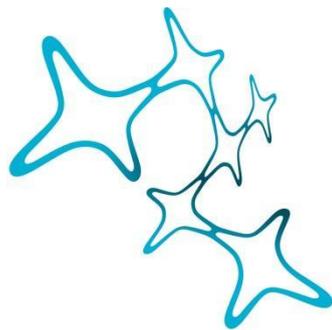
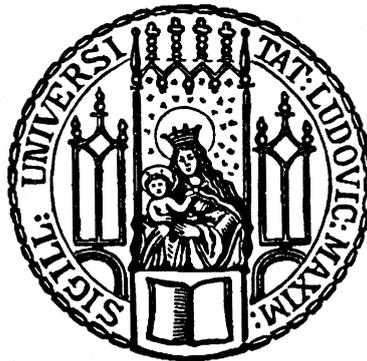

CONTEXT EFFECTS IN INTERVAL TIMING



Graduate School of
Systemic Neurosciences

LMU Munich



Yue Ren

Dissertation der
Graduate School of Systemic Neurosciences der
Ludwig-Maximilians-Universität München

20.02.2020

First Supervisor/ Reviewer:
Prof. Dr. Hermann J. Müller
Department Psychology,
General & Experimental Psychology
Ludwig Maximilian University Munich

Second Supervisor/ Reviewer:
Prof. Dr. Zhuanghua Shi
Department Psychology,
General & Experimental Psychology
Ludwig Maximilian University Munich

Third Reviewer:
Prof. Dr. Knut Drewing
Department Psychology and Sports Science,
General Psychology
Justut-Liebig University Giessen

Date of Submission: 20.02.2020
Date of Defense: 22.07.2020

Abstract

Timing in the range of milliseconds to seconds is fundamental to our everyday behavior and survival, not to mention its vital role to certain expertise such as music and dance. Timing research in this particular time range has been considerably fruitful in both theoretical and methodological developments. Although we doubtless have a time sense, we have no specific sensory organ for the passage of time. Time information has to be processed by other sensory organs (e.g., our eyes and ears). This exhibits an intriguing situation: on the one hand, the process for temporal information is susceptible to various contextual modulation, given that the sense of time is highly dependent on sensory modality and non-temporal intensity of the sensory signal; on the other hand, we are surprisingly good at timing, suggesting a well-functioning integration process to combine various temporal information from different sources together. This thesis is rooted in this situation.

While classic interval timing models, such as the scalar timing theory, or alternative intrinsic models focus on how individual interval is encoded and processed, having hot controversial debates on whether our timing system is embedded with a dedicated or intrinsic mechanism, it is much under debate how temporal contexts (e.g., multiple intervals) and non-temporal contexts (such as stimulus intensity) could influence internal estimations. Without fully examining the role of those context factors, it remains challenging to reconcile or distinguish between different timing models. Thus, this thesis focuses on various contextual effects in milliseconds to seconds interval timing.

Overview

The thesis is structured in five main chapters including three independent studies. The introduction covers the basic concepts of psychophysics and their applications in the time domain. In particular, several important ‘temporal distortions’ that are caused by various Context factors are reviewed.

The first study (Chapter 2) focuses on one basic property of interval timing - the (relative) precision of interval estimation: whether the precision of interval timing follows Weber’s law. According to classical Gibbon’s Scalar Expectancy Theory (SET), the interval timing error should conform to Weber’s law (known as scalar property in time literature), that is, the standard error increases as the time interval increases. However, violation of the scalar property, often using the observed variation of the coefficient of variation (CV), has been found in multiple studies, posing a challenge to this fundamental law. By manipulating temporal contexts (probing intervals in blocked or randomly interleaved manner), we found both conformity and violation of the observed CV. Using a Bayesian model, however, we suggest that the variation of the CV can be explained by the temporal contextual modulation, without resorting to the violation of the fundamental Weber’s law.

The second study (Chapter 3) focuses on another basic issue in interval timing- how is interval encoded internally? Classical interval models (e.g., SET) assume the interval is encoded linearly. However, this poses a serious issue of efficient encoding - an extremely long time interval requires a huge resource to store it. Would it be encoded logarithmically? It is not so easy to disentangle different encoding strategies, given that the observed behavioral responses often include an implicit decoding process. Here we adopt an intuitive process - ensemble coding - low-level fast processing to bypass potential temporal decoding and to probe the internal primitive encoding of time intervals. The results are in line with the logarithmic scaling of time (in the range of sub-seconds to second), indicating a compressed subjective timeline in humans.

The third study (Chapter 4) investigates influences of non-temporal factors (object size of visual stimuli and signal-noise-ratio of auditory stimuli) in interval discriminations,

specifically in the experimental context where visual and auditory stimuli were interleaved probed. Our results show that both the interval estimations and temporal discriminations are highly dependent on non-temporal characteristics and an asymmetric role of non-temporal characteristics between visual and auditory modality in time estimation.

The final chapter discusses the findings of interval timing from these three studies with respect to three types of contexts. And I summarize how the findings could reconcile with dominant timing models, such as scalar timing theory, and the implications of potential theoretical developments in human timing.

