies. Additionally, the number of auditory, visual, and multisensory stimuli, early and late, and low and high targets was balanced across each quarter of blocks. Thereby, we allowed for a systematic increase of temporal expectation throughout each block. Note that although balanced, the presentation within each quarter was randomized. Trials, in which participants had failed to respond in the predefined response window, were repeated at the end of each block’s quarter without the participant’s knowledge and until they gave a response to avoid trial loss. Across participants, the maximum number of repeated trials was 113 (sum of all repeated trials across conditions for 1 experiment and participant). However, the average number of repetitions in each experiment was very small: only 0-2 trials were repeated in each condition (average of 2 - 10 repeated trials across conditions).

[Figure 1 about here.]

2.4 Analysis

In accord with previous studies (Coull and Nobre, 1998; Griffin et al., 2001; Jaramillo and Zador, 2011; Lange and Röder, 2006; Lange et al., 2003; Mühlberg et al., 2014; Nobre and Rohenkohl, 2014; Sanders, 1975), only early targets were initially used for the computation of the temporal expectation effect (i.e. higher performance for expected than unexpected targets). By comparing early targets, we were also able to rule out any effects of hazard rates on our TE effects as hazard rates (i.e. the time point of target occurrence) were identical for both types of early targets (expected vs. unexpected). Additionally, we used an orthogonal task (frequency judgement) to avoid confounds by task-presentation overlaps (e.g. temporal task). Late targets were excluded from initial analysis as they might be easily expected (see also Jaramillo and Zador, 2011; Lange and Röder, 2006; Lange et al., 2003; Mühlberg et al., 2014; Nobre and Rohenkohl, 2014), and temporal attention benefits require some degree of stimulus-related uncertainty (Lange and Röder, 2006; Lange et al., 2003; Mühlberg et al., 2014; Nobre and Rohenkohl, 2014) unlike here as late targets in our study were entirely predictable. However, for completeness we computed an additional analysis for the late targets to confirm whether temporal attention is indeed absent from wholly predictable situations.

For all analyses, trials were included with RTs ranging between 150 – 3000 ms (response window) after target onset (resulting in the average exclusion of 1.8 - 3.3% of all trials across experiments). Furthermore, performance of low and high frequency targets were collapsed, as performance was adjusted to 75% accuracy across both target types; a confirmatory analysis revealed no significant difference for low vs. high frequency targets (Bayes factor - $BF_{01} = 2.648$, indicating an approximate ratio of 3:1 in favour of the null hypothesis). To quantify the effects of modality (auditory, visual, audiovisual) and temporal expectation (expected vs. unexpected), we used a perceptual sensitivity index d′ (Green and Swets, 1966) for two-alternative forced choice (2AFC) tasks. We calculated d′ as follows:

$$d' = \sqrt{2} * z(pHit), \tag{1}$$

where z denotes the normal inverse cumulative distribution function and *pHit* denotes the proportion of correct trials in the frequency judgement task. As second measure, we used mean RTs.

Matlab 2012b (Mathworks Inc.) and IBM SPSS Statistics software (version 22.0.0.1) were used for statistical analysis. RTs and d′ were subjected to repeated measures ANOVA with factors *modality* and *Temporal Expectancy* (expected, unexpected). Post-hoc tests in all analyses were one-sided t-tests due to our one-sided hypotheses (i.e. expected targets should have higher accuracy and lower RT than unexpected targets; see Introduction). P-values were Bonferroni-corrected (pBF) to account for multiple comparisons if appropriate. We used η^2 as computed in SPSS as measure of effect size (η^2 in the range of .2 to .8 can be roughly transformed into Cohen's f by doubling the value). Note that multivariate (Pillai-Spur) instead of the univariate test results will be reported as this procedure is generally suggested for strong and frequent violations of the sphericity assumption (which holds for the first 4 experiments, especially for the RTs) because multivariate results do not rely on the sphericity assumption (Stevens, 1992). Importantly, this procedure does not inflate positive results as multivariate tests tend to be more conservative than univariate.

3 Experiments 1 - 4: TE differently affects d$'$ and RTs for uni- and multisensory events under uncertainty

3.1 Experiment 1: Methods and Results

In the first experiment, we tested whether temporal expectations can be induced with unisensory visual, auditory, and audiovisual stimulation in humans, and whether these effects differ across modalities. To this end, visual, auditory, and audiovisual stimulus sequences were employed and all presented sequences contained target stimuli (low modality-specific target uncertainty). For auditory presentation, a speaker was placed in close vicinity to the visual stimuli to maximise multisensory interplay (Stein and Meredith, 1993, see top row of Fig. 1 for a depiction of the experimental design).

[Figure 2 about here.]

In Experiment 1, we tested 34 participants. Four participants were excluded (see General Methods for exclusion criteria). 30 participants (mean age: 24.5 ± 2.7 SD; 13 women, 17 men; 2 left-handed) were used for analysis. Mean d$'$ and RTs are displayed in the top panel of Fig. 2. Repeated-measures ANOVAs revealed that participants' perceptual sensitivity was enhanced (main effect of TE; d$'$ of 1.203 and 1.032, respectively; F(1,29) = 28.237, $p < .001$, $\eta^2 = .493$) and RTs were faster (RT of 1543.47 ms a 1667.71 ms, respectively; F(1,29) = 33.265, $p < .001$, $\eta^2 = .534$) for expected rather than unexpected target stimuli. Furthermore, d$'$ was increased for audiovisual compared to auditory and visual targets (main effect of modality: F(2,28) = 8.939, $p = .001$, $\eta^2 = .39$). This beneficial effect was also present for RTs, with participants responding faster on multisensory target trials (F(2,28) = 11.641, $p < .001$, $\eta^2 = .454$). The interactions for d$'$ (F(2,28) = .648, $p < .648$) failed to reach significance. For RT we found that TE effects were smaller and less significant in the visual condition compared to auditory and audiovisual conditions (F(2,28) = 4.53, $p = .02$, $\eta^2 = .244$). All post-hoc test results can be found in

Tables 2 and 3.

[Table 2 about here.]

[Table 3 about here.]

3.2 Experiment 2: Methods and Results

In Experiment 1, we maximised the effects of multisensory context by presenting visual and auditory stimuli in close proximity. In Experiment 2, we tested whether audiovisual spatial incongruence affects temporal expectation by presenting auditory stimuli via headphones, i.e. from a spatial location different from the visual stimulation. Previous neurophysiological studies on audiovisual interplay had suggested that MSI is maximal if audiovisual stimulation have a spatially congruent source. However, some studies on temporal processing suggest that spatial congruence is less relevant in temporal and identification tasks (Diederich and Colonius, 2004; Doyle and Snowden, 2001; Jones and Jarick, 2006; Kadunce et al., 2001; Keetels and Vroomen, 2007; Noesselt et al., 2005; Recanzone, 2003; Spence, 2013; Stein et al., 1996; Van der Burg et al., 2008; Vroomen and Keetels, 2006), and many studies on audiovisual interplay have in fact used headphones (Bischoff et al., 2007; Di Luca et al., 2009; Diederich and Colonius, 2004; Fujisaki and Nishida, 2007; Keuss et al., 1990; Roach et al., 2006; Soto-Faraco et al., 2005; Wada et al., 2003). Here, with auditory stimuli presented via headphones, the spatial position of the upcoming stimulus sequence was unpredictable (frontal screen and/or headphone; high spatial uncertainty) as compared to Experiment 1 (always frontal and thereby always predictable; low spatial uncertainty). Another way of inducing spatial uncertainty would have been to use several speaker position. However, this procedure might have induced the ventriloquist illusion in some participants and would have unduly increased the number of experimental conditions. We therefore adopted a different approach and used headphones instead. All other methods and analyses used were identical to the General Methods/Experiment 1. The experimental design is depicted in the top row of Fig. 1.

We tested an independent sample of 33 naive participants. Three participants were excluded (see General Methods for exclusion criteria). Data from 30 participants (mean age: 23.1 ± 3.4 SD; 18 women, 12 men; all right-handed) were used for analysis.

The bottom panel of Fig. 2 displays mean d' and RT values. Again, the repeated measures ANOVA of perceptual sensitivity revealed significant main effects of expectancy and modality; importantly, the interaction was also significant. In particular, d' was larger for expected than unexpected stimuli (1.173

and .931, respectively; F(1,29) = 24.696, $p < .001$, $\eta^2 = .46$) and larger for multi- than unisensory target stimuli (F(2, 28) = 21.192, $p < .001$, $\eta^2 = .602$). Additionally, enhanced d′ values were only found for auditory and audiovisual targets but not for visual ones (F(2,28) = 8.413, $p = .001$, $\eta^2 = .375$, see Table 2 for details of post-hoc t-tests). For RT, the pattern of results was almost identical: Responses were faster when stimuli were expected (1655.108 ms vs. 1791.894 ms; F(1,29) = 20.64, $p < .001$, $\eta^2 = .416$) and faster when stimuli were multisensory (F(2, 28) = 18.733, $p < .001$, $\eta^2 = .572$). Again, we found that TE effects – like in Experiment 1 – were smaller and less significant in the visual condition compared to auditory and audio-visual stimuli (F(2,28) = 8.415, $p = .001$, $\eta^2 = .375$, see Table 3 for details).

3.3 Experiment 3: Methods and Results

One potential explanation for the pattern of results observed in Experiment 2 could be that participants preferentially focused their attention on only one modality. This could have been the auditory modality as an effect of TE was present for unisensory auditory sequences (and audiovisual sequences) while absent in the visual modality. Thus, in the multisensory context, the TE effect might exclusively have been driven by the auditory modality. In accord, many previous studies have reported an auditory dominance in temporal tasks (Bertelson and Aschersleben, 1998; Fendrich and Corballis, 2001; Guttman et al., 2005; King and Nelken, 2009; Nobre and Rohenkohl, 2014; Recanzone, 2003; Repp and Penel, 2002; Shipley, 1964; Wada et al., 2003; Welch et al., 1986). To investigate whether modality-specific attention had an influence on the previous results, target occurrence in a particular modality (uni- and multisensory targets) was manipulated in Experiments 3 and 4. To this end, we presented only audiovisual sequences, BUT targets were as before either unisensory (auditory or visual) or redundant multisensory targets (high target uncertainty). Thus, to perform the task, participants were forced to equally monitor both modalities on each trial to be able to detect the target. The number of pure auditory, pure visual and multisensory targets was again balanced (33 percent each). As in Experiment 1, a speaker was used for auditory stimulation (low spatial uncertainty). All other methods and analyses used are identical to the General Methods. The experimental paradigm is depicted in the middle row of Fig. 1.

We tested an independent sample of 41 naive participants. Eleven participants were excluded (see General Methods for exclusion criteria). 30 participants (mean age: 24.3 ± 3.6 SD; 21 women, 9 men; 4 left-handed) were used for analysis. Note that the higher number of excluded participants could not be attributed to a specific stimulus condition, but rather to a higher number of inexperienced participants due to the beginning of a new term. Concordantly, half of the excluded participants showed low performance in auditory and half in visual conditions. Given similar average performance between Experiments 3 and 4, we suspect that the excluded individuals in Experiment 3 had to invest more effort to perform the task and did not succeed in some conditions.

The results are displayed in the top row of Fig. 3 and the repeated measures ANOVAs with the main effects of expectancy and target modality corroborated the results of Experiment 1. Main effects for both measures (d′ and RT) reached significance. In particular, responses for expected stimuli were more accurate (.893 vs. .754; F(1,29) = 17.976, $p < .001$, $\eta^2 = .383$) and faster (1647.883 ms and 1748.324 ms; F(1,29) = 21.223, $p < .001$, $\eta^2 = .423$). Furthermore, performance in the multisensory target condition exceeded performance in the auditory and visual conditions (d′ (F(2,28) = 53.543, $p < .001$, $\eta^2 = .793$); RT (F(2,28) = 57.935, $p < .001$, $\eta^2 = .805$). As in Experiment 1, the interaction term did not reach significance for d′ (F(2,28) = .352, $p = .706$), and additionally not for RT (F(2,28) = .729, $p = .492$). This pattern of results suggests that the effects found in Experiment 2 cannot be solely attributed to modality-specific attention to the auditory domain, as the multisensory TE effect remains the same and is not attenuated, if participants successfully focus on both modalities (as indexed by unisensory auditory and visual TE effects in Experiment 3).

