Computers & Graphics 124 (2024) 104097

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on EuroXR 2023 Best Papers

Psychophysiology of rhythmic stimuli and time experience in virtual reality

Stéven Picard *, Jean Botev

VR/AR Lab, University of Luxembourg, L-4364 Esch-sur-Alzette, Luxembourg

ARTICLE INFO

Keywords:
Virtual reality
Human–computer interaction
User experience
Time experience
Rhythmic stimuli
Physiological data

ABSTRACT

Time experience is an essential part of one's perception of any environment, real or virtual. In this article, from a virtual environment design perspective, we explore how rhythmic stimuli can influence an unrelated cognitive task regarding time experience and performance in virtual reality. This study explicitly includes physiological data to investigate how, overall, experience correlates with psychophysiological observations. The task involves sorting 3D objects by shape, with varying rhythmic stimuli in terms of their tempo and sensory channel (auditory and/or visual) in different trials, to collect subjective measures of time estimation and judgment. The results indicate different effects on time experience and performance depending on the context, such as user fatigue and trial repetition. Depending on the context, a positive impact of audio stimuli or a negative impact of visual stimuli on task performance can be observed, as well as time being underestimated concerning tempo in relation to task familiarity. However, some effects are consistent regardless of context, such as time being judged to pass faster with additional stimuli or consistent correlations between participants' performance and time experience, suggesting flow-related aspects. We also observe correlations between time experience with eye-tracking data and body temperature, yet some of these correlations may be due to a confounding effect of fatigue. If confirmed as separate from fatigue, these physiological data could be used as reference point for evaluating a user's time experience. This might be of great interest for designing virtual environments, as purposeful stimuli can strongly influence task performance and time experience, both essential components of virtual environment user experience.

1. Introduction

While time itself is a concept, it is also something humans can perceive. However, the perception of time is subjective, and this experience is an integral part of the overall experience of any environment, with virtual environments being no exception. Therefore, acknowledging this in their conception and actively devising virtual environments to modulate the time experience of users would be an exciting design instrument. Time perception as an interdisciplinary topic is explored in numerous scientific studies in disciplines as diverse as psychology or neuroscience. In a previous study combining cognitive science and computer science, we already examined the relationships between time perception and rhythmic stimuli with a sorting game in a two-dimensional setting, which revealed varying time experience and performance effects depending on whether single or combined stimuli were used in relation to their tempo [1]. However, that initial experiment was limited to a crowd-sourced desktop setting, and we adapted it for Virtual Reality (VR), which allows for extended control of the test environment plus extending the initial set of questions to include stimuli and time experience aspects in fully immersive environments. Therefore, while our main goal with this study is to investigate and interpret anything significant by having a correlation study process, we do come with assumptions coming from our initial study, which are different effects on time perception depending on the type of stimuli present (audio, visual, or both), tempo-related time estimation errors for combined stimuli, as well as decreased task performance by the presence of visual stimuli. A locally controlled environment furthermore allows us to let participants use hardware of our choice, including hardware with physiological sensors. Therefore, we also investigate the relationship between physiological data and subjective time experience. This article extends a paper originally published at the EuroXR 2023 conference [2], which did not comprise any physiological data analysis and correlations provided in this work to convey the full picture. Before detailing the experiment in Section 3, analyzing the data in Section 4, and discussing the results in Section 5, we will first provide the necessary background and review related work in the following Section 2. Section 6 then concludes the article with an outlook and potential impact of our findings on general virtual environment design.

2. Background and related work

The most common time perception model in literature is the clock model, which assumes an "internal clock" as a body system dedicated

* Corresponding author.
E-mail address: steven.picard@uni.lu (S. Picard).

https://doi.org/10.1016/j.cag.2024.104097
Received 13 March 2024; Received in revised form 8 August 2024; Accepted 24 September 2024
Available online 27 September 2024

to time perception [3]. This system is usually tied to a model producing "ticks" or "units", such as an oscillator or a pacemaker model, where the body keeps track of time by counting these ticks. In these models, it is assumed that time perception can be changed either through the speed at which these ticks are produced or by skipping some, with these effects potentially resulting from external stimuli unrelated to time. Counting the ticks can be delegated to attention, making attentional resources a key element of time perception [4,5]. Attention is believed to act as a gate or switch on accounting the time units, where paying less attention to time will result in compressed time experience as time units are likely to be skipped. Another source of subjective temporal distortions is the use of arousal levels, which are believed to affect the clock speed, resulting in an altered time experience [6–8]. However, using an internal clock model is not necessary to predict time perception accurately [9]: in the context of watching videos in different scenarios, time perception was accurately predicted by a classification network using changes in perceptual content and visual spatial attention (more specifically, gaze position). Nevertheless, it is a natural way to interpret time experience based on the focus on attention and arousal that gives initial directions to time perception studies.

In the literature, we can observe different types of timing tasks, which involve different processing mechanics. The most common aspects are time estimation (i.e., asking for a duration estimate of an event with units, such as seconds) and the feeling of time passage (i.e., judging if time passes by quickly or slowly), which is also referred to as time judgment.

This difference can be observed in depressed subjects underestimating time but judging it as passing slowly [6], with boredom-prone people judging boring tasks as passing slowly but not necessarily overestimating those [10], or with players of the game "Thumper" reporting faster time passing without time estimation errors [11].

Having defined time perception, we can discuss the state of "flow", one of the applications of time perception alterations. Flow is a psychological state of full attention on a task defined by Csikzentmihalyi, represented through nine dimensions: challenge-skill balance, action-awareness merging, clear goals, unambiguous feedback, concentration on task, sense of control, loss of self-consciousness, time transformation, and autotelic experience. The psychological state of flow is a research subject in itself, centered around one's relation to a task as it primarily relies on the challenge-skill ratio aspect [12]; it is often considered a desirable state, and time perception alterations are one of its manifestations. In social media, the manifestation of flow seems to be influenced by the positive effect of telepresence on enjoyment, concentration, challenge, and curiosity; flow would then influence the presence of time distortions [13]. Delving further into social context, it was found that the concentration and time distortion components of flow, but not enjoyment, were affected by working in a group of two compared to working individually in virtual worlds (within the social game platform Second Life) [14]. More in line with our work, several studies on flow and VR have been conducted. The previously mentioned study on Thumper compared flow states between VR and non-VR setups, finding that despite VR's technical immersion, both scenarios could lead to a flow state [11]. Within VR activities, a model ostensibly associates flow and playfulness, defined by a combination of intrinsic motivation, control, and freedom to suspend reality; this association then influences competence in the activity [15].

VR studies on time perception, however, are not limited to flow. VR itself affects our senses due to what is being conveyed through sensory channels, but also due to the devices used and the eventual physical discomfort we can get from it. Simply comparing the time perception of the same game in a VR versus a desktop setup leads to an underestimation bias for VR [16]. It also seems that time perception changes when bored or waiting in VR compared to real life [17]. In another study comparing time perception between VR and desktop while completing a task ranging from 30 s to 5 min, it was observed that if both situations yielded time overestimation, the VR scenario was overestimated more [18]. This seemingly contradicts the underestimation effect in [16] but may be due to VR affecting time perception in correlation to the task. However, technological immersion alone might not be a sufficient explanation, as walking in VR does not seem to affect time perception significantly [19]. In another experiment about zeitgeber on time perception while doing a task conducted both in VR and on desktop, no significant difference was observed regarding time perception, but there was a difference in task performance (with the VR participant performing better) [20]. The Thumper study also observed the effect of performing better in a VR setup compared to a non-VR one [11]. When it comes to the effect of emotional content, VR itself appears not to yield any difference to real life in time distortions when the emotional content is the same [21]. Employing VR entails possibly unique content, such as having movements represented by an avatar. Differences were observed between avatar and no-avatar conditions, where avatar presence in a retrospective paradigm leads to a significantly faster passage of time without affecting time estimation [22].

When considering VR and time perception, we can thus regard both the technological immersion, i.e., the effect of being in VR through its dedicated hardware, as well as the virtual environment stimuli and transformations that can be induced through VR. A specific stimulus type we want to employ in VR is rhythmic stimuli, which we already identified as having notable effects in a desktop scenario [1]. Rhythmic stimuli and music generally have a high potential to modulate one's time experience. With music, it was observed that higher tempos induce longer subjective time, but emotional valence decreases (but does not suppress) the effect of tempo and affects time perception. These effects on time perception might be due to their impact on arousal; interestingly, using a different orchestration (piano only or full orchestra) does not affect time judgment and pleasantness while affecting arousal [23]. On timing evaluations of instrumental excerpts of Disco songs (including estimation and judgment of time passage) over different tempos, it was observed that faster tempos were correlated to longer reproduction duration; however, no effect on estimations was observed alongside the necessity of a tempo difference of at least 20 BPM required for timing measurement differences to appear [24]. By varying cognitive load through tasks and arousal levels through music choice while keeping music tempo constant, it was found that (1) time is judged as passing faster under higher cognitive load (presence of a math task), (2) presence of a concurrent motor task (tapping the music's tempo) yield shorter subjective durations, and (3) for the motor task, for the same music tapping to half notes instead of eighth notes ended up with smaller time estimations and time passage rated as faster [25]. Regarding timing and spatial movement, rhythmic auditory stimuli (RAS) have been observed to improve motor performance when vision is unavailable [26].