Table of Contents

Chapter I. General Introduction	8
1.1 Approaches to the sensory world of Time	10
1.1.1 Weber's law and scalar timing	10
1.1.2 Scalar Timing and Internal-Clock Mechanism	11
1.1.3 (Non) Scalar Timing	12
1.1.4 Linear versus Logarithmic scale of timeline	13
1.2 Context Effect in Interval Timing	14
1.2.1 Temporal context: 'Central Tendency' Effect	15
1.2.2 The case of Ensemble Perception - incorporating knowledge about the overall temporal context	16
1.2.3 Modality Difference and Intensity effect	17
1.3 Perception as Bayesian Inference	18
1.3.1 Bayesian Formulation	19
1.3.2 Bayesian Integration and SET	20
1.3.3 Bayesian Inference and Modality difference	21
1.4 Open Issues and Research Questions	22
Chapter II. The role of Temporal Context in Duration Reproductions	23
2.1 Summary	25
2.2 Reference	25
Chapter III. The Internal Subjective Timeline: Linear or Logarithmic	37
3.1 Summary	39
3.2 Reference	39
Chapter IV. Influences of non-temporal magnitudes in duration perception	53
4.1 Abstract	55
4.2 Introduction	57
4.3 Experiment 1: Auditory and visual reproductions within one session (mixing)	58
4.4 Experiment 2. Auditory and visual reproductions in separate sessions	62
4.5 Experiment 3. Auditory and visual bisection within auditory SNR manipulation	66
4.6 General Discussion	68
4.7 Author Contributions	71
4.8 Acknowledgement	71
4.9 References for Chapter IV	72
Chapter V. General Discussion	77

5.1 Summary of findings	77
5.2 Temporal Context and Bayesian Inference Model	79
5.3 Modality difference between Vision and Audition	80
5.4 Implications of logarithmic encoding of Time	80
5.5 Conclusions	81
References	83
Acknowledgment	97
Curriculum Vitae	99
List of Publications	100
Eidesstattliche Erklärung / Affidavit	101
Declaration of author contributions:	102

Chapter I. General Introduction

Waiting for the sunset at the peak of the Olympia hill, I can see this moment when the whole city gradually immersed into the golden afterglow, I can hear this moment when the church bell strikes the most times of the day, but how can I know this moment, how can I know what this moment represents -- time?

Time is a mystery, or as Professor Callender opened about his book: ‘time is a big invisible thing that will kill you’ (Callender, 2019). Because time is intangible, temporal information heavily relies on concrete existence or external reference to express (Fraisie, 1984). For example, we can use the location of the sun to inform the time of the day, and we use a stopwatch to time the duration of a certain event. Forming the time sense, however, faces even more challenges, as no apparent organ or neural underpinnings could account for how and where time is processed in the human body (Gibson, 1975; Vroomen & Keetels, 2010). Thus, to obtain a proper timing sense and to make accurate time judgments, we highly depend on alternative intermediates carried through variant sensory channels (Simon Grondin, 2010; Matthews & Meck, 2014; Mauk & Buonomano, 2004; van Rijn, 2016). Here, all those dependent factors were included within the general concept of *Context*, of how temporal information is expressed and how the time sense is formed, which is the main object of investigation in this thesis.

To study the time sense, one has to consider how temporal information is mapped to our senses and how the resulting sensation (the subjective sensation) leads to perception and timed behaviors. The subject of psychophysics, which measures sensory magnitudes ‘resting upon evidence and mathematical relations of empirical facts (Fechner, 1860)¹’, provides an ideal tool to bridge the process of temporal information and the timing responses. In the past decades, abundant timing theories have been developed within the scope of psychophysics, yet the role of the ‘Context’ factors has not been fully addressed. In this thesis, I report three studies (Chapters 2 to 4) that focus on three aspects of context factors which can potentially

¹ Translation by Dr. F.H. Petzschnier in her doctoral thesis ‘Magnitue estimation in Humans - a Bayesian approach to characteristic behavior in path integration (Frederike Hermi Petzschnier, 2012)’, original quote ‘...auf Erfahrung und mathematischer Verknüpfung erfahrungsmäßiger Tatsachen [...] zu fußen’ (Fechner, 1860).

affect the observed timing responses and discuss their implications on the existing timing models.

In the following of the Introduction, I will first introduce the fundamental mathematical relationship between absolute (external) magnitudes and our perceived (interval) measurements, which formed the core idea of the psychophysical laws. Second, I will focus on the major statistical properties, such as accuracy and precision, in evaluating timing performances in repetitive examination and experimental environment and their application in subsecond to seconds human timing. Third, I will focus on several classic findings of the temporal distortions in past research, such as the central tendency. Finally, I will introduce the idea of Bayesian inference theory in modeling sensory magnitudes and how the Bayesian model can function as a perfect tool to incorporate information from temporal context into our internal representation of magnitudes.