[Figure 3 about here.]

3.4 Experiment 4: Methods and Results

In the last two experiments (Experiments 2 and 3), we tested if introducing either spatial or modality-specific uncertainty in isolation would affect temporal expectations in multisensory contexts, but failed to find any effects. In Experiment 4, we combined both uncertainties and tested whether temporal expectation is affected by high spatial plus high target uncertainty conditions. To this end, we presented only audiovisual sequences with unisensory and multisensory targets (high modality-specific target uncertainty) and used headphones (high spatial uncertainty). All other methods and analyses are identical to the General Methods. The experimental paradigm is depicted in the middle row of Fig. 1.

Again, 33 naive participants were tested and three of them were excluded (see General Methods for exclusion details). 30 participants (mean age: 23.9 ± 3.7 SD; 22 women, 8 men; 2 left-handed) were used for analysis.

The results are displayed in the bottom row of Fig. 3. As with all previous experiments, expected targets led to higher d′ values (.921 vs. .835; F(1,29) = 6.23, $p = .018$, $\eta^2 = .177$) and faster RTs (1609.76 ms vs. 1689.055 ms; F(1,29) = 16.723, $p < .001$, $\eta^2 = .366$). d′ was increased for multi- compared to unisensory stimuli (F(2,28) = 34.113, $p < .001$, $\eta^2 = .709$) and responses were also faster (F(2,28) = 35.467, $p < .001$, $\eta^2 = .717$). Furthermore, we found an interaction effect for d′, and this time the temporal expectation effect was only carried by multisensory stimuli (F(2,28) = 5.339, $p = .011$, $\eta^2 = .276$) — with both unisensory visual and auditory targets expressing a reduced effect of temporal expectancy (post-hoc test results can be found in Tables 2). The interaction for RTs was not significant (F(1,28) = 1.664, $p = .208$). Together, the pattern of results suggest that with increased level of uncertainty, TE effects for multisensory contexts remain stable, while they are reduced if less information is available.

4 Control Experiment 5-6: TE effects scale with early-late target ratio but are unaffected by specific target position

4.1 Experiment 5: Methods and Results

The previous experiments provided robust evidence that temporal attention was directed to (expected) or away from (unexpected) particular instants in time. However, in the previous experiments, we only used one predefined ratio of early and late target occurrences. This experimental design does not rule out that temporal attention in our paradigm operates on a rather global level and just computes early vs. late likelihood on a coarse scale. If, on the other hand, temporal attention is based on a fine-grained analysis of probabilities, we would predict that performance systematically decreases when the likelihood of early targets decreases. To this end, we conducted an experiment in which we varied the likelihood of early targets across blocks. As Experiments 1 through 4 revealed robust TE effects for audiovisual stimuli with audiovisual targets, we restricted the following experiments to audiovisual stimuli. Note that we still varied the spatial certainty (speakers: Exp. 5 ; headphones: Exp. 6) to confirm that the effects in purely audiovisual context are – as in Exp. 1-4 – unaffected by spatial proximity.

In Experiment 5, we tested an independent sample of 32 naive participants. Two participants were excluded. 30 participants (mean age: 21.7 ± 2.9 SD; 20 women, 10 men; 6 left–handed) were used for analysis. The stimulation protocol was identical to the General Methods except for the following changes. In the main experiment, we presented only audiovisual sequences with audiovisual targets.

Instead of presenting 2 block types (expect early and expect late), we presented 6 different block types (168 trials each) with varying early-late target ratios. The probability of early targets was set to 14%, 29%, 43%, 57%, 71%, or 86%. The probability of late targets was set to 100% minus the probability of early targets. We balanced the early target probability of the first block across participants and randomized the order of the remaining probabilities. RTs and d′ were analysed with 1-factorial repeated measures ANOVA with factor *early target probability* (14% to 86% early targets).

Average d′ and RT values are displayed in the top panel of Fig. 4. The results show an almost perfect linear trend (see Fig. 4). d′ systematically decreased with decreasing early target probability (F(5,25) = 7.102, $p < .001$, $\eta^2 = .587$; evidence for linear relationship: F(1,29) = 35.429, $p < .001$, $\eta^2 = .55$) while RTs systematically increased (F(5,25) = 8.944, $p < .001$, $\eta^2 = .641$; evidence for linear relationship: F(1,29) = 40.564, $p < .001$, $\eta^2 = .583$). Hence, the pattern of results strongly suggests that TE is based on a fine-grained analysis of the probability of early target presentations.

[Figure 4 about here.]

4.2 Experiment 6: Methods and Results

In all previous experiments, only a single early and one late target position were used. However, Jaramillo and Zador (2011) reported effects of temporal expectancy for unisensory auditory streams using 2 adjacent target positions (3rd and 4th position). This indicates that temporal expectancy does not necessarily foster a single point in time but may be spanned over a larger time period. In our last experiment, we jittered the early target position to investigate the effect of target position on temporal expectancy. If temporal expectancy operates over a larger time window, we should see similar temporal expectancy effects across target positions. However, if temporal expectancy operates only in a narrow time window, temporal expectancy effects should either be absent or largest for the centre of the temporal positions.

We tested an independent sample of 34 naive participants. Four participants were excluded (see General Methods for exclusion criteria). 30 participants (mean age: 23.5 ± 3.5 SD; 21 women, 9 men; 5 left-handed) were used for analysis. The stimulation protocol and analyses were identical to the General Methods except for the following changes. In the main experiment, we presented only audiovisual sequences with audiovisual targets. Importantly, targets could appear in the sequence at positions 2, 3, or 4 (early positions) and 8, 9, or 10 (late positions). We balanced the number of trials of each position across blocks' quarters. Furthermore, the trial number was balanced across positions within each position type (early and late positions). Note, that for statistical analyses, the factors *temporal expectancy* and *target position* (position 2, 3, or 4) were used.

Average d′ and RT values are displayed in the bottom panel of Fig. 4. The only d′ effects were found for the factor *temporal expectation*: values for expected stimuli were higher than for unexpected stimuli (1.511 and 1.284, respectively; F(1,29) = 17.068, $p < .001$, $\eta^2 = .37$). d′ did not differ for *target position* (F(2,28) = 2.207, $p = .129$) and we found no interaction (F(2,28) = .681, $p = .514$). RTs were also different for *temporal expectation*: values for expected stimuli were lower than for unexpected stimuli (1386.518 ms and 1555.215 ms, respectively; F(1,29) = 70.957, $p < .001$, $\eta^2 = .71$). Again we found no interaction (F(2,28) = 2.089, $p = .143$) but a significant main effect of *target position* (F(2,28) = 20.575, $p < .001$, $\eta^2 = .595$). Post-hoc t-tests indicated that responses times were faster when target position increased (see Table 3).

5 Summary late target results

To confirm that temporal expectancy is only relevant if there is any uncertainty with regard to target presentation, we also analysed the late targets. Note that late targets are always expected whenever an early target is not presented and perceived (Coull and Nobre, 1998; Griffin et al., 2001; Jaramillo and Zador, 2011; Lange and Röder, 2006; Lange et al., 2003; Mühlberg et al., 2014; Nobre and Rohenkohl, 2014; Sanders, 1975). All results and plots can be found in the supplementary material (Supplement_LateTargets.pdf; url: `osf.io/4m26y`; Ball, 2017). Here we highlight only the significant findings.

In Experiments 1-4 we found neither an TE effect nor and interaction of TE and modality for late target d′ and RT. In all 4 Experiments, we found an effect of modality which was due to faster and more accurate responses in the audio-visual condition compared to the auditory and visual conditions. Thus, although the TE effect vanished for late trials, the multisensory interplay still enhanced performance in general.

In Control-Experiment 5, we again found no effects for RT and d′. However, in Experiment 6, late target d′ was influenced by the position of the target with highest performance at the 9th position. Here, late target positions varied between the 8th to 10 th position – hence temporal predictability was decreased in this case – which resulted in an position effect for the late targets. As for early targets, RT decreased with increasing target position. There was also an interaction of Position and TE, indicating that TE might have had an effect as long as the target was presented at the 9th position. A closer look unveiled that the TE effect for the 8th position was reversed (unexpected trials faster then expected) which might be due to the lower number of trials for the unexpected late targets. Together, the results from the first five experiments suggest that TE require at least some temporal unpredictability to occur, in accord with earlier studies (Coull and Nobre, 1998; Griffin et al., 2001; Jaramillo and Zador, 2011; Lange and Röder, 2006; Lange et al., 2003; Mühlberg et al., 2014; Nobre and Rohenkohl, 2014; Sanders, 1975) and late target data of Experiment 6. Accordingly, TE effects for late targets can only be observed if temporal predictability of late targets is reduced (for example by jittering target position, as we did in Experiment 6 or by introducing catch trials as in Mühlberg et al., 2014).

5.1 Re-analysis of Experiments 1 - 6

On reviewers' request, we re-analysed the data to test whether the choice of our response time restriction could affect the pattern of results. To this end, we used only trials in which response times were in the range of $RT_{mean} \pm 2 * STD$. The results were virtually identical to our original analyses and can be found in the supplementary material (Supplement_AlternativeRTRestriction.pdf; url: `osf.io/4m26y`; Ball, 2017). A minor difference was a slightly less significant main effect of factor *temporal expectation* in Experiment 6 (p = .056).

6 General Discussion

In this study, we tested whether participants are able to built up temporal expectations (TE) from temporal regularities hidden in the stimulus stream, whether TE is modulated by audiovisual stimulation, and whether target and spatial uncertainty would further affect TE in multisensory contexts. In all experiments, participants were more accurate and faster in discriminating the frequency of expected relative to unexpected targets, as predicted. Furthermore, we found a benefit for multisensory over unisensory stimulation irrespective of temporal regularities. Most importantly, multisensory stimulation had a protective effect on perceptual sensitivity based on temporal regularities when tasks became more difficult and spatial and target reliability decreased. Finally, results from control experiments indicate that TE operates by weighting the actual probabilities of target occurrence at a given time and that temporal attention window covered multiple possible target positions (> 500 ms; for further information see below).

Our consistent finding in Experiments 1-3 of enhanced processing of auditory targets – based on temporal regularities within stimulus sequences – translates previous work in non-human animals (Jaramillo and Zador, 2011) and demonstrates that Jaramillo and Zadors's paradigm can be successfully applied to study the effect of auditory temporal expectation in humans. Importantly, temporal expectation effects were also observed for visual stimuli, hence are not restricted to the auditory modality. Our findings are in line with previous studies on temporal expectations in relatively simple unisensory contexts (Correa et al., 2004; Coull and Nobre, 1998; Cravo et al., 2013; Griffin et al., 2001, 2002; Jepma et al., 2012; Jones et al., 2002; Lange and Röder, 2006; Lange et al., 2003; Mathewson et al., 2010; Miniussi et al., 1999; Niemi and Näätänen, 1981; Rohenkohl et al., 2012, 2014; Rolke and Hofmann, 2007; Sanabria et al., 2011; Westheimer and Ley, 1996), which also reported enhanced processing of expected stimuli. In addition, our study corroborates rarely investigated topics by showing that temporal expectations can be studied in more complex and ecologically valid paradigms (Jaramillo and Zador, 2011; Shen and Alain, 2011, 2012) and in the absence of prior knowledge about the manipulation of temporal regularities (in line with findings by Beck et al., 2014).