Rhythm, however, is not only tied to music and audio. Concerning temporal judgments, audio was believed to be dominant over visuals [27]. However, a later study found visual stimuli dominance using Point-Light-Display (PLD) dance motions compared to simple audio tempos [28]. Participants were presented with both the dance motion and the audio tempo and had to give a globally fitting tempo. The result of this study suggests that under the right conditions, visual stimuli can dominate audio in terms of timing, which may be due to the quantity of temporal information on a sensory channel rather than the channel itself.

While our focus here is primarily on VR-related stimuli and associated in-environment monitoring, the physiological aspects of time perception are equally interesting.

Temperature is among the stimuli believed to be linked to arousal and affect time perception. Indeed, as shown by the literature review of Wearden and Penton-Voak [29], multiple previous works suggest "a parametric relation between the rate of subjective time and body temperature". Wearden and Penton-Voak further suggests that increases and decreases of the rate of subjective time that happens in temperature

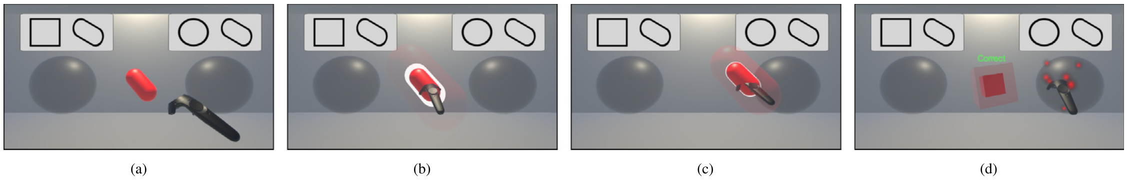

Fig. 1. Sorting example from trial task, in sequential order from left to right.

situations could be assumed to correspond to equivalent arousal level changes (with higher body temperature leading to arousal increase and thus time dilation, gradual cooling reducing arousal and therefore leading to time compression but cold stress leading to time dilation could be explained by an arousal increase).

Still on the topic of body temperature, with an experiment involving head-out water immersion conditions, Kingma et al. investigated the relation between time perception, reaction time, gastrointestinal temperature, heart rate and thermal sensation. From their observations, a change in the accuracy of an expanded judgment task was significantly associated with a change in gastrointestinal temperature, thermal sensation and heart rate. It was also observed that a change in response time was significantly associated with a change in gastrointestinal temperature. Finally, changes in time estimations were associated with a change in thermal discomfort or gastrointestinal temperature but not with any other measured physiological or subjective variables [30].

Dormal et al. investigated physiological arousal/heart rate and time processing in an offline arousal manipulation manner (separated from duration processing). According to the authors, the overestimation observed is not due to physiological arousal but to a distortion of memorized standard duration caused by time lag rather than a physiological arousal effect. Heart rate/arousal alone does not seem sufficient to explain time distortions at a supra-second scale, and other studies online (concurrently with duration processing) show that arousal manipulation's effect on time perception due to the allocation of attentional resources [31].

In the context of VR experiments, Cho et al. designed a stress-level classification algorithm using K-ELM based on physiological signals (Heart rate variability, skin conductance, measured by photoplethysmogram (PPG), electrodermal activity (EDA) and skin temperature (SKT)). In their presentation study, the algorithm had over 95% accuracy [32].

Emotional response seems correlated to pupil dilation, likely due to increased arousal. In this context, Chen et al. designed methods to use measure and normalize pupil dilation values depending on the brightness of the screen [33].

As for exercising and time perception and the relation of these two with arousal, Lambourne observed a leftward shift of temporal generalization gradients during exercise when compared to rest, indicating that the subject perceived intervals to elapse more slowly. They also observed no significant differences in the episodic timing or working memory tasks. These findings support the notion that exercise influences the internal clock similar to other arousal-inducing manipulations [34].

The stopped-clock illusion, or saccadic chronostasis, is a phenomenon that occurs when making a saccade (here referred to as an eye movement) to a silent clock. This causes the first–second feeling to last longer than the next ones [35,36].

Yarrow et al. [35] investigated through multiple experiments the effect of the duration of the saccade on chronostasis (using an incremental digital counter with the first value being displayed for a time between 400 ms and 1600 ms and the subsequent values being a constant 1s) if the illusion of chronostasis is due to eye movement or shift of the locus of visual attention. By comparing situations in which the participant had to either simply do a saccade to the target or shift their attention first towards the subject and then move their eyes, they observed what happens when moving the target of the saccade. The results of Yarrow et al.'s experiments suggest that the extra-time judgment of chronostasis depends on the time used to move the eye, that the shift of attention does not affect the length of chronostasis and regarding the target of the saccade moving, if the shift was noticed then no illusory effect was observed but if it was not noticed then the illusion had a magnitude between a control scenario and a full illusion scenario. It is then suggested that "chronostasis is an illusion occurring to fill in the perceptual gap during saccadic suppression" and "moving the target unpredictably during the saccade breaks the spatial continuity and thus, the illusion disappears" [35].

3. Experiment

We recruited 30 participants from a public, science-related event in Luxembourg City and from students and staff at the University of Luxembourg. The female-to-male ratio was 46.66%–53.33%, with ages ranging from 19 to 45 (mean age 25.63/median 24). Participants were received at the VR/AR Lab's test space at the University of Luxembourg and briefed about the experiment before the session, both verbally and via an informed consent form, in which we also collected demographic data. The study was approved by the local ethics committee in compliance with all ethical standards and guidelines. Among others, participants were allowed to abort immediately at any time, e.g., in case of cybersickness. After the setup and familiarization phase, the participants performed the tasks, with each participant able to take breaks between trials if desired. Each session lasted a total of approximately one hour.

3.1. Trial task and design

Participants had to complete trials in which they had to sort three-dimensional objects according to their shape (spheres, capsules, or cubes). As shown in the screenshot sequence in Fig. 1, the objects must be grabbed with a VR controller and dragged into one of the two larger sinks, which only accept the specific shapes displayed on a scoreboard above them (cf. Figs. 1(a)–1(c)). While cubes and spheres can only be placed in one of the sinks, capsules can go in either of the two. Once sorted, an object disappears with a small animation, indicating if the sorting is correct, and a new object to be sorted appears in the center of the virtual environment (cf. Fig. 1(d)).

The sorting attempt has a predefined duration unknown to participants. After this duration, the experiment ends with a questionnaire in which the participants are asked to estimate the time in seconds and rate on two Likert scales how fast the experiment felt and how tired they were (cf. Fig. 2).

Trials were subjected to conditions that were a combination of these parameters:

- *trial length*: how long the trial lasted in seconds (either 40, 50, or 60)
- *tempo*: the rhythm of the audio/visual stimuli if present in beats per minute (either 100, 140, or 180; forced to 0 without stimuli)
- *visual stimuli*: whether or not there are flashing visual pulses around the object
- *audio stimuli*: whether or not metronome click sounds are produced

Each participant went through all 30 possible combinations in random order.

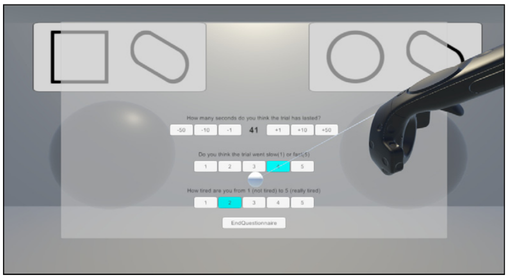

Fig. 2. Post-trial questionnaire (time estimation, time perception, fatigue).

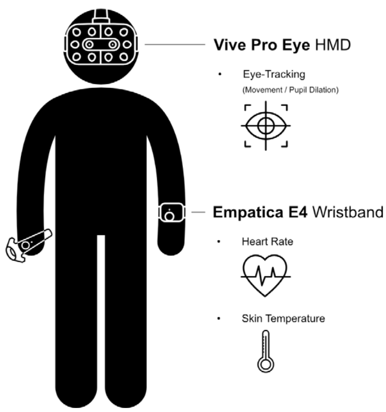

Fig. 3. Hardware used for its associated physiological data collection. Vive Pro Eye in the case of eye-tracking related data, Empatica E4 otherwise.

3.2. Technical specifications

Participants used the VIVE Pro Eye head-mounted display (HMD) with one of its controllers, allowing them to enter VR and control virtual objects with six degrees of freedom (DoF). This particular HMD also allows precise eye-tracking data to be collected during the trials, specifically on gaze and pupil dilation. During the experiments, participants were also instructed to wear an Empatica E4 wristband, as illustrated in Fig. 3, to record further physiological data; our experiment included heart rate variability and skin temperature. Our custom application reads the data by receiving messages from Empatica's E4 Streaming Server software. The experimental application itself was developed in Unity and used SteamVR. In addition to the application, OVR advanced settings were employed to adjust the participant's height position.