Most importantly, the most robust TE effects were found for multisensory stimulation with redundant multisensory target stimuli, extending previous unisensory research on TE (for an overview see Nobre and Rohenkohl, 2014). Our results also extend our understanding of multisensory interplay. In particular, previous crossmodal TE research focused solely on the transfer of TE across different modalities (i.e. can TE be transfered from vision or audition to touch, and vice versa; Bolger et al., 2013; Jones, 2015; Lange and Röder, 2006; Miller et al., 2012; Mühlberg et al., 2014), and the weighting of visual and auditory inputs in a purely multisensory speech paradigm (no unisensory stimulation was applied; Menceloglu et al., 2016). While these previous studies have important implications (see below), none of these studies addressed the critical question of whether redundant multisensory stimulation – which is known to enhance performance via enhanced sensory representations, as indicated by an increase in d' or accuracy (Alais and Burr, 2004; Driver and Noesselt, 2008; Forster et al., 2002; Gondan et al., 2005; Jaekl and Harris, 2009; Noesselt et al., 2010; Parise et al., 2012; Sinnett et al., 2008; Stevenson et al., 2014; Talsma et al., 2007; Van der Burg et al., 2008) – also interacts with statistical learning based on temporal regularities.

We are the first to show that TE interacts with target modality (auditory vs. visual vs. audio-visual) in experiments with increased levels of uncertainty. In Experiment 1, without uncertainty, TE effects occurred in unisensory as well as multisensory conditions. In Experiment 2, TE effects were reduced for unisensory visual stimulus sequences when introducing spatial uncertainty by presenting visual and auditory stimuli from different position (high spatial uncertainty). However, it could be argued that participants simply focused on the auditory stream as the auditory modality provides a better temporal resolution, and auditory stimuli are thus better suited for the extraction of temporal regularities and may dominate in temporal tasks (Bertelson and Aschersleben, 1998; Fendrich and Corballis, 2001; Guttman et al., 2005; King and Nelken, 2009; Lechelt, 1975; Nobre and Rohenkohl, 2014; Philippi et al., 2008; Recanzone, 2003; Repp and Penel, 2002; Shipley, 1964; Wada et al., 2003; Welch et al., 1986). If such a strategy would have always been chosen, we would expect to observe reduced TE effects for visual targets in Experiment 3, in which visual, auditory or audiovisual targets were presented in audiovisual streams (high target uncertainty). In contrast, a general TE effect was observed, rendering an explanation based on attention to the auditory domain less likely. In accord, the results from Experiment 4 do not support an explanation based on modality-specific attention; there, both high target and spatial uncertainty were introduced. If spatial uncertainty would have led to a focusing of the auditory domain, we would have expected a pattern of results similar to Experiment 2, i.e. reduced TE effects for the visual targets. In contrast, in Experiment 4, both visual and auditory targets expressed reduced TE effects. Only for audiovisual targets was a TE effect on perceptual sensitivity still present. This pattern of results suggests that the effects of multisensory interplay may help to preserve statistical learning of temporal regularities in noisy environments. More specifically, participants might utilize unsupervised learning strategies as they were naive about temporal regularities, upcoming target modalities and spatial position. While target modality and spatial position were rendered unpredictable by design (especially in the high uncertainty experiments), temporal regularities underwent statistical changes across blocks (more or less early targets). The higher informational content of the redundant multisensory target allowed participants to perceive targets more easily (more clearly or more often), and thereby allowing them to make inferences about the most likely time point of target occurrence. In turn, participants were able to create some form of summary statistics within blocks (when do targets occur more often) to guide their attention in time. We propose that this statistical learning is reflected by the temporal expectation effects found in our study.

Control Experiments 5 and 6 further corroborated this notion. In both experiments, we again replicated the robust TE effects, even under high spatial uncertainty (Exp. 6). In addition, the last two experiments provided further in-depth evidence how temporal attention operates in our paradigm. Experiment 5 revealed that performance decreased linearly with decreasing early target probability. Hence, temporal attention acts on a rather fine-grained level as the ratio of early and late targets shaped performance gradually. This was true even though the early-late likelihoods changed with beginning of each block. Thus, temporal attention is not only capable of a fine-grained analysis of temporal regularities, it can also adapt rather quickly to new situations. This finding is in good agreement with findings from cueing studies in which temporal attention has to be adapted for each trial (Correa et al., 2004; Coull and Nobre, 1998; Griffin et al., 2001, 2002; Jepma et al., 2012; Miniussi et al., 1999) and with studies using explicit temporal tasks (Akdoğan and Balcı, 2016; Balci et al., 2009; Balcı et al., 2011; Bogacz et al., 2006; Çavdaroğlu et al., 2014; Çoşkun et al., 2015; Freestone et al., 2015).

In addition, the results of Experiment 6 provide further insights into the time interval on which temporal attention operates. Earlier studies had reported that the temporal estimates rely heavily on exogenous (paradigm induced) and endogenous (participant specific) uncertainties (Akdoğan and Balcı, 2016; Balci et al., 2009; Balcı et al., 2011; Bogacz et al., 2006; Çavdaroğlu et al., 2014; Çoşkun et al., 2015; Freestone et al., 2015). Thus, the precision with which temporal regularities can be extracted and used is variable. There are at least 4 scenarios that can explain our findings. In the first, the focus of temporal attention is divided and operates in small time windows around each stimulus presentation (Fig. 5 A). In the second scenario, the temporal attention window is broadened and spans across multiple stimuli. Here, stimuli are attended equally and the on- and offsets of the window could either be smooth (Fig. 5 B1) or sharp (rectangular function, Fig. 5 B2). In the third scenario (Fig. 5 C), the attentional window is broadened but stimuli are not attended equally. Here the average stimulus position (i.e. the 3rd position which is flanked by the 2nd and 4th) is attended more than the flanker positions (which are attended equally). Finally, in the fourth scenario, temporal attention operates differently across stimuli. Again, temporal attention rises until it peaks for the mean target duration (3rd position) but attention for the last stimulus position falls below all others (Fig. 5 D). By visually inspecting the d-prime data in Experiment 6, the overall performance trajectory for early and late targets (see Fig. 4 and supplementary material late targets, first figure, bottom row) favours the fourth scenario (skewed Gaussian distribution). While performance is highest for the middle positions (3rd and 9th), it is lower for the first (2nd and 8th), and even lower for the last positions (4th and 10th). Thus, participants seem to pool information of target occurrence over a larger time interval, and shift attention to the middle position. This might be attributed to endogenous timing uncertainties, as stimuli are presented in close succession and might not be easily perceived as distinct events. Furthermore, upholding attention is resource demanding, so to optimize resources allocation, attention would be distributed asymmetrically. If this suggestion is true, paying mainly attention to 2nd position would result in a release of attention and therefore, a drop for the 3rd and 4th position. If one would mainly attend the 4th position, attention would either have to be uphold (resource demanding), or the window would be shifted so that relevant positions (2nd) would be ignored and irrelevant positions (5th) would be attended. Hence, the optimal trade-off between resource allocation and performance increase is to attend the average target onset time while focussing less on the flanking onset times. Thereby, effects of endogenous timing uncertainty would also be reduced as the timing uncertainty would be centred in the middle of the overall target interval. In fact, the idea of broader time window, and estimation of the most likely target position to reduced effects of timing uncertainty (i.e. a decrease of performance) is in line with studies investigating optimal behaviour in temporal studies (Akdoğan and Balcı, 2016; Balci et al., 2009; Balcı et al., 2011; Bogacz et al., 2006; Çavdaroğlu et al., 2014; Çoşkun et al., 2015; Freestone et al., 2015).

We also found that participants in Experiment 6 responded slower when targets occurred at the 2nd position. This might have been due to a response strategy, as participants apparently tended to withhold their response until the end of the sequence. Hence, “target to response” times would be slower for earlier positions in the sequence. Although this might have been the general strategy used by participants, it did not affect or interact with the temporal expectation effects, strongly suggesting that participants always responded slower when targets were unexpected — irrespective of target position (note that expected and unexpected early targets used for analyses always occurred at the same target positions). This pattern of results indicates that temporal expectation effects in multisensory contexts are, to a large extent, unaffected by response strategies.

[Figure 5 about here.]

While our results indicate that discrimination sensitivity is more sensitive to capture the cognitive processes underlying TE, previous research on TE had often relied on differences in RT to characterize these perceptual and cognitive processes (for an recent overview see Nobre and Rohenkohl, 2014). However, a modulation of RT could reflect differential motor preparation, while a difference in discrimination sensitivity should reflect enhanced sensory representations (Green and Swets, 1966; Prinzmetal et al., 2005; van Ede et al., 2012). In our studies the pattern of results differed for the two behavioural measures (i.e. perceptual sensitivity and RT). In particular, the critical interaction effect of modality and TE was only observed for the sensitivity measure, but not for the RT measures indicating that multisensory interplay allowed participants to extract temporal regularities in noisy environments. The selectivity of the sensitivity measure for the interplay of multisensory stimulation and TE extends previous studies on multisensory interplay (e.g. Jaekl and Harris, 2009; Noesselt et al., 2010) and suggests that sensory representation were indeed altered. Thereby, our results significantly extend the one previous study on the interaction of TE and MSI (Menceloglu et al., 2016), as they only reported an interaction of TE and MSI for RTs. In contrast, the RT decrease in our experiments for expected stimuli was observed for all conditions and might therefore reflect enhanced response preparation for expected stimuli regardless of their particular modality or modality combination (see below Section 6.1. for further discussion of potential response strategies). This difference in RTs between our study and the study by Menceloglu and colleagues might be due to the fact, that participants withhold their response until the end of the stimulus sequence in our paradigm, thereby reducing differential effects. If this is the case, our data does not support a generalizable mechanisms proposed by Menecoglu et al. It might rather be that response facilitation of visual stimuli occurs in cross-modal TE paradigms whenever a rather “simple” paradigm is used. There, the detriments of the visual condition could be compensated by TE to increase overall performance. However, this might not be possible when visual targets are not easily identifiable as in our experiments.

Moreover, auditory and visual stimulation may differ in their ability to aid participants to extract temporal regularities. Several studies reported that auditory perception outperforms visual perception in temporal tasks which led to the notion of auditory dominance for temporal processing (Bertelson and Aschersleben, 1998; Fendrich and Corballis, 2001; Guttman et al., 2005; King and Nelken, 2009; Nobre and Rohenkohl, 2014; Recanzone, 2003; Repp and Penel, 2002; Shipley, 1964; Wada et al., 2003; Welch et al., 1986), as auditory perception has higher temporal resolution and might therefore be in a privileged position to extract temporal regularities. This auditory dominance is not restricted to the implicit extraction of temporal regularities but extends to situations in which durations (Akdoğan and Balcı, 2016; Balcı et al., 2011; Bogacz et al., 2006; Freestone et al., 2015) or even the number of incidents (e.g. how many flashes have been presented) has to be judged (Lechelt, 1975; Philippi et al., 2008) and has been more recently conceptualised by computational models using Bayesian approaches (Maiworm and Röder, 2011).

The aforementioned studies as well as our results question the idea that TE preferentially modulates auditory processing by visual information (Menceloglu et al., 2016). Recall that Menceloglu and colleagues presented auditory targets with congruent or incongruent visual stimuli, and visual targets with congruent or incongruent auditory stimuli in a temporal attention task. When targets were expected, RT slowing due to incongruent stimulation in the second modality was more pronounced for visual distractors than for auditory distractors. The authors concluded that temporal expectation increases the weight of visual signals, thus, temporal expectation would favour performance in the visual condition.

Furthermore, they showed that TE decreases the impact of auditory distractors on visual performance and increases the impact of visual distractors on auditory performance. A result of such findings would be that performance in the auditory condition is decreased compared to the visual condition, and that TE effects are stronger or at least more robust in the visual condition especially under high target uncertainty (i.e. incongruent condition). One could argue that high target uncertainty in our Experiments 3 and 4 resemble at least to some extent the incongruent condition (e.g. auditory target with incongruent visual target) in Menceloglu et al.'s experiment. Here, targets were not always redundant and sometimes flanked by a non-target (distractor) in the second modality. However, our results indicate that the visual condition was not favoured in these Experiments. If it would have been, we should have found higher performance and/or TE effects in the visual condition in Experiments 3 and 4. In general (across all experiments), our results revealed overall decreased performance in the visual condition relative to auditory and audiovisual conditions, and less incidences of TE. Thus, our findings are in clear contradistinction to Menceloglu et al. but are in line with findings implicating auditory dominance in temporal tasks. However, to reconcile these apparently contradictory findings, it could be argued that semantic audiovisual stimulation as used by Menceloglu represents a special case of audiovisual integration (Doehrmann and Naumer, 2008), and thus interacts differently with temporal regularities.

6.1 Is behaviour in our temporal expectation task optimal?

More complex experimental designs, as used here, usually manipulate exogenous uncertainty. However, endogenous uncertainty (i.e. noisy internal representations of external stimulus probabilities) might also have impacted our results. Several studies reported that for explicit timing tasks performance is close to optimal in line with statistical decision theories (e.g. Balcı et al., 2011; Bogacz et al., 2006; Çoşkun et al., 2015; Freestone et al., 2015). This indicates that participants take into account uncertainties introduced by the experimental design (exogenous; e.g. likelihood of target position and pay-offs) but also intrinsic uncertainties (endogenous) such as the precision of temporal judgements. In these human and animal studies RT tasks were often used with and without the risk to loose rewards when responses were too fast or too slow (see e.g. Çoşkun et al., 2015). In our experiment, we asked participants to respond as accurately and quickly as possible. However, our RT results strongly suggest that instead of making speeded responses, participants relied on choice responses to increase their performance. Mean RTs were situated around 1600 ms after early target presentation which amounts to a button press around 2000 ms after sequence onset which is almost the end of the sequence. Furthermore, a post-hoc questionnaire supports the notion that participants used this strategy; as almost all participants stated that they withhold their response till the end of the sequence to confirm their percept and response choice. Such strategy might be often been chosen when a task is difficult (see Berkay et al., 2016, for suboptimal performance under noise in rats) and response speed is neither punished nor enforced but is clearly suboptimal if insufficient response speed would be linked to detrimental effects (such as the loss of reward or, in more ecological context, an accident in traffic due to slow reaction).

Another suboptimal strategy we observed in our experiments is to shift attention to instances in time when target likelihood is maximal. The best strategy one could choose in the current experiment to maximise task performance would be to sequentially sample each stimulus and to make a decision when evidence of all stimuli of a particular sequence is accumulated. Recall, that participants had to determine the target on the basis that it is different from all other stimuli (distractors). However, the aforementioned strategy would lead to diminished TE effects as temporal information would become irrelevant when using an unbiased sequential sampling strategy. In contrast, we observed TE effects for early targets strongly suggesting that participants shifted their attentional focus to the later position in late target blocks – which is in principle suboptimal. This pattern was most prominent in Experiment 5, in which we observed an decrease in accuracy for early targets which scaled with the ratio of early vs. late target likelihoods. Given that the late target always occurred after the early target, there was no obvious need to shift attention in the first place as it only decreases performance. Our data suggests that it is unlikely that participants actively sampled the individual stimuli but created some form of intrinsic, implicit knowledge about the time point at which target likelihood is highest. This time point might be subject to endogenous timing uncertainty (Akdoğan and Balcı, 2016; Balci et al., 2009; Balcı et al., 2011; Bogacz et al., 2006; Çavdaroğlu et al., 2014; Çoşkun et al., 2015; Freestone et al., 2015), which might lead to a temporal focus that can encompass multiple items, as observed in Experiment 6. Additionally, the reference time given by the experimental design might shift with different proportions of early and late targets (Çoşkun et al., 2015), at least in the case of early targets. Thus, presenting a balanced amount of early and late targets might shift perceived target timing to the middle of the sequence and potentially broadens the perceived temporal window of target occurrence, while presenting more late targets shifts perceived target timing to the end of the sequence. In cases of high uncertainty, as in experiments incorporating distractor sequences, and without active engagement (e.g. sequential sampling) and knowledge about the temporal manipulation, participants seem to integrate and use as much information as is provided by the experimental design to optimize their performance.

Such optimization might affect the speed with which evidence about target presence is accumulated. TE could e.g. prepare the neural system for incoming information which in turn would increase perceptual sensitivity, an idea supported by our data. If the system is prepared, evidence can be accumulated faster. Given that we also found a general increase in performance and decrease in RT for multisensory compared to unisensory stimuli, it is likely that multisensory target evidence is accumulated faster. This assumption could be tested by means of drift-diffusion models which have successfully been applied to explain performance in temporal task (e.g Akdoğan and Balcı, 2017; Balcı et al., 2011; Balcı and Simen, 2014). If evidence accumulation is fastest for expected and multisensory trials, the drift rate (parameter representing evidence accumulation) should be highest. However, the implementation of such model is beyond the scope of this paper and future research is needed which could model these effects and quantify by how much our results deviate from the optimal performance of an ideal observer.

While our current results are in line with some previous studies on optimal performance, it should be noted that our task regimes are not directly comparable to previous studies investigating optimal performance (Akdoğan and Balcı, 2016; Balci et al., 2009; Balcı et al., 2011; Bogacz et al., 2006; Çavdaroğlu et al., 2014; Çoşkun et al., 2015; Freestone et al., 2015): Here we did not reward or punish participants based on their responses. This might lead to completely different outcomes as the aforementioned studies usually defined optimal behaviour on the basis of speeded RT. Furthermore, we used only an implicit timing task. Remarkably, participants appeared to have been oblivious of the temporal manipulation at the beginning of the experiment, and most of them were oblivious even at the end of the experiments. To assess the participants' explicit knowledge of temporal regularities, we asked all participants after the experiment ended whether they noticed any regularities in general and if they negated that, we enquired whether they noticed any regularities about target position and further if they could specify this position. Out of the 180 participants, only 65 noticed any position regularity (13 stated regularities immediately). 41 participants could identify the second or third position as target position while the remaining stated that targets "occurred mostly early" or "mostly early and late". Out of the 65 participants, 15 made their statements specifically for the auditory but not visual stimulus, again supporting the notion that auditory information might be the more reliable source in temporal tasks. Given that these 65 participants were randomly distributed across experiments, TE effects shown in our study seem to be independent of explicit knowledge about the target position. Future research may use a trial-based test procedure (e.g. asking to judge the target position on every trial) to characterize the influence of explicit knowledge on TE. Nevertheless, the TE effects observed here seem not to be based on active counting or voluntary shifts of attention to more likely target intervals, suggesting that participants performed primarily the frequency discrimination task which was orthogonal to the manipulation of temporal context. Thus, we addressed the question whether participants made optimal use of temporal regularities to improve their discrimination performance, rather than investigating optimal performance in a temporal task (Akdoğan and Balcı, 2016; Balci et al., 2009; Balcı et al., 2011; Bogacz et al., 2006; Çavdaroğlu et al., 2014; Çoşkun et al., 2015; Freestone et al., 2015). And indeed, our data suggests that participants made use of most of the information based on the experimental design (use of temporal regularities and multisensory information) and adapted their response strategy for a optimal decision of frequency (wait till sequence end and compare the percept to all frequencies presented in the stream).

Above we have linked the behavioural benefits to the successful extraction of temporal regularities. In principle, however, different strategies could have been used to extract this type of information in most of our experiments. First, participants could have used the time point of occurrence (400 ms) to focus their temporal attention, as intended. However, in our first five experiments, the ’early’ time point was always identical with the 3rd stimulus of the stimulus train. Thus, it is conceivable that some participants’ strategy to solve the task was based on counting stimuli instead of focusing on a specific time range. As mentioned above, most participants were unaware of the target position and even those with explicit knowledge reported that they rather relied on the early time range and did not count as they found this strategy impossible with the fast succession of stimuli. This choice might also be due to the task demands which required a stimulus discrimination rather than judging the time point or position of a particular stimulus. Hence, in our experiments counting would inevitably result in a dual-task paradigm, reducing valuable cognitive resources for the discrimination task which might be detrimental for discrimination performance (Han and Marois, 2013). Additionally, counting should result in sequential sampling of events and as outlined before this should diminish any TE effects. For example, if one always count to 3 because targets more frequently appear at this position, one should detect expected and unexpected targets at the third position equally likely. There is also no reason to assume that the 3rd position would be completely ignored when people start to actively count, even if they count to 9.

However, one might argue that numerosity might be easily encoded and retrieved without explicit counting and knowledge, like temporal estimates (Coull and Nobre, 2008; Shen and Alain, 2012). This could imply that some participants used numerosity, other time and others a mixture of both quantities for their judgements. Subject-specific performance would then be limited to the resolution of the individual domain, and different numbers of “numerosity vs. time-based participants” across experiments could explain the differential effects across experiments in this case. However, previous research indicated that the two domains have similar psychophysical properties (Çoşkun et al., 2015; Gallistel and Gelman, 2000; Meck and Church, 1983; Meck et al., 1985; Whalen et al., 1999). For instance, Meck and Church (1983) suggested that the mental representation of 1 second is equal to a count of five. Thus, counting a 5 Hz stimulus would have the same precision as judging the timing of a 5 Hz stimulus, implying that even if participants used one or the other domain, precision of judgements would be similar and would not obscure effects. Moreover, it is still an ongoing debate whether time and numerosity are mediated by different, similar or even the same mechanism(s) (Balci and Gallistel, 2006; Çoşkun et al., 2015; Fetterman and Killeen, 2010; Gallistel and Gelman, 2000; Meck and Church, 1983; Meck et al., 1985; Whalen et al., 1999). With regard to the popular pacemaker theory (e.g. Gibbon, 1991; Gibbon et al., 1984; Treisman, 1963) which posits that an internal clock or pacemaker generates beats which are accumulated to estimate duration one could even argue that time estimation is always based on counting. In our paradigm, the rhythmic stimulus train could be conceptualised as an external pacemaker which constantly resets or at least informs the internal pacemaker, in accord with studies focussing on rhythmic stimulation for external pacemaker updating (McAuley and Fromboluti, 2014). Future research is needed to disentangle these two potential mechanisms.

6.2 Potential underlying cognitive mechanisms

The pacemaker theory led to the assumption that temporal judgements and timing are supervised by an internal clock, a supramodal, centralized timing mechanism (see e.g. Gibbon et al., 1984; Treisman, 1963). If TE could be transferred across different modalities, this would strengthen this notion. Accordingly, some studies reported cross-modal TE transfer with faster RTs for expected trials in both attended and unattended modalities (Bolger et al., 2013; Jones, 2015; Lange and Röder, 2006; Miller et al., 2012). However, this was only observed for short cue-target intervals while there was no difference found in long interval trials (see Lange and Röder, 2006). Mühlberg et al. (2014) replicated the RT effects for short cue-target intervals – but more importantly – showed different effects for late cue-target intervals when these were unpredictable (by including catch trials without target presentation). For late target intervals, effects for attended and unattended modalities were inversed (hence, not driven by TE of the more frequently attended modality), questioning the general transferability of TE across modalities, and favouring the idea of modality-specific temporal networks. In our experiments we should have observed TE effects in all our conditions, if cross-modal transfer of TE exists and there would be a common network for temporal predictions. Depending on the choice of response measure, both interpretations could be drawn from our results. The pattern of RTs indicates that TE speeds responses similarly for visual, auditory and audiovisual targets. However, this may simply reflect enhanced response preparation (Green and Swets, 1966; Prinzmetal et al., 2005; van Ede et al., 2012). For discrimination sensitivity, TE effects were not always present in the unisensory conditions, thereby suggesting that sensory representations are not always affected by TE. In particular, we observed visual TE effects to be impaired which is in line with the notion of auditory dominance for temporal processing (Bertelson and Aschersleben, 1998; Fendrich and Corballis, 2001; Guttman et al., 2005; King and Nelken, 2009; Nobre and Rohenkohl, 2014; Recanzone, 2003; Repp and Penel, 2002; Shipley, 1964; Wada et al., 2003; Welch et al., 1986). Auditory dominance is also present when not durations but rather the numerosity of events (e.g. how many flashes have been presented) has to be judged (Lechelt, 1975; Philippi et al., 2008). In the latter case, mainly the judgement of visual numerosity is impaired. Hence, these findings and our data favour rather or at least the presence of modality-specific, distributed temporal networks enhancement of sensory representations rather than a single common pacemaker (see also Coull et al., 2011; Johnston et al., 2006). However, presenting evidence for higher TE-induced accuracy in the auditory domain by no means implies that TE effects cannot occur in the visual domain and that e.g. duration judgements in the visual domain are impossible. Indeed, most of the work conducted in the temporal domain have been visual experiments. However, usually those task are quite simple (e.g. matching two durations or detecting a single stimulus after a certain cue-target interval). Our task required participants to orient their attention to instances in time in noisy environments without prior knowledge of potential temporal regularities, detect the target and identify/discriminate the target. Hence, detrimental effects in the visual modality might be linked to the high task requirements and would be absent if only simple detection is required (see e.g. Correa et al., 2004), or if more complex, experimental designs are used.