4. Analysis and results

4.1. General methods and data

As discussed previously, various data were collected from each trial; in this subsection, we will describe what data we effectively used for our analysis and how.

4.1.1. Disclaimed trials

Some trials were disclaimed from the data set depending on our notes during trials. The reasons for removing trials were:

- Misunderstanding of task controls by the participant
- Misunderstanding of trial questionnaire by the participant (verbally checked when giving incoherent values such as negative time estimation).
- Disturbance or interruption during the trial, either from the participants themselves (i.e., asking a question or talking during a trial) or external sources (i.e., technical issues, noise from a nearby room).

4.1.2. Variables/functions

Variables and functions extracted from trial data are:

- $t \in T$: trial identifier t from all available trials T.
- $p \in P$: participant identifier p from all available participants P.
- *trials(p)*: all trial identifiers of a participant p.
- *participant(t)*: participant identifier of a trial t.
- *correct(t)*: number of correct sorts at the end of trial t.
- *trialLength(t)*: length of trial t in seconds.
- *reportedLength(t)*: reported length of trial t in seconds.
- *reportedFatigue(t)*: self-reported fatigue of a participant after performing trial t in an ordinal scale from 1 to 5.
- *reportedSpeedPerception(t)*: subjective participant rating of trial t's speed in an ordinal scale from 1 (slow) to 5 (fast).
- *trialIndex(t)*: the index of trial t, indicates how many trials were performed before t and, therefore, global repetition from the experiment session.

Since *reportedSpeedPerception(t)* and *reportedFatigue(t)* are purely subjective questions to the participant, we can use these values directly. However *correct(t)*, the task performance direct variable depends on the participants' individual performance and *reportedLength(t)*, while the direct time estimation variable depends on both the trial's duration as well as the participants' individual representation of a second. Therefore, these two variables need to undergo a normalization process. Normalizing the performance variable (*correct(t)*) for our analysis goes through a three-step process involving the following extracted variables:

- *correctPerSecondTrial(t)*: average number of correct answers per second during trial t.
$$\frac{correct(t)}{trialLength(t)}$$
- *correctPerSecondParticipant(p)*: average number of correct answers per second of a participant during trials.
$$\frac{\sum_{t' \in trials(participant(t))} correct(t')}{\sum_{t' \in trials(participant(t))} trialLength(t')}$$
- *correctNormalized(t)*: amount of correct answers per second of trial t normalized with 1, i.e., the average number of correct answers per second of a participant among all performed trials.
$$\frac{correctPerSecondTrial(t)}{correctPerSecondParticipant(p)}$$

For time estimation (*reportedLength(t)*), we employed the following process:

- *secondBias(p)*: ratio of the total of seconds of a participant p's trials and the total reported time, defining what the participant considers a second.
$$\frac{\sum_{t \in trials(p)} trialLength(t)}{\sum_{t \in trials(p)} reportedLength(t)}$$
- *deltaTimePerception(t)*: averaged delta per second between reported time (accounting participant bias) and trial length of a trial t.
$$\frac{secondBias(p) * reportedLength(t)}{trialLength(t)} - 1$$

However, in addition to *deltaTimePerception(t)* we also use its absolute *abs(deltaTimePerception(t))* as it represents the magnitude of time perception delta of a trial.

4.1.3. Outcome variables

The specific variables relevant to the analysis performed in this study are:

- *deltaTimePerception(t)*: this variable represents a participant's variation in perception, i.e, the difference between reported time and trial time. A negative value indicates that the seconds of the trial *t* were reported as shorter than the other trial performed by this participant. A positive value means that the seconds were reported as longer.
- *abs(deltaTimePerception(t))*: instead of denoting how much longer or shorter a second is interpreted for a trial *t* compared to other trials performed by a participant, the absolute value represents the magnitude of the eventual time distortion.
- *correctNormalized(t)*: to simplify the analysis, we do not take into account incorrect answers to evaluate performance but only the amount of correct answers. A smaller number of correct answers compared to other trials would indirectly reflect the number of incorrect answers due to the time lost. Similar to a participant's variation in perception, this variable represents the variation in performance instead of the pure performance, with values < 1 indicating worse and > 1 better performances.
- *reportedSpeedPerception(t)*: the subjective interpretation of whether time drags or flies after performing trial *t* in an ordinal scale from 1 to 5.
- *reportedFatigue(t)*:reported fatigue should mostly depend on the participants, independent of trial parameters.
- *trialIndex(t)*: the index of trial *t*, indicates how many trials were performed before *t* by the participant and, thus, is an indicator of global repetition from the experiment session.

4.1.4. Parameters

The parameters used in this experiment are:

- *stimulusTrial(t)*: the type of stimulus used in trial *t*, possible values are: *None, VisualsOnly, AudioOnly, Both*.
- *hasAudioTrial(t)*: whether or not the trial *t* contains an audio stimulus.
- *hasVisualTrial(t)*: whether or not the trial *t* contains a visual stimulus.
- *hasStimulusTrial(t)*: whether or not the trial *t* contains any type of stimulus.
- *tempo(t)*: the tempo in beats per minute (BPM) of a trial *t*, possible values are: 0, 100, 140, 180. A tempo of 0 means that the trial had no stimuli.

4.1.5. Trial filters

When performing analyses, we may want to include only subsets of the trials to investigate specific effects. The filters used are:

- *filterAUDIOONLY*: considers only trials that only have an audio stimulus, is equivalent to saying *t where stimulusTrial(t) == AudioOnly*.
- *filterVISUALSONLY*: considers only trials that have a visual stimulus, is equivalent to saying *t where stimulusTrial(t) == VisualsOnly*.
- *filterBOTH*: considers only trials that have both audio and visual stimuli, is equivalent to saying *t where stimulusTrial(t) == BOTH*.

4.1.6. Physiological variables and functions

The raw eye-tracking data recorded and used in this study are the pupil position in the sensor area and the pupil dilation. These data were being recorded at every frame the application was running, so aimed at about 90hz frequency, as this is the headset's refresh rate. The pupil dilation was used for each trial as follows: recording errors were filtered (detected with values below 0), and then values were weighted depending on the time difference between it and the previously recorded data point. We thus collect this Pupil dilation and use the mean *PupilDilation.mean* and variance *PupilDilation.var*. The pupil position was used to derive what we call the "eye speed", which refers to the distance per second the pupils are traveling. Similar to pupil dilation, we filter the incorrect recordings (using the interquartile method) and then get weighted values from subsequent values in time. We thus also get the mean (*EyeSpeed.mean(t)*) and variance (*EyeSpeed.var(t)*) of this eye speed. While the Empatica E4 also allows for the collection of electrodermal activity (EDA/GSR) data and interbeat interval (IBI), we restricted our collection to heart rate and skin temperature. The reason is that the Empatica software already uses IBI to derive heart rate, and EDA/GSR would require minimal user movement as well as a 15-minute waiting phase, which was incompatible with the task. The temperature value uses the same filtering and weighting as the eye-tracking values and is used with both its mean and variance (*Temperature.mean(t)*, *Temperature.var(t)*); we also keep the temperature difference between the start and end of the trial as it shows the change across the trial (*Temperature.start-end(t)*. As heart rate is already derived from the Empatica's blood value pressure (BVP) over a longer period, we do not have enough data points. Thus, we take the last recorded value in a trial (*Heartrate(t)*) and filter trials where the value is not updated. Due to numerous technical issues with the Empatica hardware and software, 463 (44.88% of the data is retained) of the trials' heart rate data and 42 (95% of the data is retained) for the temperature have been removed.

4.1.7. Performed tests

Using the variables, parameters, and filters detailed in Section 4.1.2 to 4.1.6, we performed statistical tests on various data subsets to examine the effect of a stimuli presence on performance and time estimation with the following parameters:

- *stimulusTrial(t)*: ANOVAs on *deltaTimePerception(t)*, *abs(deltaTimePerception(t))* and *correctNormalized(t)* to see if any significant difference appears between possible stimuli situations.
- *hasAudioTrial(t), hasVisualTrial(t), hasStimulusTrial(t)*: t-tests on *deltaTimePerception(t), abs(deltaTimePerception(t))* and *correctNormalized(t)* to see if there is an effect on the presence or absence of a specific stimulus (since our t-tests are not pairwise, no t-test corrections have been performed).

To examine this effect on time judgment, we did the following as the time judgment variable is ordinal:

- *stimulusTrial(t)*: Kruskal–Wallis test on *reportedSpeedPerception(t)* to see if any significant difference appears between possible stimuli situations.
- *hasAudioTrial(t), hasVisualTrial(t), hasStimulusTrial(t)*: Wilcoxon test on *reportedSpeedPerception(t)* to see if there is an effect on the presence or absence of a specific stimulus.