6.3 Potential underlying neural mechanisms

The identification of the neural mechanisms underlying TE might open an avenue to disentangle whether it is based on timing or counting and on the use of uni-, supra- or even amodal timing networks. Entrainment of cortical oscillations could be one key mechanism underlying TE for rhythmic stimulation as used here. Concordantly, several authors have indeed observed that rhythmic stimulation creates entrained brain oscillations (for review see Merchant et al., 2015). Furthermore, Lakatos et al. (2008, 2009) have linked MSI to entrainment of cortical areas. Cravo et al. (2013) observed for visual stimuli that the amount of entrainment was related to perceptual discrimination sensitivity. Given the differences of auditory and visual temporal precision, entrainment for short inter-stimulus-intervals (ISI) sequences might be hampered in the visual modality while it might be better in the auditory modality. In turn, the perception of individual events in the entrained auditory modality is boosted, making it more likely to perceive the target. Thereby, participants could explicitly or implicitly calculate the likelihood that targets occur in a given interval within the stream. The facilitation of TE through multisensory input could be explained by direct connections between the primary sensory areas (see Driver and Noesselt, 2008). The entrainment of the auditory cortex could drive entrainment of the visual cortex (Lakatos et al., 2008), making information processing more reliable and enabling the robust extraction of visual information in our paradigm. Hence, participants could use information of both modalities, providing richer information on target presence, and making TE effects in the multisensory context more robust.

However, entrainment alone cannot account for all effects, as TE effects were reduced for visual and both visual and auditory targets in purely multisensory Experiments 3 and 4, respectively. One reason for this reduction in performance might be that performance in the auditory and visual conditions was reduced by endogenous and exogenous uncertainty which may shift the weights for preferential processing of incoming information (Rohe and Noppeney, 2016). Note that uncertainty pertains to the combined properties of visual and auditory information in Exp 3-4 , while the informational content per modality remained unchanged. This higher-order uncertainty might affect higher frequency oscillations coupled with lower-delta-band modulations (Lakatos et al., 2008) and these higher frequencies might be under control from higher multisensory (Lakatos et al., 2009) and timing areas such as the posterior Superior Temporal Sulcus (pSTS; see Driver and Noesselt, 2008; Marchant et al., 2012; Noesselt et al., 2007, 2010) or the posterior parietal cortex (Coull et al., 2011). The posterior parietal cortex has been implicated in explicit timing tasks (Coull et al., 2011) and weighting of visual and auditory information (Rohe and Noppeney, 2016). The pSTS has been related to the integration of audio-visual information especially when stimuli are presented in an isochronous rhythm, and activity in this region has been linked to performance benefits (Marchant et al., 2012). Hence, processing of multisensory stimulation in supramodal areas specialized on timing and MSI would explain robust multisensory effects in our study, while unisensory effects would be restricted to the timing precision of the individual unisensory neural networks. Finally, if our suggestions about the distinctive qualities of accuracies (indicating perception) and RTs (indicating motor preparation) are valid, one should most likely find that RT variation relates more strongly to activity in areas involved in motor activity (for overview see Coull et al., 2011). However, future studies are required to test these assumptions.

In addition to these networks in involved in sensory processing, it might also be the case that amodal temporal networks may play a role here. Recent findings suggest that pupillatory activity (i.e. pupil dilation) in a visual task increased shortly before temporally expected stimuli were presented (Akdoğan et al., 2016; Wierda et al., 2012). There is also evidence that pupil dilation occurs for visual, auditory and audiovisual events (for overview see Wang and Munoz, 2015), and that activity for these different modalities differs with faster responses for auditory stimuli and larger responses for audiovisual stimuli. Hence, pupil dilation might be used as an index of temporal and modality-specific processing. Given our results and previous findings, one should observe anticipatory pupil dilation whenever targets are expected and dilation should be stronger in the multisensory condition. Furthermore, pupil dilation differences between expected and unexpected trials in the visual condition would be less pronounced in our experiments. Although, pupillatory responses could serve as an objective measure for TE their neural underpinnings are less clear. Akdoğan et al. (2016) suggested that pupil dilation is related to the amodal norepinephrine (NE) system and activity in the locus coeruleus (LC), and that this activity represents the time interval between a cue and a target stimulus. However, although they showed anticipatory pupil dilation, they could not relate individual pupil dilation with behavioural benefits. Furthermore, while the causal role of LC-NE system in pupil dilation is often proposed there is very little empirical support for this notion(for review Wang and Munoz, 2015). Alternatively, pupil dilation might be linked to activity in the superior colliculi which also have multisensory properties(Kadunce et al., 2001; Meredith and Stein, 1983, 1986a,b; Stein and Meredith, 1993; Wallace et al., 1998, 1996). However, evidence for the involvement of the SC in multisensory integration is mostly derived from anaesthetized cats, while there is little evidence that this structure is involved in the increase in perceptual sensitivity in humans as found here. Thus, the most likely brain network underlying our effects might therefore include sensory-specific plus multisensory areas, including posterior parietal cortex and pSTS which may be instrumental in forming a multisensory event or object for which temporal regularities can be extracted more easily.

7 Conclusion

In a series of experiments, we consistently observed that hidden temporal regularities can be reliably extracted and used to successfully direct temporal attention. These temporal expectations enhance not only RTs but also discrimination sensitivity, thus pointing at a TE-induced change in sensory representations. Furthermore, TE linearly scales with early/late target likelihood and can operate over larger time windows. Most importantly, temporal expectations seem to interact with multisensory stimulation more frequently than with unisensory stimuli. This emphasises the special – yet only rarely investigated – role of multisensory interplay on temporal expectation. We propose that enhanced informational content (multisensory stimulation) protects statistical learning of temporal regularities, particularly in unreliable stimulus contexts.

Acknowledgements

This work was funded by the SFB-TR31-TPA08. FB and TN wrote the manuscript. LEM revised the manuscript. FB, LEM, and CT acquired and analysed the data. We thank Svea C. Schröder and Alexander Waite for proof-reading and revising the manuscript.

Supplementary material - Data Archiving

The data related to this article as well as the remaining supplementary material can be found on Open Science Framework (OSF; Ball, 2017). Relevant information: contributors name(s) - Felix Ball, dataset title - "The role of multisensory interplay in enabling temporal expectations: Data archive", data repository - OSF, year - 2017, and global persistent identifier - `osf.io/4m26y`.

References

Akdoğan, B. and Balcı, F. (2016). Stimulus probability effects on temporal bisection performance of mice (mus musculus). *Animal cognition*, 19(1):15–30.

Akdoğan, B. and Balcı, F. (2017). Are you early or late?: Temporal error monitoring. *Journal of Experimental Psychology: General*, 146(3):347.

Akdoğan, B., Balcı, F., and van Rijn, H. (2016). Temporal expectation indexed by pupillary response. *Timing & Time Perception*, 4(4):15–30.

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, 14(3):257–262.

Alsius, A., Navarra, J., Campbell, R., and Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, 15(9):839–843.

Balci, F., Freestone, D., and Gallistel, C. R. (2009). Risk assessment in man and mouse. *Proceedings of the National Academy of Sciences*, 106(7):2459–2463.

Balcı, F., Freestone, D., Simen, P., Desouza, L., Cohen, J. D., and Holmes, P. (2011). Optimal temporal risk assessment. *Frontiers in integrative neuroscience*, 5:56.

Balci, F. and Gallistel, C. R. (2006). Cross-domain transfer of quantitative discriminations: Is it all a matter of proportion? *Psychonomic bulletin & review*, 13(4):636–642.

Balcı, F. and Simen, P. (2014). Decision processes in temporal discrimination. *Acta psychologica*, 149:157–168.

Balcı, F. and Simen, P. (2016). A decision model of timing. *Current Opinion in Behavioral Sciences*, 8:94–101.

Ball, F. (2017). The role of multisensory interplay in enabling temporal expectations: Data archive. Archive url: `osf.io/4m26y`.

Ball, F., Bernasconi, F., and Busch, N. A. (2015). Semantic relations between visual objects can be unconsciously processed but not reported under change blindness. *J Cogn Neurosci*, 27(11):2253–2268.

Ball, F. and Busch, N. A. (2015). Change detection on a hunch: Pre-attentive vision allows "sensing" of unique feature changes. *Attention, Perception, & Psychophysics*, 77(8):2570–2588.

Ball, F., Elzemann, A., and Busch, N. A. (2014). The scene and the unseen: manipulating photographs for experiments on change blindness and scene memory. *Behav Res Methods*, 46(3):689–701.

Barakat, B. K., Seitz, A. R., and Shams, L. (2013). The effect of statistical learning on internal stimulus representations: Predictable items are enhanced even when not predicted. *Cognition*, 129(2):205–211.

Baylis, G. C. and Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology. Human Perception and Performanc*, 19(3):451–470.

Beauchamp, M. S., Pasalar, S., and Ro, T. (2010). Neural substrates of reliability-weighted visual-tactile multisensory integration. *Frontiers in systems neuroscience*, 4:25.

Beck, M. R., Hong, S. L., van Lamsweerde, A. E., and Ericson, J. M. (2014). The effects of incidentally learned temporal and spatial predictability on response times and visual fixations during target detection and discrimination. *PloS one*, 9(4):e94539.

Behrmann, M., Zemel, R. S., and Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. *Journal of Experimental Psychology. Human Perception and Performanc*, 24(4):1011–1036.

Bendixen, A., SanMiguel, I., and Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: a review. *International Journal of Psychophysiology*, 83(2):120–131.

Berkay, D., Freestone, D., and Balcı, F. (2016). Mice and rats fail to integrate exogenous timing noise into their time-based decisions. *Animal cognition*, 19(6):1215–1225.

Bertelson, P. and Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, 5(3):482–489.

Bertelson, P., Vroomen, J., De Gelder, B., and Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & psychophysics*, 62(2):321–332.

Bischoff, M., Walter, B., Blecker, C., Morgen, K., Vaitl, D., and Sammer, G. (2007). Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologia*, 45(3):578–586.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, 113(4):700.

Bolger, D., Trost, W., and Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. *Acta psychologica*, 142(2):238–244.

Brainard, D. H. (1997). The psychophysics toolbox. *Spat Vis*, 10(4):433–436.

Çavdaroğlu, B., Zeki, M., and Balcı, F. (2014). Time-based reward maximization. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1637):20120461.

Chen, Z. (2000). An object-based cost of visual filtering. *Perception & Psychophysics*, 62(3):482–495.

Correa, Á., Lupiáñez, J., Madrid, E., and Tudela, P. (2006). Temporal attention enhances early visual processing: A review and new evidence from event-related potentials. *Brain research*, 1076(1):116–128.

Correa, Á., Lupiáñez, J., Milliken, B., and Tudela, P. (2004). Endogenous temporal orienting of attention in detection and discrimination tasks. *Perception & Psychophysics*, 66(2):264–278.

Çoşkun, F., Sayalı, Z. C., Gürbüz, E., and Balcı, F. (2015). Optimal time discrimination. *The Quarterly Journal of Experimental Psychology*, 68(2):381–401.

Coull, J. and Nobre, A. (2008). Dissociating explicit timing from temporal expectation with fmri. *Current opinion in neurobiology*, 18(2):137–144.

Coull, J. and Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both pet and fmri. *The Journal of Neuroscience*, 18(18):7426–7435.

Coull, J. T., Cheng, R.-K., and Meck, W. H. (2011). Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology*, 36(1):3–25.

Cravo, A. M., Rohenkohl, G., Wyart, V., and Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience*, 33(9):4002–4010.

Di Luca, M., Machulla, T.-K., and Ernst, M. O. (2009). Recalibration of multisensory simultaneity: cross-modal transfer coincides with a change in perceptual latency. *Journal of vision*, 9(12):7–7.

Diederich, A. and Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, 66(8):1388–1404.

Doehrmann, O. and Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain research*, 1242:136–150.

Doherty, J. R., Rao, A., Mesulam, M. M., and Nobre, A. C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. *The Journal of Neuroscience*, 25(36):8259–8266.

Doyle, M. C. and Snowden, R. J. (2001). Identification of visual stimuli is improved by accompanying auditory stimuli: The role of eye movements and sound location. *Perception*, 30(7):795–810.

Driver, J. and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron*, 57(1):11–23.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology. General*, 113(4):501–517.

Fendrich, R. and Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, 63(4):719–725.

Fetterman, J. G. and Killeen, P. R. (2010). Categorical counting. *Behavioural processes*, 85(1):28–35.

Forster, B., Cavina-Pratesi, C., Aglioti, S. M., and Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research*, 143(4):480–487.

Freestone, D. M., Balcı, F., Simen, P., and Church, R. M. (2015). Optimal response rates in humans and rats. *Journal of Experimental Psychology: Animal Learning and Cognition*, 41(1):39.

Fujisaki, W. and Nishida, S. (2007). Feature-based processing of audio-visual synchrony perception revealed by random pulse trains. *Vision research*, 47(8):1075–1093.

Gallistel, C. R. and Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in cognitive sciences*, 4(2):59–65.

Gibbon, J. (1991). Origins of scalar timing. *Learning and motivation*, 22(1):3–38.

Gibbon, J., Church, R. M., and Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of sciences*, 423(1):52–77.

Gondan, M., Niederhaus, B., Rösler, F., and Röder, B. (2005). Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. *Perception & psychophysics*, 67(4):713–726.

Grant, K. W. and Greenberg, S. (2001). Speech intelligibility derived from asynchronous processing of auditory-visual information. In *AVSP 2001-International Conference on Auditory-Visual Speech Processing*.

Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley, New York.

Griffin, I. C., Miniussi, C., and Nobre, A. C. (2001). Orienting attention in time. *Frontiers in Bioscience*, 6:660–671.

Griffin, I. C., Miniussi, C., and Nobre, A. C. (2002). Multiple mechanisms of selective attention: differential modulation of stimulus processing by attention to space or time. *Neuropsychologia*, 40(13):2325–2340.

Guttman, S. E., Gilroy, L. A., and Blake, R. (2005). Hearing what the eyes see auditory encoding of visual temporal sequences. *Psychological science*, 16(3):228–235.

Han, S. W. and Marois, R. (2013). The source of dual-task limitations: Serial or parallel processing of multiple response selections? *Attention, Perception, & Psychophysics*, 75(7):1395–1405.

Jaekl, P. M. and Harris, L. R. (2009). Sounds can affect visual perception mediated primarily by the parvocellular pathway. *Visual neuroscience*, 26(5-6):477–486.

Jaramillo, S. and Zador, A. M. (2011). Auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nature Neuroscience*, 14(2):246–251.

Jepma, M., Wagenmakers, E.-J., and Nieuwenhuis, S. (2012). Temporal expectation and information processing: A model-based analysis. *Cognition*, 122(3):426–441.

Johnston, A., Arnold, D. H., and Nishida, S. (2006). Spatially localized distortions of event time. *Current Biology*, 16(5):472–479.

Jones, A. (2015). Independent effects of bottom-up temporal expectancy and top-down spatial attention. an audiovisual study using rhythmic cueing. *Frontiers in integrative neuroscience*, 8:96.

Jones, J. A. and Jarick, M. (2006). Multisensory integration of speech signals: The relationship between space and time. *Experimental Brain Research*, 174(3):588–594.

Jones, M. R., Moynihan, H., MacKenzie, N., and Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological science*, 13(4):313–319.

Kadunce, D. C., Vaughan, W. J., Wallace, M. T., and Stein, B. E. (2001). The influence of visual and auditory receptive field organization on multisensory integration in the superior colliculus. *Experimental Brain Research*, 139(3):303–310.

Keetels, M. and Vroomen, J. (2007). No effect of auditory–visual spatial disparity on temporal recalibration. *Experimental Brain Research*, 182(4):559–565.

Keuss, P., Van der Zee, F., and Van den Bree, M. (1990). Auditory accessory effects on visual processing. *Acta psychologica*, 75(1):41–54.

King, A. J. and Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nature neuroscience*, 12(6):698–701.

Klemmer, E. T. (1956). Time uncertainty in simple reaction time. *Journal of experimental psychology*, 51(3):179.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS one*, 2(9):e943.

Kovalenko, L. Y. and Busch, N. A. (2016). Probing the dynamics of perisaccadic vision with eeg. *Neuropsychologia*, 85:337–48.

Kramer, A. F., Weber, T. A., and Watson, S. E. (1997). Object-based attentional selection: Grouped arrays or spatially invariant representations? comment on vecera and farah (1994). *Journal of Experimental Psychology. General*, 126(1):3–13.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *science*, 320(5872):110–113.

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., and Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron*, 64(3):419–430.

Lange, K. and Röder, B. (2006). Orienting attention to points in time improves stimulus processing both within and across modalities. *Journal of Cognitive Neuroscience*, 18(5):715–729.

Lange, K., Rösler, F., and Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. *Psychophysiology*, 40(5):806–817.

Lechelt, E. C. (1975). Temporal numerosity discrimination: Intermodal comparisons revisited. *British Journal of Psychology*, 66(1):101–108.

Luck, S., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., and Hawkins, H. L. (2004). Effects of spatial cueing on luminance detectability: psychophysical and electrophysiological evidence for early selection. *J Exp Psychol Hum*, 20:887–904.

Luria, A. R. (1968). *The mind of a mnemonist: A little book about a vast memory*. Harvard University Press.

Maiworm, M. and Röder, B. (2011). Suboptimal auditory dominance in audiovisual integration of temporal cues. *Tsinghua Science & Technology*, 16(2):121–132.

Marchant, J. L., Ruff, C. C., and Driver, J. (2012). Audiovisual synchrony enhances bold responses in a brain network including multisensory sts while also enhancing target-detection performance for both modalities. *Human brain mapping*, 33(5):1212–1224.

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., and Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition*, 115(1):186–191.

McAuley, J. D. and Fromboluti, E. K. (2014). Attentional entrainment and perceived event duration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658):20130401.

Meck, W. H. and Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9(3):320.

Meck, W. H., Church, R. M., and Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 11(4):591.

Menceloglu, M., Grabowecky, M., and Suzuki, S. (2016). Temporal expectation weights visual signals over auditory signals. *Psychonomic Bulletin & Review*, pages 1–7.

Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., and Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Phil. Trans. R. Soc. B*, 370(1664):20140093.

Meredith, M. A. and Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221(4608):389–391.

Meredith, M. A. and Stein, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res*, 365(2):350–354.

Meredith, M. A. and Stein, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol*, 56(3):640–662.

Miller, J. E., Carlson, L. A., and McAuley, J. D. (2012). When what you hear influences when you see listening to an auditory rhythm influences the temporal allocation of visual attention. *Psychological science*, page 0956797612446707.

Miniussi, C., Wilding, E. L., Coull, J., and Nobre, A. C. (1999). Orienting attention in time. *Brain*, 122(8):1507–1518.

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., and Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental brain research*, 184(1):39–52.

Mühlberg, S., Oriolo, G., and Soto-Faraco, S. (2014). Cross-modal decoupling in temporal attention. *European Journal of Neuroscience*, 39(12):2089–2097.

Näätänen, R. and Merisalo, A. (1977). Expectancy and preparation in simple reaction time. *Attention and performance VI*, pages 115–138.

Näätänen, R., Muranen, V., and Merisalo, A. (1974). Timing of expectancy peak in simple reaction time situation. *Acta Psychologica*, 38(6):461–470.

Niemi, P. and Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, 89(1):133.

Nobre, A. C. (2001). Orienting attention to instants in time. *Neuropsychologia*, 39(12):1317–1328.

Nobre, A. C. and Rohenkohl, G. (2014). Time for the fourth dimension in attention. In Nobre, A. C. and Kastner, S., editors, *The Oxford Handbook of Attention*, pages 676–724. Oxford University Press.

Noesselt, T., Fendrich, R., Bonath, B., Tyll, S., and Heinze, H.-J. (2005). Closer in time when farther in space–spatial factors in audiovisual temporal integration. *Brain research. Cognitive brain research*, 25(2):443–458.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., and Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience*, 27(42):11431–11441.

Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H. J., and Driver, J. (2010). Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *J Neurosci*, 30(41):13609–13623.

Parise, C. V., Spence, C., and Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Current Biology*, 22(1):46–49.

Philippi, T. G., van Erp, J. B., and Werkhoven, P. J. (2008). Multisensory temporal numerosity judgment. *Brain research*, 1242:116–125.

Posner, M., Snyder, C. R., and Davidson, B. J. (1980). Attention and the detection of signals. *J Exp Psychol Gen*, 109:160–174.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32:3–25.

Prinzmetal, W., McCool, C., and Park, S. (2005). Attention: reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General*, 134(1):73.

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of neurophysiology*, 89(2):1078–1093.

Reisberg, D., Mclean, J., and Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd, B. and Campbell, R., editors, *Hearing by eye: The psychology of lip-reading*, pages 97–114. Lawrence Erlbaum Associates, Hillsdale.

Repp, B. H. and Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5):1085.

Risberg, A. and Lubker, J. (1978). Prosody and speechreading. *Speech Transmission Laboratory Quarterly Progress Report and Status Report*, 4:1–16.

Roach, N. W., Heron, J., and McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1598):2159–2168.

Rohe, T. and Noppeney, U. (2016). Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Current Biology*, 26(4):509–514.

Rohenkohl, G., Coull, J. T., and Nobre, A. C. (2011). Behavioural dissociation between exogenous and endogenous temporal orienting of attention. *PLoS One*, 6(1):e14620.

Rohenkohl, G., Cravo, A. M., Wyart, V., and Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *The Journal of Neuroscience*, 32(24):8424–8428.

Rohenkohl, G., Gould, I. C., Pessoa, J., and Nobre, A. C. (2014). Combining spatial and temporal expectations to improve visual perception. *Journal of Vision*, 14(4):8.

Rolke, B. and Hofmann, P. (2007). Temporal uncertainty degrades perceptual processing. *Psychonomic Bulletin & Review*, 14(3):522–526.

Sanabria, D., Capizzi, M., and Correa, Á. (2011). Rhythms that speed you up. *Journal of Experimental Psychology: Human Perception and Performance*, 37(1):236.

Sanders, A. (1975). The foreperiod effect revisited. *The Quarterly Journal of Experimental Psychology*, 27(4):591–598.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in cognitive sciences*, 12(3):106–113.

Shams, L. and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences*, 12(11):411–417.

Shen, D. and Alain, C. (2011). Temporal attention facilitates short-term consolidation during a rapid serial auditory presentation task. *Experimental Brain Research*, 215(3):285–292.

Shen, D. and Alain, C. (2012). Implicit temporal expectation attenuates auditory attentional blink. *PLoS ONE*, 7:1–6.

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, 145(3638):1328–1330.

Shore, D. I. and Simic, N. (2005). Integration of visual and tactile stimuli: top-down influences require time. *Experimental Brain Research*, 166(3-4):509–517.