In order to observe the effect of tempo on performance and time estimation, we performed ANOVAs between *tempo(t)* and the variables *deltaTimePerception(t), abs(deltaTimePerception(t))*, and *correctNormalized(t)*. The ANOVAs were also repeated across the filters *filterAUDIOONLY, filterVISUALSONLY*, and *filterBOTH* to see if differences in tempo appear only within stimuli conditions. For the effect of tempo on time judgment (*reportedSpeedPerception(t)*), we again replaced the ANOVAs with Kruskal–Wallis tests, including the repeated ones under filters. Correlations between time estimation variables (*deltaTimePerception(t), abs(deltaTimePerception(t))*) and performance (*correctNormalized(t)*) were investigated with Pearson tests. As the time judgment variable (*reportedSpeedPerception(t)*) is ordinal, Pearson tests were used to determine its correlation with time estimation variables (*deltaTimePerception(t), abs(deltaTimePerception(t))*) and performance (*correctNormalized(t)*). The confounding effect of fatigue was investigated by considering *reportedFatigue(t)* both as a nominal and as an ordinal variable; the former allows us to eventually observe differences

Table 1
Subsets for which a significant p-value (•) or tendency (∘) is observed for a combination of stimulus dimension and outcome variable group.

	Performance	Time Estimation	Time Judgment
Stimuli Presence	–	–	• Full Set • Trials 11-20 ∘ Fatigue Levels 1-2 • Fatigue Levels 3-4-5
Stimuli Type	• Full Set • Trials 1-10 • Trials 21-30 • Fatigue Levels 3-4-5	–	• Full Set • Trials 11-20 • Fatigue Levels 1-2
Stimuli Tempo	∘ Full Set • Trials 11-20 • Fatigue Levels 1-2	• Full Set ∘ Trials 11-20 ∘ Fatigue Levels 3-4-5	• Full Set • Fatigue Levels 1-2

Table 2
Physiological data for which a significant p-value (•) or tendency (∘) is observed for a target variable.

	Fatigue	Performance	Time Estimation	Time Judgment
Eye-Tracking	• EyeSpeed.std ∘ PupilDiameter.mean	• PupilDiameter.mean	∘ PupilDiameter.mean (abs)	• PupilDiameter.std
Temperature	• Temperature.mean • Temperature.std • Temperature.start-end	–	• Temperature.mean (abs)	• Temperature.std • Temperature.start-end
Heartrate	–	–	• Heartrate (signed)	–

between specific ratings and has been considered through ANOVAs with time estimation variables (*deltaTimePerception(t)*, *abs(deltaTimePerception(t))*), and performance (*correctNormalized(t)*), the latter was considered through Spearman tests on time estimation variables, performance, and time judgment (*reportedSpeedPerception(t)*). Regarding trial index's confounding effect, *trialIndex(t)* is ordinal data, so it has been investigated with Spearman tests on time estimation variables (*deltaTimePerception(t)*, *abs(deltaTimePerception(t))*), performance (*correctNormalized(t)*) as well as time judgment (*reportedSpeedPerception(t)*). Each ANOVA with a p-value below 0.1 would lead to a subsequent Tukey HSD, Kruskal–Wallis tests would lead to subsequent paired Wilcoxon tests. Table 1 provides an overview of the different effects per subset with a significant p-value or tendency, while each is discussed in detail in the following sections. For physiological data, all retained outcome variables (*PupilDilation.mean(t), PupilDilation.var(t), EyeSpeed.mean(t), EyeSpeed.var(t), Temperature.mean(t), Temperature.var(t), Temperature.start-end(t), Heartrate(t)*) have been investigated in correlation tests with the time estimation variables (Pearson tests with *deltaTimePerception(t)* and *abs(deltaTimePerception(t))*), the performance outcome variable (Pearson test with *correctNormalized(t)*), time judgment (Spearman test with *reportedSpeedPerception(t)*), and fatigue (*reportedFatigue(t)*).

Table 2 provides an overview of the observed correlations related to our physiological variables. The complete data from our tests, including confidence intervals and average values, are available online [37].

4.2. Across all trials

4.2.1. Effects of stimuli on performance

One of the aims of this study is to investigate the effects of stimuli on task performance. Looking at performance across all trials, the ANOVA between task performance (*correctNormalized(t)*) and stimulus type (*stimulusTrial(t)*) revealed no significant difference. However, when performing a t-test between task performance (*correctNormalized(t)*) and the presence of visual stimuli (*hasVisualTrial(t)*), a significant difference (p=0.027) can be observed alongside decreased performance when a visual stimulus is involved, as the mean with the stimulus is lesser than without. No effect is observed when considering the t-test with the presence of an audio stimulus (*hasAudioTrial(t)*,$p = 0.167$) or any stimulus (*hasStimuluslTrial(t)*,$p = 0.431$). Therefore, we can only observe a decrease in performance due to the presence of a visual stimulus but no effect on performance from the sole presence of any or an audio stimulus. When stimuli have the dimension of type, they also have the dimension of tempo. The ANOVA between participant performance (*correctNormalized(t)*) and stimuli tempo (*tempo(t)*) generally finds no effect of tempo. However, tempo might have an effect under a specific stimulus type. Therefore, we performed the same ANOVA but only considering subsets of data where trials contained either audio stimuli only (*filterAUDIOONLY*), visual stimuli only (*filterVISUALSONLY*), or both simultaneously (*filterBOTH*). We can then observe a tendency when trials have audio stimuli only (*filterAUDIOONLY*,$p = 0.088$, F=2.461). The Tukey HSD of this ANOVA reveals that the effect is significant between 180-100 (p=0.070, *diff*=0.04), with a diff value indicating that the faster tempo leads to better trial performance with audio stimuli only. In the absence of interference from visual stimuli, the faster tempo for audio stimuli may have implicitly stimulated the participant to sort objects faster.

4.2.2. Effects of stimuli on time estimation

Similar to performance, we evaluated the effect of stimuli type and tempo on time estimation variables. Likewise, ANOVAs were effectuated regarding the type of stimuli (*stimulusTrial(t)*) and tempo (*tempo(t)*) on both the normalized time estimation error (*deltaTimePerception(t)*) and its magnitude (*abs(deltaTimePerception(t))*). The only significant result is a tendency between time estimation error (*deltaTimePerception(t)*) and tempo (*tempo(t)*) (p=0.073, F=2.328). Tukey's HSD of this ANOVA reveals that the effect is a tendency only between tempi of 180 and 140 (p=0.65, *diff*=0.067), with trials under a tempo of 180 being rated with a longer time per second than trials under a tempo of 140. We performed similar ANOVAs involving tempo, considering subsets of data in which the trials had either only audio stimuli (*filterAUDIOONLY*), only visual stimuli (*filterVISUALSONLY*), or both at the same time (*filterBOTH*). The only significant result comes from the ANOVA between time estimation error (*deltaTimePerception(t)*) and tempo (*tempo(t)*) across trials within the AudioOnly condition (*filterAUDIOONLY*) (p=0.039, F=3.288), where the Tukey HSD follow-up reveals a near-significant difference between tempi of 140 and 100 (p=0.051, *diff*=-0.106) and a near tendency between 180 and 140 (p=0.107,*diff*=0.091). This means that in trials with only an audio stimulus, trials with a BPM of 140 were evaluated as faster than others, which contradicts the analysis under all types of stimuli. This contradiction may indicate the confounded effect of tempo in time perception depending on stimuli types. Finally, t-tests between our time estimation variables (*deltaTimePerception(t), abs(deltaTimePerception(t))*) and the presence of audio stimuli (*hasAudioTrial(t)*), visual (*hasVisualTrial(t)*), or any (*hasStimuluslTrial(t)*) yielded no significant result, meaning no effect of any type of stimuli present can be observed on time estimation here.

4.2.3. Effects of stimuli on time judgment

As the time judgment variable (*reportedSpeedPerception(t)*) is ordinal, we produced Kruskal–Wallis tests between the variable and the type of stimuli (*stimulusTrial(t)*) and tempo (*tempo(t)*). In the case of the test between time judgment (*reportedSpeedPerception(t)*) and the type of stimuli (*stimulusTrial(t)*), we can see a significant effect (*p=0.016, chi2=10.267*); however, a follow-up paired Wilcoxon test reveals statistical difference only between the stimulus "None" and each of the other stimuli types (*p=0.048 for None-AudioOnly,p = 0.015 for None-Both,p = 0.015 for None-VisualsOnly*). As for the test on tempo (*tempo(t)*), we observe another correlation (*p=0.001, chi2=15.432*) that, after a paired Wilcoxon, shows significant differences between 0–140 (*p=0.006*), 0–180 (*p=0.005*), 100-180 (*p=0.031*) as well as a tendency between 100-140 ((*p=0.064*) and a near-tendency between 0–100 (*p=0.12*). These two tests highlight a significant difference in time judgment depending on the presence of any stimuli (both by the differences from the "None" stimulus in the first test and the "0" BPM tempo in the second, which correspond to trials without stimuli). This is also verified by the Wilcoxon test between time judgment (*reportedSpeedPerception(t)*) and the presence of any stimulus (*hasStimuluslTrial(t)*) (*p=0.003, mean(TRUE)>mean(FALSE)*); considering the mean values, we can say that the presence of a stimulus has a significant impact in making a trial judged as passing faster than one without any. The same test has been done on the presence of audio (*hasAudioTrial(t)*) (*p=0.323*) and visuals (*hasVisualTrial(t)*) (*p=0.023, mean(TRUE)>mean(FALSE)*), meaning no significant difference in the presence or absence of an audio stimulus is observed but a fast-inducing effect is observed on the presence of a visual stimulus is recorded. In the case of tempo, the results of the paired Wilcoxon discussed earlier also indicate a significant between 100BPM and other (non-0) tempi across all types of stimuli. However, running the same Kruskal–Wallis test under subsets on "AudioOnly" trials (*filterAUDIOONLY*), "VisualOnly" trials (*filterVISUALSONLY*), and trials with both (*filterBOTH(t)*) highlights a significant difference only across trials with both stimuli (*filterBOTH(t)*) (*p=0.027, chi2=7.2343*) meaning that meanwhile tempo may have an effect across all stimuli, that effect might only be due to the combined stimuli scenario. Follow-up paired Wilcoxon tests indicate a significant difference between 100–180 (*p=0.032*) and 140-180 (*p=0.091*), the same conclusion as the tests without subsets.

4.2.4. Correlations between outcome variables

Spearman tests were used to investigate the correlation between outcome variables when time judgment (*reportedSpeedPerception(t)*) was involved as the data is ordinal; otherwise, Pearson tests were used. When comparing time estimation error (*deltaTimePerception(t)*) and performance (*correctNormalized(t)*), we see no correlation (*p=0.218,cor=-0.043*), but we see a significant negative correlation with the magnitude of time estimation error (*abs(deltaTimePerception(t))*) (*p=0.016, cor=-0.083*). This means the performance is correlated to the magnitude of time estimation errors but not to the direction; in other words, participants may generally be more error-prone in their estimations depending on their performance. When it comes to time judgment (*reportedSpeedPerception(t)*), it is negatively correlated to time estimation errors (*deltaTimePerception(t)*) (*p=1.724e−13, rho=-0.251*) and positively correlated to the magnitude of said error (*abs(deltaTimePerception(t))*) (*p=0.002, rho=0.107*). This means that the bigger the error, the faster the time is perceived, and underestimated trials are rated as passing faster. As for time judgment (*reportedSpeedPerception(t)*) and performance (*correctNormalized(t)*), better performance is associated with faster passing trials (*p=1.265e−05, rho=1.50*).

4.2.5. Confounding effect of fatigue

While running the experiment, we noticed that participants were often exhausted at the end of the session. As exhaustion affects time perception and performance, we verified if it affected our outcome variables. For its effect on performance, a Spearman test between performance (*correctNormalized(t)*) and fatigue *(reportedFatigue(t))* reveals a significant correlation *(p=2.924e−08, rho=0.190)*. By considering the fatigue variable *(reportedFatigue(t))* nominal and performing an ANOVA with performance (*correctNormalized(t)*), we retrieve this correlation (*p=1.68e−11, F=14.56*). Subsequent Tukey HSD reveals that fatigue values of "3,4,5" are statistically different from values of "1,2" as the *p*-value is below 0.001 in all these situations. Other situations (i.e., "3-4", "1-2", …) have a *p*-value above 0.48. As for time estimations, signed error (*deltaTimePerception(t)*) is not correlated if we look through a Spearman test *(p=0.191, rho=0.045)*, but we retrieve statistical differences with the ANOVA *(p=0.004, F=3.86)*. Subsequent Tukey HSD indicates statistical differences between "3-2" *(p=0.001)*, "5-2" *(p=0.045)* and a tendency between "4-2" *(p=0.06)*. No correlation is observed for the absolute error (*abs(deltaTimePerception(t))*) with both the Spearman test (*p=0.842, rho=0.107*) and the ANOVA (*p=0.893, F=0.277*); however, it is observed for the Spearman test with time judgment (*reportedSpeedPerception(t)*) (*p=2.88e−11, rho=0.227*). From the results of the ANOVAs involving performance (*correctNormalized(t)*) and time estimation (*deltaTimePerception(t)*), we can identify two groups of reported fatigue values: "1-2" and "3-4-5". We thus decided to perform the same tests on subsets of our data according to these two groups on Section 4.4.

4.2.6. Confounding effect of trial index

Similarly to fatigue, repeated trials can affect both performance and time perception due to learning effects and repetition. We thus evaluated correlations through Pearson tests between the number of a trial across the session *(trialIndex(t))* as a continuous variable and time estimation variables (*deltaTimePerception(t), abs(deltaTimePerception(t))*) as well as performance (*correctNormalized(t)*). Regarding performance (*correctNormalized(t)*), the test reveals a correlation (*p=2.626e−14, cor=0.259*), which indicates a learning effect. Trial repetition also seems to affect time estimation as we retrieve a significant correlation with the signed time estimation error (*deltaTimePerception(t)*) (*p=1.407e−04, cor=0.131*) and a tendency with its absolute (*abs(deltaTimePerception(t))*) (*p=0.089, cor=-0.059*). Therefore, we decided to investigate different phases (beginning, middle, end) in the experiment defined by three subsets of the data based on the trial index, as shown in Fig. 4 and discussed in detail in the following Section 4.3.

4.3. Trial index subsets

Due to our results on the confounding effects on trial index as described in Section 4.2.6, we decided to investigate three subsets of the data based on the trial index with steps of ten (1–10, 11–20, 21–30). For each subset, we performed all the tests like on the full trial set, which are available for download [37] and detailed in Section 4.1. However, the normalization process only considered the targeted subset when using the sum of data on trials. We go through each subset in the following subsections, focusing on the significant results.

4.3.1. Trials 1–10

This subset corresponds to each participant's first ten trials of the experiment, constituting a discovery phase. Regarding stimuli effects on performance, the results indicate a positive effect of audio stimulus presence (t-test performance∼audiopresence, *p=0.016*; TukeyHSD performance∼stimuli, *p=0.099* on worse performance between visuals∼both). This can be linked to the results across the entire experiment as we have seen a negative impact of visual stimulus presence and a tendency for trials with just audio to have their performance led by the tempo (see Section 4.2.1). This difference

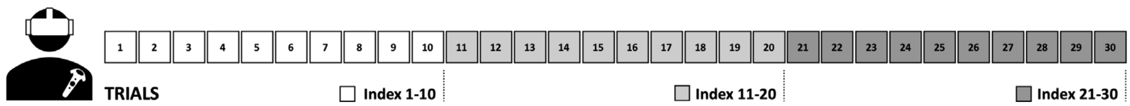

Fig. 4. Trial subset allocation for a participant by index.

might be due to a learning effect on the trials where the participants are not proficient enough to lose enough performance from visuals but may be eased by the presence of any leading audio rhythmic stimulus for this repetitive task. As for stimuli on time estimation, here we only observe a potential novelty effect on trials without stimuli as they are rarer than trials with any stimuli (t-test time estimation error stimuli presence, *p=0.093*). The most notable difference with the analysis on all trials regarding this aspect is the absence of the effect of tempo on the time estimations. Surprisingly, no effect of stimuli concerning time judgment is observed from any of our tests. When it comes to correlation between performance and time estimations through Pearson correlation tests, contrary to the full set of trials, we observe a (negative) correlation with the signed time estimation error (*p=0.025, cor=-0.142*) but not on the absolute error. Regarding time judgment concerning both time estimations and performance, we lost the correlation with the absolute time error; however, we retrieve the positive correlation from the Spearman tests with performance (*p=0.095, rho=0.106*) and the signed time estimation error (*p=-0.183, rho=-0.184*). Finally, regarding results on confounding effects of trial index and fatigue, we retrieve correlations of fatigue on performance and time experience, the correlation of performance and trial index, but none between trial index and time experience. This means that with this subset, we should have isolated an experiment phase based on trials for time perception but not for performance, which is expected as the participants were likely learning how to perform better during the first few trials.

4.3.2. Trials 11–20

This subset corresponds to each participant's ten trials in the middle of the experiment, representing a neutral phase as they no longer learn the task while not being in the experiment long enough to be bored. Regarding performance and stimuli, in this trial, we observe a performance increase from higher tempo within trials using combined stimuli (Tukey HSD 180-100, *p=0.044, diff=0.055*; 180-140, *p=0.075, diff=0.053*) As for the effect of stimuli on time estimation, we only observed a tendency between tempo 140-100 across trials solely using a visual stimulus (Tukey HSD 140-100, *p=0.100, diff=-0.083*). When it comes to time judgment, strong evidence shows that under this subset, the presence of any stimulus heavily alters it (paired Wilcoxon on time judgment and type of stimuli, *p<0.002 for all situations with "None"*; paired Wilcoxon on time judgment and tempo, *p<0.02 for all cases with "0"*; Wilcoxon time judgment and stimuli presence, (*p=8.939e–05, higher mean with stimulus*). We also observe an effect of visual stimulus specifically with the same Wilcoxon test on visual stimulus presence (*p=0.041, higher mean with stimulus*) but not on audio presence. Therefore, the effect of stimuli on time judgment in this subset is consistent with the full set regarding the effect of the present stimuli type, but we lost the effect of the tempo. This time, no correlation has been observed between time estimation and performance. However, we retrieve the time judgment correlations from the full set with Spearman tests on the performance (*p=0.004, rho=0.168*), the signed time estimation error (*p=1.988e–04, rho=0.215*), and its absolute (*p=4.20e–05, rho=-0.236*). Finally, on confounding effects, we find a correlation between fatigue and time judgment, which is expected, yet we also see a correlation tendency between trial index and signed time estimation error through a Pearson test (*p=0.071, cor = 0.105*). Still, the absence of correlation with performance indicates a proper subset division on the trial index.