Sinnett, S., Soto-Faraco, S., and Spence, C. (2008). The co-occurrence of multisensory competition and facilitation. *Acta psychologica*, 128(1):153–161.

Soto-Faraco, S., Morein-Zamir, S., and Kingstone, A. (2005). On audiovisual spatial synergy: The fragility of the phenomenon. *Perception & psychophysics*, 67(3):444–457.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? evaluating the spatial rule. *Annals of the New York Academy of Sciences*, 1296(1):31–49.

Stein, B. E., London, N., Wilkinson, L. K., and Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *Journal of cognitive neuroscience*, 8(6):497–506.

Stein, B. E. and Meredith, M. A. (1993). *The merging of the senses*. The MIT Press, Cambridge, MA, US.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Routledge, Hillsdale, NJ: LEA.

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., Kurela, L. R., Siemann, J. K., James, T. W., and Wallace, M. T. (2014). Identifying and quantifying multisensory integration: A tutorial review. *Brain Topography*, 27(6):707–730.

Strybel, T. and Fujimoto, K. (2000). Minimum audible angles in the horizontal and vertical planes: effects of stimulus onset asynchrony and burst duration. *J Acoust Soc Am*, 108(6):3092–3095.

Sumby, W. H. and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of america*, 26(2):212–215.

Talsma, D., Doty, T. J., and Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cerebral cortex*, 17(3):679–690.

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the" internal clock". *Psychological Monographs: General and Applied*, 77(13):1.

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., and Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *J Exp Psychol Hum Percept Perform*, 34(5):1053–1065.

van Ede, F., de Lange, F. P., and Maris, E. (2012). Attentional cues affect accuracy and reaction time via different cognitive and neural processes. *Journal of Neuroscience*, 32(30):10408–10412.

Vecera, S. P. and Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology. General*, 123(2):146–160.

Vroomen, J., Bertelson, P., and De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Attention, Perception, & Psychophysics*, 63(4):651–659.

Vroomen, J. and Keetels, M. (2006). The spatial constraint in intersensory pairing: No role in temporal ventriloquism. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4):1063.

Wada, Y., Kitagawa, N., and Noguchi, K. (2003). Audio–visual integration in temporal perception. *International journal of psychophysiology*, 50(1):117–124.

Wallace, M. T., Meredith, M. A., and Stein, B. E. (1998). Multisensory integration in the superior colliculus of the alert cat. *Journal of Neurophysiology*, 80(2):1006–1010.

Wallace, M. T., Wilkinson, L. K., and Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology*, 76(2):1246–1266.

Wang, C.-A. and Munoz, D. P. (2015). A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Current opinion in neurobiology*, 33:134–140.

Wang, P. and Nikolić, D. (2011). An lcd monitor with sufficiently precise timing for research in vision. *Front Hum Neurosci.*, 5(85):1–10.

Welch, R. B., DutionHurt, L. D., and Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics*, 39(4):294–300.

Werkhoven, P. J., van Erp, J. B., and Philippi, T. G. (2009). Counting visual and tactile events: the effect of attention on multisensory integration. *Attention, Perception, & Psychophysics*, 71(8):1854–1861.

Werner, S. and Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *Journal of Neuroscience*, 30(7):2662–2675.

Westheimer, G. and Ley, E. (1996). Temporal uncertainty effects on orientation discrimination and stereoscopic thresholds. *J. Opt. Soc. Am. A*, 13(4):884–886.

Whalen, J., Gallistel, C., and Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10(2):130–137.

Wierda, S. M., van Rijn, H., Taatgen, N. A., and Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, 109(22):8456–8460.

Yeshurun, Y. and Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, 396:72–75.

List of Figures

List of Tables

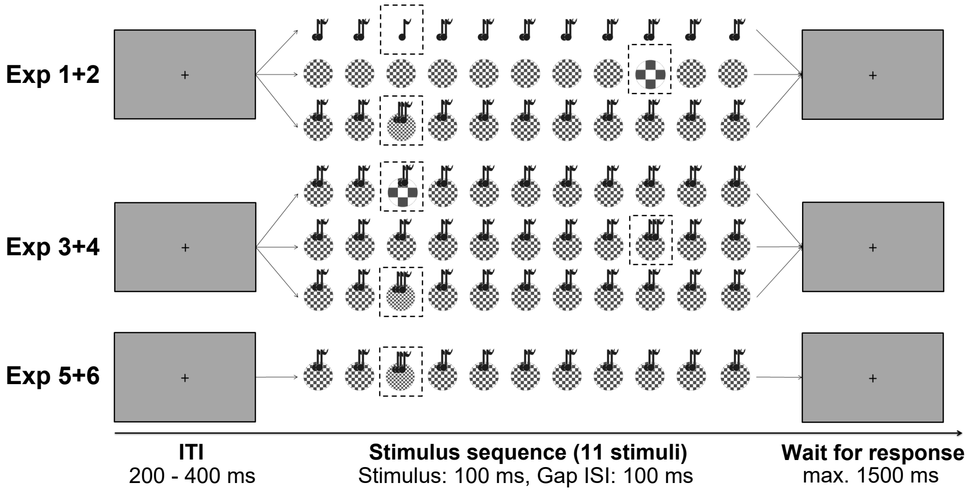

Figure 1: Experimental Design. Each trial started with a blank screen (inter-trial-interval) lasting for 200–400 ms followed by a sequence of 11 auditory (Exp 1 + 2), visual (Exp 1 + 2), or audiovisual stimuli (all Exp). Stimuli were presented for 100 ms with a 100 ms gap in between. After the stimulus sequence, a blank screen was displayed for a maximum of 1500 ms. A response within this time range terminated the blank screen immediately. *Top row:* Design of Experiments 1 and 2 with the three experimental conditions from top to bottom: auditory, visual, and audiovisual. Targets were either presented at the 3rd or 9th position. Note, that squares highlight the target (lower or higher frequency than distractor items) for illustrative purposes only and were not present in the experiment. *Middle row:* Design of Experiments 3 and 4 with three experimental conditions from top to bottom: audiovisual sequences with unisensory auditory, visual, or audiovisual target. *Bottom row:* In Experiments 5 and 6, only multisensory streams with redundant multisensory targets were used. In Experiment 6, six different target positions were used (2,3,4 vs. 8,9,10). For auditory presentation, either headphones (Exp 2, 4, 6) or speakers were used, the latter in close vicinity to the visual stimulation (Exp 1, 3, 5) in order to manipulate audiovisual spatial uncertainty between experiments.

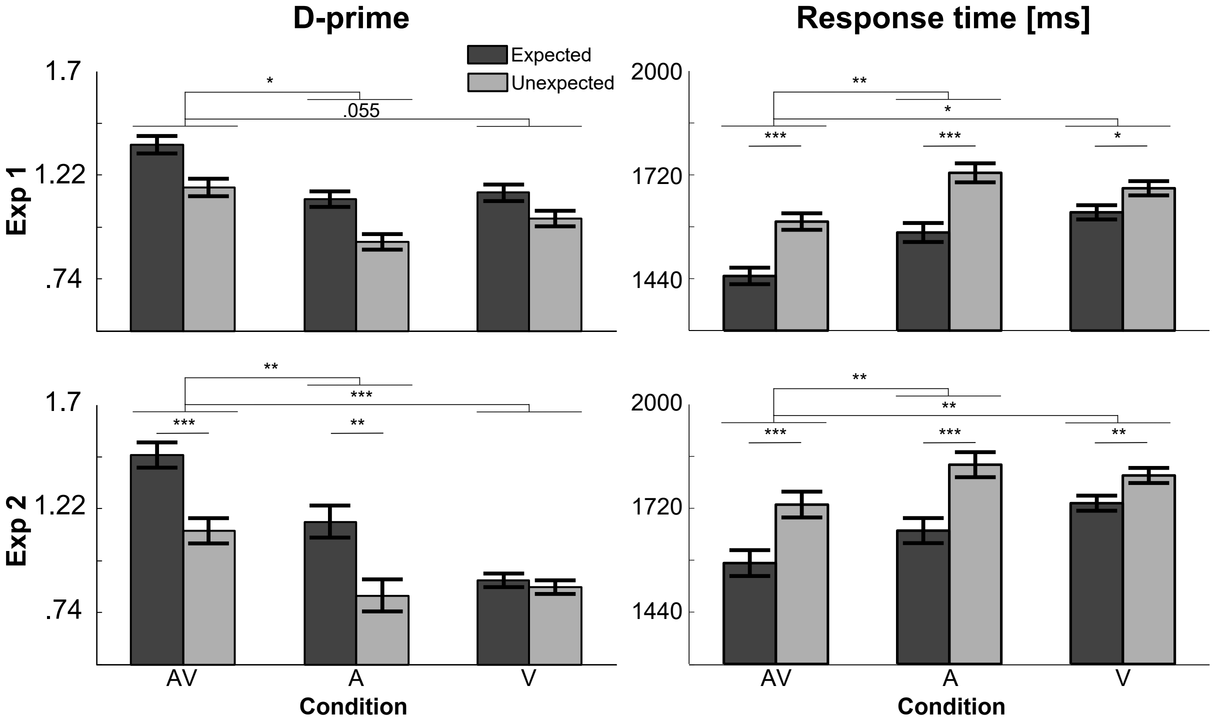

Figure 2: d′ and RT values for Experiments 1 and 2. d′ values are displayed in the left column and RTs in the right column, separately for auditory (A), visual (V), and audiovisual (AV) conditions. *Top row:* Results Experiment 1. *Bottom row:* Results Experiment 2. Error bars depict standard errors of the difference “expected - unexpected”. Asterisks denote significant effects (*** = <.001, ** = <.01, * = <.05) for main effects of modality, and individual TE effects (bar above each modality) in case the interaction of modality and TE was significant. Note that modality-specific effects were only tested and depicted if the interaction of TE and Modality was significant (though the main effect of TE was always significant, see main text).

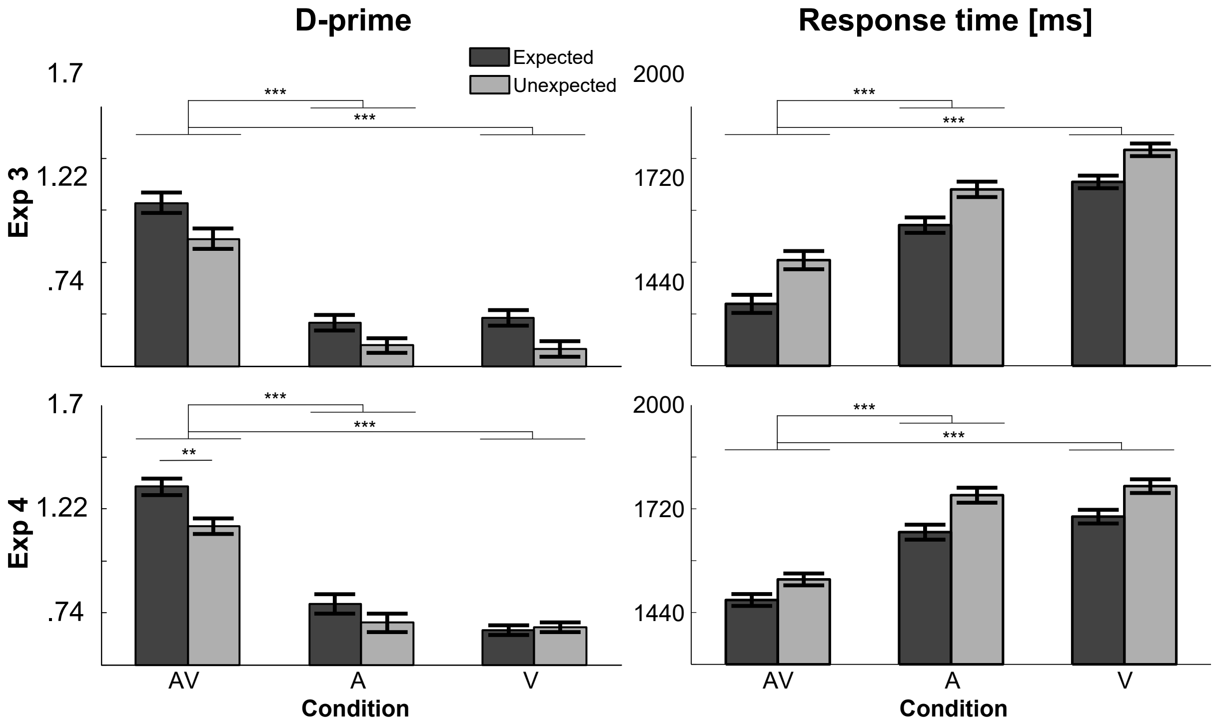

Figure 3: d′ and RT measures for Experiments 3 and 4. d′ scores are depicted in the left column and RTs in the right column, separately for auditory (A), visual (V), and audiovisual (AV) conditions. *Top row:* Results Experiment 3. *Bottom row:* Results Experiment 4. Error bars are standard errors of the difference "expected - unexpected". Asterisks denote significant effects (*** = <.001, ** = <.01, * = <.05) for main effects of modality, and individual TE effects (bar above each modality) in case the interaction of modality and TE was significant. Note that modality-specific effects were only tested and depicted if the interaction of TE and Modality was significant (though the main effect of TE was always significant, see main text).