4.3.3. Trials 21–30

This subset corresponds to each participant's last ten trials, representing the end of the experiment and, thus, a phase where the participant is possibly tired or bored. Here, investigation of performance suggests that when there are stimuli, the presence of visual stimulus leads to worse task performance (Tukey HSD performance between stimuli modes "Both" and "AudioOnly", *p=0.010, diff=-0.045*; "VisualsOnly" and "AudioOnly", *p=0.026, diff=-0.041*; t-test on performance and visual presence, *p=1.449e–04, lower mean when the stimulus is present*). Under this subset, nothing significant has been observed in the relation between stimuli (type or tempo) and time perception (time estimation and judgment). Similarly to the previous set (Trials 11–20), no correlation between time estimation and performance is observed. The relation from Spearman tests between time judgment with both performance and time estimation is similar to the subset at the beginning of the experiment (Trials 1–10), where performance is positively correlated (*p=0.002, rho=0.176*) and time estimation error is negatively correlated (*p=4.795e–04, rho=-0.202*) but the absolute error is not. On confounding effect, while finding effects of fatigue on time judgment as expected, unfortunately, we see tendencies on the effects of the trial index on both time estimation error (*p=0.067, cor=0.107*) and task performance (*p=0.073, cor=-0.104*) from Pearson tests. This may indicate a transition between phases of boredom and tiredness relative to the time spent in the experiment.

4.4. Fatigue subsets

Having obtained the results on the confounding effects of fatigue described in Section 4.2.5, we decided to investigate two subsets depending on the participants' answers on fatigue, one for fatigue at 1 or 2, and one for fatigue at 3, 4, or 5 (see Fig. 5). Similar to the previous subsets, for each, we performed all tests on the complete trial set, which can be found online [37], and the same modification on the variable normalization process by only considering the targeted subset when using the sum of data on trials. In the following sections, we will again focus exclusively on significant test results and will not re-elaborate the methodology.

4.4.1. Fatigue levels 1–2

This subset corresponds to the participant experiencing "low" fatigue. First, concerning performance and stimuli tempo, we can observe the lesser performance of stimuli with a tempo of 100 (Tukey HSD on performance and tempo between 140 and 100 bpm, *p=0.056, diff=0.065*; 180 and 100 bpm, *p=0.020, diff=0.081*). This finding can be aligned to results from the full set (performance dependent on tempo for audio stimuli) and trials 11–20 (180 bpm leading to better performance under combined stimuli). No general effect of stimuli type on performance is observed, either from the specific situations possible or the presence of a modality. This subset yielded no significant insights regarding stimuli dimensions (type and tempo) and time estimation. Regarding time judgment and tempo, however, we observe significant differences between 180 and 100 bpm across all stimuli (Paired Wilcoxon on time judgment and tempo between 100 and 180 bpm, *p=0.023*) as well as an effect of the presence of 180 bpm (Wilcoxon between 0 and 180, *p=0.025*). When considering only trials with combined stimuli, the paired Wilcoxon shows a significant

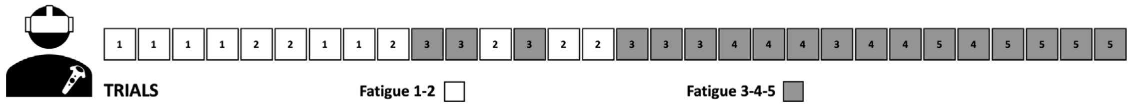

Fig. 5. Example trial subset allocation for a participant by fatigue.

difference between 180-100 (*p=0.017*) and a tendency between 100–140 (*p=0.094*), which is consistent with the time judgment effects results on the complete set of trials. As for time judgment and stimuli type, we observe another consistent result from the full set as stimuli tend to be judged faster when there is any stimulus (Wilcoxon on the presence of any stimulus, *p=0.073*) or if there is at least a visual one (*p=0.020*). Looking for a correlation between performance, time estimation, and time judgment yielded similar results to the subset of Trials on index 1–10. With a tendency of a negative correlation (*p=0.059, cor=-0.141*) from Pearson between time estimation error and performance, a significant negative correlation (*p=0.022, rho=-0.171*) out of the Spearman between time estimation error and time judgment, a positive one (*p=0.033, rho=0.157*) between performance and time judgment, but no correlations from the absolute time estimation error. Confounding effects of trial index on performance (Pearson test, *p=0.086, cor=0.129*) are similar to trials 1–10, which is not surprising as early trials probably are low fatigue trials. A confounding effect of fatigue is not observed for both time estimation and judgment; however, we can observe it for performance (Spearman test, *p=0.037, rho=0.157*; Tukey HSD (more of a t-test considering we have two values in this subset), *p=0.048, F=3.977, diff2-1=0.046*). We can assume that higher fatigue trials in this subset would occur after the learning phase when the participant is more proficient.

4.4.2. Fatigue levels 3–4–5

This subset corresponds to the participant having a higher fatigue level. Concerning performance and stimuli type, like for the full set and trials 21–30, we observe a negative impact from the presence of visual stimuli (TukeyHSD on stimuli type and performance between VisualsOnly and Audio Only, *p=0.080, diff=-0.024*; t-test between performance and presence of visuals, *(p=0.020)*). Some observations converge towards contextual effect on tempo depending on the type of stimulus of the trial (TukeyHSD on absolute time estimation error between tempos 140 and 180 for audio trials, *p=0.086, diff=0.1*; TukeyHSD on signed time error between tempos 140 and 180 across all, *p=0.080, diff=0.087*). As for time judgment and stimuli, we only found evidence indicating an effect of general stimulus presence (Paired Wilcoxon on stimuli type and time judgment, *p<0.07* for pairs involving "None"; Wilcoxon on time judgment and stimulus presence, *p<0.011*; paired Wilcoxon on tempos, *p=0.032*) between 0–140 and *p=0.040* for 0–180). Looking for a correlation between performance, time estimation, and time judgment yielded similar results to the full set of trials. From Spearman tests with time judgment, we retrieve the negative correlation with the time estimation error (*p=2.217e–11, rho=-0.256*), the positive correlation with the absolute error (*p=4.366e–04, rho=0.136*) and with the performance (*p=3.915e–05, rho=0.159*). We do not retrieve the significant *p*-value on the Pearson test between performance and absolute error, but a near-tendency (*p=0.110, cor=-0.062*). While we do not observe a confounding effect from the ANOVAs between fatigue and time estimation variables, we see a tendency (*p=0.089, rho=0.066*) from the Spearman test on the absolute time estimation error. The effect of fatigue is more pronounced on performance (Pearson test, *p=0.029, rho=0.085*) and from the subsequent Tukey HSD of the ANOVA (*p=0.064, F=2.768*) the difference appears to be between 3–5 (*p=0.050, diff=0.023*). Fatigue also seems to significantly affect time judgment (*p=2.185e–08, rho=0.215*). In this subset, the fatigue levels 3 and 5 may be significantly different on both performance and time judgment; however, this is apparently due to the normalization on the subset and was not observable across all trials. The confounding effects of the trial index observed from Pearson tests are similar to those of the full set, which is not too surprising as the subset is rather large and was not made to minimize the effect of the index.

4.5. Results on physiological data

As stated in earlier sections, one of our objectives is to investigate the interrelationship of physiological measures, fatigue, task performance, and time perception. To this end, we systematically performed a Pearson test between all extracted values discussed in Section 4.1.6 and *deltaPerception(t), abs(deltaPerception(t))*, and *performance(t)*, and a Spearman test with these values and both *subjFatigue(t)* and *subjTimeRating(t)*. Contrary to the analysis on trial parameters, we could not perform this analysis under the different subsets due to the amount of data lost on the Empatica data and because we are trying to detach physiological data from the activity itself. Looking at the correlation on subjective fatigue, we find correlations on the eye speed variance (*p=9.43e–5, rho=0.134*), all temperature values (mean *p=0.013, rho=0.088*; variance *p=1.38e–10, rho=-0.225*; start-end *p=0.005, rho=-0.099*) and a tendency for pupil diameter (*p=0.097, rho=-0.057*). Regarding temperature, we can observe correlations between the temperature mean and the magnitude of time estimation error (*p=0.029, cor=-0.077*), between temperature variance and subjective time rating (*p=2.11e–5, cor=-0.150*), as well as between the subjective time rating with temperature difference at the end and start of the trial (*p=0.001, cor=-0.114*). Pupil diameter also has a tendency on its mean with the magnitude of time estimation error (*p=0.088, cor=0.059*) and a significant correlation with participant performance (*p=0.023, cor=-0.079*). However, while the mean pupil diameter is tangled between multiple confounded variables, its variance only correlates to the subjective time rating (*p=1.71e–5, cor=0.148*). As for the heart rate, we only observe a tendency with the signed time error (*p=0.071, cor=-0.130*), which, while not being a significant *p*-value, is, with the variance of pupil diameter, one of the only values that are not also correlated with fatigue.

5. Discussion

We conducted a VR experiment in which participants repeatedly performed a simple sorting task subjected to different stimuli conditions. This allowed us to gather numerous data, including information related to task proficiency, subjective data from questionnaires, and physiological data. The overarching goal was to explore relationships between time experience, task performance, environment/stimuli conditions, and physiological cues. However, we found that for some of the results on the complete data set, it was necessary to investigate closer multiple subsets, which we will discuss together with their implications for VR application design. Looking at the entire data set, we can observe specific effects of stimuli type and tempo on different aspects of time perception and performance, as well as some interesting correlations between those variables, which appeared to be also heavily impacted by the trial index and fatigue through the experiment. Therefore, we defined subsets of data based on the trial index and difference values from ANOVAs for fatigue. As indicated in Table 1, we can observe effects of stimuli presence, type, and tempo on performance and time experience depending on the subset. Some of the data and correlations align between subsets while others do not, which may indicate contextual effects of stimuli on performance and time experience depending on task repetition and fatigue.

5.1. Observations on task performance and stimuli

A central result from the analysis of the complete set of trials is how the presence of visual stimuli negatively impacts task performance. This is coherent with our previous study and is to be anticipated as the task requires visual attention, and those stimuli may be disturbing. However, within subsets, this result is observed only for later trials (index 21–30) and high fatigue (fatigue 3-4-5). Surprisingly, we see a positive effect on performance from the presence of audio but only in the early trials (index 1–10) and no effect of stimuli type presence in between (index 11–20). This could be interpreted as the disturbance of visual stimulus not being impactful enough when one is learning the task or not physically tired. We can also interpret the presence of audio stimuli as beneficial for this task only when the participant is in a learning phase. Effects of tempo are observed on trials within trials with only the audio stimulus when considering all trials and within trials with combined stimuli within the subset of trials from index 11–20. In both cases, the faster tempo led to faster performance, which indicates an invitation to go faster in the task from the faster stimuli; however, the participant's interpretation depends on the context.

5.2. Observations on time estimation and stimuli

Time estimation variables are defined from the difference between (normalized) participants' estimation of time taken for a trial and the actual time of a trial; we thus talk about the time estimation error and its absolute, which represents the magnitude of error regardless of whether the participant under- or overestimated the length of a trial. A global effect of tempo can only be observed with the complete set of trials between 180–140 (with 180 being overestimated). As for differentiation within stimuli situations, we see a time estimation error difference on the audio stimuli for the entire trial set and high-fatigue trials, and an absolute error difference in visuals for trials 11–20 as well as on combined stimuli for high fatigue. These results show a tendency of the 140 bpm tempo leading to fewer (absolute) estimation errors and being underestimated compared to 100–180. Another overestimating effect from stimulus presence is observed for trials 1–10. Overall, we also observe context-dependent effects of stimuli as the type of stimuli will affect one's time perception differently depending on the index or fatigue.

5.3. Observations on time judgment and stimuli

Time judgment or time passage refers to the subjective evaluation of a participant on whether they think a trial is going by fast or slow. It differs from time estimation in that the participant gives their subjective feeling about the time spent, whereas time estimation is an attempt by the participant to be objective about time. Time judgment has semi-constant results of the presence of any stimuli inducing faster perception; this is observed across all trials for both subsets on fatigue and the subset on stimuli 11–20. We can also observe a specific fastening effect of visual stimuli on all these sets affected by the presence of any stimuli except for the high fatigue one. The absence of these observations on subsets of trials either at the beginning or at the end of the experiment might indicate that the participant needs to get used to and, over time, gets too used to the presence of stimuli to be noticeable, regardless of fatigue levels. Another effect observed only on the complete set and for low fatigue is a difference between tempo in general and within trials with combined stimuli.

5.4. Time judgment, time estimation, and performance balance

Two correlations were consistent across all sets: a negative correlation between time estimation error and time judgment and a positive correlation between performance and time judgment. The first means that a subjective faster trial reflects under-estimation of time. The second means that when the participant rated the trial as faster than usual, they would perform better. This could directly be tied to the notion of flow as two elements of flow states are the challenge-skill balance and time transformation. The similarity between time estimation error and time judgment is indicative that our time transformation was a general time experience shift and not a side effect of disorientation (i.e., a participant judging a trial as fast because they thought it was a higher amount of time that actually passed). Among low fatigue and early trials, we also retrieve a negative correlation between performance and performance, reinforcing the flow approach. The time estimation error magnitude and performance are negatively correlated under all trials and high fatigue, which means that possibly, in a specific context, higher time transformation was generally detrimental to performance. However, this is against the flow definition, and combined with previous observations, it may imply that we are approaching flow states only with time transformations that are an underestimation. We also observed positive correlations between this magnitude and time judgment with all trials, the 11–20, and high fatigue subsets, which could be interpreted as the presence of any time transformation potentially leading to faster time passage in general.

5.5. Observations on physiological values

The results on physiological values are strongly affected by fatigue, as most extracted values correlate with it. Overall, our results imply that the same physiological outputs may represent time estimation and feeling of time passage differently. However, we cannot overlook how fatigue confounds with all other target variables of time perception and performance. So, while we do observe correlations with time perception variables on the mean of pupil dilation and temperature-related variables, the fact that these physiological values are also correlated with fatigue weakens their significance. While we can consider these values promising for assessing one's time experience, especially in the case of body temperature, additional research designed around tasks that are not fatigue-inducing would benefit this. These findings may also allow for looking at the problem from the other side, as fatigue is easier to assess than time perception through noninvasive sensors such as eye-tracking and body temperature; due to its influence, fatigue may be used as a proxy for time perception. However, we have two variables that are not observed as correlated with fatigue, which are correlated with time perception measurements: variance of pupil diameter (significant with subjective time rating) and heart rate (tendency with time estimation error).

5.6. Limitations

It is important to remember that the effect of the rhythmic stimuli in our experiment is contextualized in the particular scenario of our sorting task. We can also see some limits from the confounding effects of task familiarity and fatigue, and even with the subsets, which unfortunately implies using less data and thus having lower statistical power relevance (especially in the case of low fatigue), we can isolate the effect of at most one confounding effect but not both at the same time. In the case of physiological data, as shown in Table 3, the inter-correlation between fatigue, time estimation, time passage, and performance limits the scope of these findings, as discussed in Section 5.5. Individual per-participant differences are also to be considered, as through casual talks with the participants, we know of varying degrees of VR experience between participants; however, this data was not recorded and is thus not included in our analysis. It should also be mentioned that due to the fact that the self-assessment of fatigue had to be answered quickly in this study, standard tests such as the NASA Task Load Index could not be employed. Consequently, we cannot distinguish between mental and physical fatigue, which could be considered in future studies.

6. Conclusion

In this article, we used a simple sorting task to explore how rhythmic stimuli affect time experience and task performance in VR. We

Table 3
Results of correlation tests on investigation of cofounding factors. With annotation of significant p-values (•).

	deltaPerception	Abs (deltaPerception)	performance	subjTimeRating	subjFatigue
deltaPerception	–	–	P = 0.2177 Cor = −0.04257087	• P = 1.724e−13 Rho = −0.2505479	P = 0.1918 Rho = 0.04507991
Abs (deltaPerception)	–	–	• P = 0.01602 Cor = −0.08308206	• P= 0.001995 Rho = 0.1065027	P = 0.8419 Rho = 0.1065027
performance	–	–	–	• P = 1.265e−05 Rho = 0.1499229	• P = 2.924e−08 Rho = 0.1897809
subjTimeRating	–	–	–	–	• P = 2.924e−08 Rho = 0.1897809

found that the context concerning the trial index (repetition of the action) and fatigue affected these aspects of the user experience. Depending on the familiarity with the task, the presence of a particular type of rhythmic stimulus under possible tempos will affect either performance or time experience. Both aspects can contribute significantly to a flow experience or even well-being in general, and the results of this study can thus inform the design of future interactive VR applications. We also found promising physiological variables to assess one's time experience, which would benefit from future research to untangle them from fatigue's effect or use fatigue as a proxy to assess time perception.

While the familiarity or repetition of a task or action can be easily assessed in any interactive application, using fatigue as a modulator could be a growing opportunity for VR developers as newer HMDs incorporate advanced sensors, e.g., for eye-tracking. We observed effects of rhythmic stimuli under some fatigue and task familiarity, yet the crucial finding is the presence of effect variation rather than the specific effect itself, highlighting the need for studies of time perception concerning context- and subject-dependent time modulations.

CRediT authorship contribution statement

Stéven Picard: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jean Botev:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964464 (ChronoPilot).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cag.2024.104097.