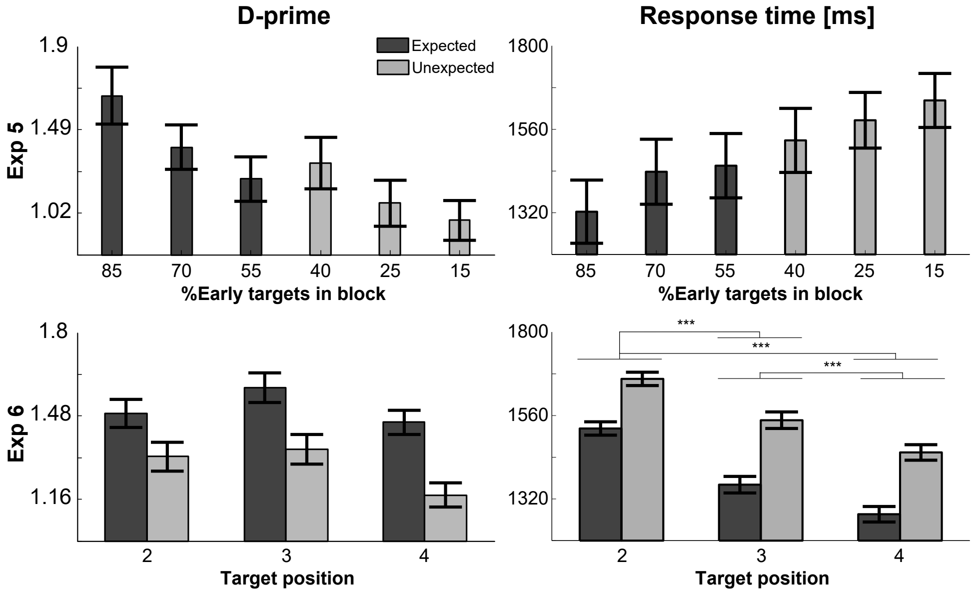

Figure 4: d′ and RT values for Experiments 5-6. d′ values are depicted in the left column and RTs are shown in the right column. *Top row:* Results Experiment 5. Error bars are standard errors separately for all probabilities. *Bottom row:* Results Experiment 6. Error bars of Experiments 6 are standard errors of the difference expected - unexpected. Significant condition differences are only depicted for Experiment 6. Asterisks denote significant effects (*** = <.001, ** = <.01, * = <.05) for main effect of position only.Note that modality-specific effects were only tested and depicted if the interaction of TE and target Position was significant (though the main effect of TE was always significant, see main text).

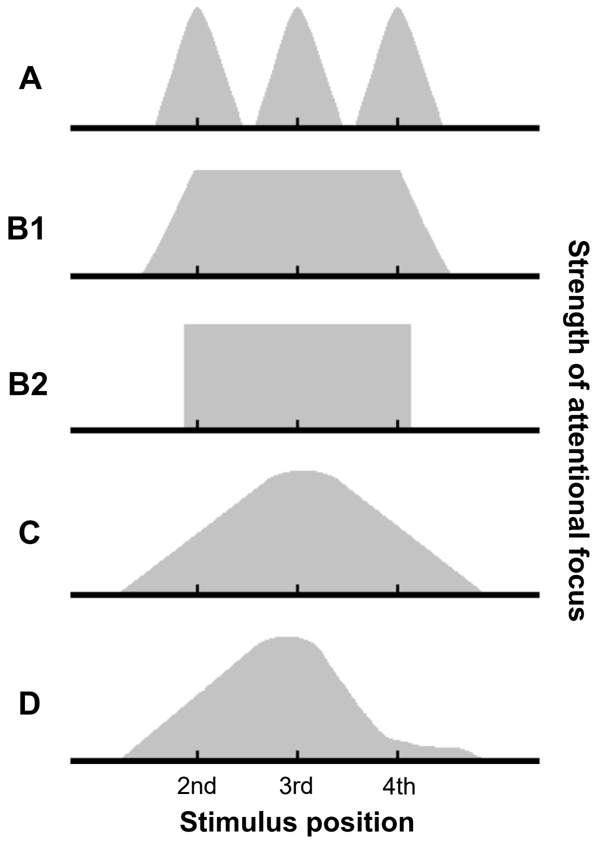

Figure 5: Possible scenarios for the shape of the strength of temporal attentional focus (denoted on y-axis) operating across several target positions in Experiment 6 (denoted on x-axis).

Modality	Experiment	$ETlow$	$EThigh$	$LTlow$	$LThigh$
Auditory	Exp1	2901 ± 101	3134 ± 80	2905 ± 65	3088 ± 68
[in Hz]	Exp2	2883 ± 110	3141 ± 97	2885 ± 103	3111 ± 87
	Exp3	2873 ± 87	3158 ± 136	2872 ± 114	3107 ± 111
	Exp4	2879 ± 91	3140 ± 62	2886 ± 90	3115 ± 72
	Exp5	2888 ± 69	3112 ± 72	2917 ± 43	3083 ± 60
	Exp6	2866 ± 99	3136 ± 85	2843 ± 109	3116 ± 89
Visual	Exp1	3.75 ± 0.32	6.17 ± 0.66	3.74 ± 0.55	5.78 ± 0.47
[in cycles/degree]	Exp2	3.82 ± 0.46	5.97 ± 0.48	3.76 ± 0.51	5.88 ± 0.47
	Exp3	3.87 ± 0.41	6.18 ± 0.46	3.89 ± 0.51	6.09 ± 0.48
	Exp4	3.9 ± 0.22	6.06 ± 0.43	3.81 ± 0.59	5.95 ± 0.39
	Exp5	3.81 ± 0.26	6.29 ± 0.51	3.64 ± 0.4	6.16 ± 0.51
	Exp6	4.01 ± 0.39	5.95 ± 0.42	3.81 ± 0.5	5.83 ± 0.47

Table 1: Mean target frequencies of all experiments: Mean target frequencies plus/minus standard deviations are listed for each modality (auditory, visual), early (ET) and late (LT) targets, and each target frequency (low and high). Distractor frequencies ranged from 2.04-2.33 cycles per degree and 2975-3025 Hz. Note that mean target frequencies did not differ between experiments (see main text for details).

Exp	Effect	$C1$	$C2$	$meanC1$	$meanC2$	$t(29)$	pBF	SD
Exp 1	Modality	AV	A	1.264	1.010	2.624	.021	.530
		AV	V	1.264	1.078	2.192	.055	.464
		A	V	1.010	1.078	-.453	1	.824
Exp 2	Modality	AV	A	1.294	.989	3.269	.004	.511
		AV	V	1.294	.874	5.376	<.001	.428
		A	V	.989	.874	.899	1	.701
	Interaction	$AV_{expected}$	$AV_{unexpected}$	1.470	1.118	5.118	<.001	.377
	(Mod x TE)	$A_{expected}$	$A_{unexpected}$	1.160	.817	3.757	.001	.5
		$V_{expected}$	$V_{unexpected}$	.889	.858	.591	.839	.288
Exp 3	Modality	AV	A	1.172	.647	7.832	<.001	.367
		AV	V	1.172	.651	5.269	<.001	.541
		A	V	.647	.651	-.028	1	.714
Exp 4	Modality	AV	A	1.232	.737	6.313	<.001	.429
		AV	V	1.232	.666	5.818	<.001	.533
		A	V	.737	.666	.587	1	.670
	Interaction	$AV_{expected}$	$AV_{unexpected}$	1.323	1.141	3.392	.006	.293
	(Mod x TE)	$A_{expected}$	$A_{unexpected}$	.781	.694	1.263	.65	.379
		$V_{expected}$	$V_{unexpected}$	.660	.671	-.284	1	.213

Table 2: Post-hoc tests for d′: The table presents post-hoc tests for all experiments (Exp) in which main or interaction effects (effects) of the repeated-measures ANOVAs were significant. We list the two conditions (C1, C2) which were compared and their mean d′ values (mean C1/C2), together with t-values, Bonferroni corrected p-values (pBF), and the standard deviation of the difference (SD). Abbreviations used: AV = audiovisual, A = audio, V = visual, Mod = Modality, TE = Temporal Expectation.

Exp	Effect	$C1$	$C2$	$meanC1$	$meanC2$	$t(29)$	pBF	SD
Exp 1	Modality	AV	A	1520.571	1644.84	-3.111	.006	218.806
		AV	V	1520.571	1651.356	-2.841	.012	252.1
		A	V	1644.84	1651.356	-.095	1	374.977
	Interaction	$\text{AV}_{expected}$	$\text{AV}_{unexpected}$	1447.225	1593.918	-5.332	<.001	150.692
	(Mod x TE)	$\text{A}_{expected}$	$\text{A}_{unexpected}$	1564.377	1725.343	-6.019	<.001	146.512
		$\text{V}_{expected}$	$\text{V}_{unexpected}$	1618.836	1638.875	-2.309	.042	154.303
Exp 2	Modality	AV	A	1650.96	1748.531	-3.341	.003	159.956
		AV	V	1650.96	1771.011	-3.997	.001	164.524
		A	V	1748.531	1771.011	-.47	1	261.946
	Interaction	$\text{AV}_{expected}$	$\text{AV}_{unexpected}$	1572.042	1729.879	-4.271	<.001	260.805
	(Mod x TE)	$\text{A}_{expected}$	$\text{A}_{unexpected}$	1659.671	1837.391	-4.59	<.001	497.036
		$\text{V}_{expected}$	$\text{V}_{unexpected}$	1733.61	1808.412	-3.123	.006	346.200
Exp 3	Modality	AV	A	1526.274	1728.146	-6.007	<.001	184.074
		AV	V	1526.274	1839.886	-7.607	<.001	225.821
		A	V	1728.146	1839.886	-1.905	.2	321.338
Exp 4	Modality	AV	A	1501.427	1706.781	-5.524	<.001	203.603
		AV	V	1501.427	1740.014	-6.394	<.001	204.377
		A	V	1706.781	1740.014	-.622	1	292.555
Exp 6	Position	2nd	3rd	1594.518	1453.729	6.115	<.001	125.694
		2nd	4th	1594.518	1364.805	6.405	<.001	196.052
		3rd	4th	1453.729	1364.805	4.396	<.001	110.805

Table 3: Post-hoc tests for RTs: The table denotes post-hoc test measures for all experiments (Exp) in which main or interaction effects (effects) were significant. Conditions (C1, C2) which were compared are listed plus their average RT values (mean C1/C2), t-value, the Bonferroni corrected p-value (pBF), and the standard deviation of the difference (SD). Abbreviations used: AV = audiovisual, A = audio, V = visual, Mod = Modality, TE = Temporal Expectation.