References

[1] Picard S, Botev J. Rhythmic stimuli effects on subjective time perception in immersive virtual environments. In: Proceedings of the 14th international workshop on immersive mixed and virtual environment systems. MMVE '22, New York, NY, USA: Association for Computing Machinery; 2022, p. 5–11. http://dx.doi.org/10.1145/3534086.3534330.
[2] Picard S, Botev J. Rhythmic stimuli and time experience in virtual reality. In: Proceedings of the 20th euroXR international conference. EuroXR '23, Springer, Cham; 2023, p. 53–75. http://dx.doi.org/10.1007/978-3-031-48495-7_4.
[3] Hoagland H. The physiological control of judgments of duration: Evidence for a chemical clock. J Gen Psychol 1933;9(2):267–87. http://dx.doi.org/10.1080/00221309.1933.9920937.
[4] Ghaderi A, Niemeier M, Crawford JD. Linear vector models of time perception account for saccade and stimulus novelty interactions. bioRxiv 2021. http://dx.doi.org/10.1101/2020.07.13.201087.
[5] Gorea A. Ticks per thought or thoughts per tick? A selective review of time perception with hints on future research. J Physiol Paris 2011;105(4):153–63. http://dx.doi.org/10.1016/j.jphysparis.2011.09.008.
[6] Droit-Volet S, Fayolle S, Lamotte M, Gil S. Time, emotion and the embodiment of timing. Timing Time Percept 2013;1(1):99–126. http://dx.doi.org/10.1163/22134468-00002004.
[7] Droit-Volet S, El-Azhari A, Haddar S, Drago R, Gil S. Similar time distortions under the effect of emotion for durations of several minutes and a few seconds. Acta Psychol 2020;210:103170. http://dx.doi.org/10.1016/j.actpsy.2020.103170.
[8] Makwana M, Srinivasan N. Intended outcome expands in time. Sci Rep 2017;7(1):6305. http://dx.doi.org/10.1038/s41598-017-05803-1.
[9] Roseboom W, Fountas Z, Nikiforou K, Bhowmik D, Shanahan M, Seth AK. Activity in perceptual classification networks as a basis for human subjective time perception. Nature Commun 2019;10(1). http://dx.doi.org/10.1038/s41467-018-08194-7.
[10] Watt JD. Effect of boredom proneness on time perception. Psychol Rep 1991;69(1):323–7. http://dx.doi.org/10.2466/pr0.1991.69.1.323, PMID: 1961817.
[11] Rutrecht H, Wittmann M, Khoshnoud S, Igarzábal FA. Time speeds up during flow states: A study in virtual reality with the video game thumper. Timing Time Percept 2021;9(4):353–76. http://dx.doi.org/10.1163/22134468-bja10033.
[12] Jackson SA, Eklund RC. Assessing flow in physical activity: The flow state scale–2 and dispositional flow scale–2. J Sport Exerc Psychol 2002;24(2):133–50. http://dx.doi.org/10.1123/jsep.24.2.133.
[13] Pelet J-É, Ettis S, Cowart K. Optimal experience of flow enhanced by telepresence: Evidence from social media use. Inf Manag 2017;54(1):115–28. http://dx.doi.org/10.1016/j.im.2016.05.001.
[14] Nah F, Eschenbrenner B. Flow experience in virtual worlds: Individuals versus Dyads. In: SIGHCI 2015 proceedings. 2015, URL: https://aisel.aisnet.org/sighci2015/19.
[15] Reid D. A model of playfulness and flow in virtual reality interactions. Presence Teleoperators Virtual Environ 2004;13(4):451–62. http://dx.doi.org/10.1162/1054746041944777.
[16] Mullen G, Davidenko N. Time compression in virtual reality. Timing Time Percept 2021;9:1–16. http://dx.doi.org/10.1163/22134468-bja10034.
[17] Igarzábal FA, Hruby H, Witowska J, Khoshnoud S, Wittmann M. What happens while waiting in virtual reality? A comparison between a virtual and a real waiting situation concerning boredom, self-regulation, and the experience of time. Technol Mind Behav 2021-07-22;2(2). http://dx.doi.org/10.1037/tmb0000038, URL: https://tmb.apaopen.org/pub/what-happens-while-waiting-in-virtual-reality.
[18] Mallam SC, Ernstsen J, Nazir S. Accuracy of time duration estimations in virtual reality. Proc Hum Factors Ergon Soc Annu Meet 2020;64(1):2079–83. http://dx.doi.org/10.1177/1071181320641503.
[19] Bruder G, Steinicke F. Time perception during walking in virtual environments. In: 2014 IEEE virtual reality. VR, 2014, p. 67–8. http://dx.doi.org/10.1109/VR.2014.6802054.

[20] Schatzschneider C, Bruder G, Steinicke F. Who turned the clock? Effects of manipulated zeitgebers, cognitive load and immersion on time estimation. IEEE Trans Vis Comput Graphics 2016;22(4):1387–95. http://dx.doi.org/10.1109/tvcg.2016.2518137.

[21] van der Ham IJ, Klaassen F, van Schie K, Cuperus A. Elapsed time estimates in virtual reality and the physical world: The role of arousal and emotional valence. Comput Hum Behav 2019;94:77–81. http://dx.doi.org/10.1016/j.chb.2019.01.005.

[22] Unruh F, Landeck M, Oberdörfer S, Lugrin JL, Latoschik ME. The influence of avatar embodiment on time perception - towards VR for time-based therapy. Front Virtual Real 2021;2. http://dx.doi.org/10.3389/frvir.2021.658509.

[23] Droit-Volet S, Ramos D, Bueno JLO, Bigand E. Music, emotion, and time perception: the influence of subjective emotional valence and arousal? Front Psychol 2013;4:417. http://dx.doi.org/10.3389/fpsyg.2013.00417.

[24] Hammerschmidt D, Wöllner C, London J, Burger B. Disco time: The relationship between perceived duration and tempo in music. Music Sci 2021;4:2059204320986384. http://dx.doi.org/10.1177/2059204320986384.

[25] Wöllner C, Hammerschmidt D. Tapping to hip-hop: Effects of cognitive load, arousal, and musical meter on time experiences. Atten Percept Psychophys 2021;83(4):1552–61. http://dx.doi.org/10.3758/s13414-020-02227-4.

[26] Peters CM, Glazebrook CM. Rhythmic auditory stimuli heard before and during a reaching movement elicit performance improvements in both temporal and spatial movement parameters. Acta Psychol 2020;207:103086. http://dx.doi.org/10.1016/j.actpsy.2020.103086.

[27] Lukas S, Philipp AM, Koch I. Crossmodal attention switching: Auditory dominance in temporal discrimination tasks. Acta Psychol 2014;153:139–46. http://dx.doi.org/10.1016/j.actpsy.2014.10.003.

[28] Wang X, Wöllner C, Shi Z. Perceiving tempo in incongruent audiovisual presentations of human motion: Evidence for a visual driving effect. Timing Time Percept 2021;10(1):75–95. http://dx.doi.org/10.1163/22134468-bja10036.

[29] Wearden J, Penton-Voak I. Feeling the heat: Body temperature and the rate of subjective time, revisited. Q J Exp Psychol Sect B 1995;48(2):129–41. http://dx.doi.org/10.1080/14640749508401443.

[30] Kingma BR, Roijendijk LM, Maanen LV, Rijn HV, Beurden MHV. Time perception and timed decision task performance during passive heat stress. Temp Austin 2020;8(1):53–63. http://dx.doi.org/10.1080/23328940.2020.1776925.

[31] Dormal V, Heeren A, Pesenti M, Maurage P. Time perception is not for the faint-hearted? Physiological arousal does not influence duration categorisation. Cogn Process 2017;19(3):399–409. http://dx.doi.org/10.1007/s10339-017-0852-3.

[32] Cho D, Ham J, Oh J, Park J, Kim S, Lee NK, Lee B. Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine. Sens Basel 2017;17(10):2435. http://dx.doi.org/10.3390/s17102435.

[33] Chen H, Dey A, Billinghurst M, Lindeman RW. Exploring pupil dilation in emotional virtual reality environments. In: ICAT-EGVE 2017 - International conference on artificial reality and telexistence and eurographics symposium on virtual environments. The Eurographics Association; 2017, http://dx.doi.org/10.2312/EGVE.20171355.

[34] Lambourne K. The effects of acute exercise on temporal generalization. Q J Exp Psychol Hove 2012;65(3):526–40. http://dx.doi.org/10.1080/17470218.2011.605959.

[35] Yarrow K, Haggard P, Heal R, Brown P, Rothwell JC. Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. Nature 2001;414(6861):302–5. http://dx.doi.org/10.1038/35104551.

[36] Knöll J, Morrone MC, Bremmer F. Spatio-temporal topography of saccadic overestimation of time. Vis Res 2013;83:56–65. http://dx.doi.org/10.1016/j.visres.2013.02.013.

[37] Picard S, Botev J. Rhythmic Stimuli and time experience in virtual reality - complete data analysis. Zenodo; 2023, http://dx.doi.org/10.5281/zenodo.10804031.