1.1 Approaches to the sensory world of Time

1.1.1 Weber's law and scalar timing

In psychophysics, the original approach to the sensory world was to construct a quantitative relation between the physical stimuli and perceived measurements via experimental facts, and to reveal this relation to an observable level (Fechner, 1860). Ideally, physical magnitudes (such as weight, distance, luminance, etc) can be mapped onto internal psychological scales with elegant mathematical formulations, which were later known as the main psychophysical laws (Baird & Noma, 1978; Luce & Edwards, 1958). To access to our perceived magnitudes in a quantitative fashion, the earliest method was through discrimination of magnitude estimations. For example, by asking participants to discriminate between stimulus magnitudes with a small increment, the discrimination threshold (or known as the *just noticeable difference*, *JND*) could thus be determined. It has been shown that the JND is approximately proportional to the absolute magnitude, which is later known as **Weber's law**², that is,

$$\frac{\Delta T}{T} = k, \quad (1)$$

where k is a constant known as *Weber's fraction*. Weber's fraction can be easily applied to intuitive perceptions in a wide range of daily situations. For example, when a crowd consists of 5 persons, it is easy for one to detect one more person joining; However, when a crowd of a hundred people, one person leaves or one more person appears does not seem to be noticeable (assuming no additional salient motion or color introduced). In the repetitive measurements, Weber's fraction can be calculated as the ratio between the standard deviation and the mean of magnitude estimations. Dubbed in the time domain, this relation is defined as the *Coefficient of Variance (CV)*.

$$\frac{\Delta T}{T} = \frac{s.d.(T)}{mean(T)} = k, \quad (2)$$

Until today, the constancy of the Weber's fraction (in time literature it is known as the Scalar property) still serves as the hallmark of one of the most influential timing theories — Gibbon's *Scalar Expectancy Theory (SET)*. Gibbon's early description of SET was to test a

² Weber's law was first constructed by Fechner; based on Weber's experimental data. Fechner dedicated this formula after his mentor's name.

'Weber-like assumption' based on meta dataset and secondary data analyses of timing behavior in animal timing (John Gibbon, 1971, 1972; John Gibbon & Church, 1981). The later mature theory has been developed with expansion to both animal timing and human timing on wider applications such as a fixed-interval schedule of reinforcement and discrimination between stimulus duration (Lorraine G. Allan & Gibbon, 1991; John Gibbon et al., 1984). Secondary data analysis had come to the following consequences of scalar timing: 1) Proportional timing. The mean of an interval estimation increased linearly with the duration to be timed; 2) The scalar variance along with the mean, which indicates a constancy of CV irrespective of the duration to be timed. 3) 'Superimposition' where psychometric functions fall on top of each other when plotted on a relative timescale. The third is another type of validation of the scalar property.

1.1.2 Scalar Timing and Internal-Clock Mechanism

The scalar timing theory (or SET) is integrated with the 'internal clock' model that was established by Treisman in the 60s (M. Treisman et al., 1990; Michel Treisman, 1963) and has been expanded by Gibbon, Church, and Warren in the 70s (R. Church, 2003; R. M. Church et al., 1994; Warren H. Meck et al., 1984; W. H. Meck & Church, 1983). The manifest of SET, also known as the Information Process (IP) model, introduces a three-stage model: clock, memory, and decision (See Figure 1, left panel). The clock functions with three parts: a pacemaker, a switch, and an accumulator. The pacemaker generates continuous pulses to the accumulator at a mean rate that is highly related to the timing values, while the switch decides the onset and offset of such a process. When a temporal stimulus occurs (event t), it permits a certain amount of pulses to pass on to the accumulator ($a = \lambda t$, where λ is the pace of pulses being numbered) in the working memory; thus, the perceived time could be recorded and restored as a form of the number of pulses.

Gibbon and colleagues (John Gibbon et al., 1984) introduced a memory transfer variable, which transferred accumulated ticks to the reference memory ($m = k * a$, where $k *$ is a random variable along with the occurrence of memory transformation) for later comparison, which conforms to scalar properties (i.e., Weber's law). For the decision process, a comparator judges information from two sources: the time value recorded in working memory and the stored value of pulses in the accumulator. When these two values are 'close enough' (a threshold b), a response occurs. To be more specific, the decision of whether or not to make a response is made by calculating the relative discrepancy between the

above-mentioned two sources ($|m - a|/m$) with the threshold. Thus, if $|m - a|/ < b$, response; otherwise, continuing the process of updating temporal memory.

It is worth noting that there are several sources of variances that could potentially affect both the accuracy (measured by mean estimation) and the sensitivity (measured by variance) of temporal performances. The internal clock is a model of time perception stating that time across all modalities is measured against a centralized clock that uses a pacemaker with a variable tick rate changing on a trial-to-trial basis (R. M. Church et al., 1994; John Gibbon, 1971). There are three perspectives of variances that should be considered as the sources of variance, which can arise from either stage of the model (clock, memory or decision). The clock stage is a Poisson process whose pulses are accumulated in the working memory until the occurrence of an important event, such as reinforcement or update, in which the number of clock pulses accumulated is transferred from the working (short-term) memory and stored in a reference (or long-term) memory. According to the SET, a response is produced by computing the ratio between the value stored in the reference memory and the current accumulator total. To account for the scalar property of interval timing, i.e., the variability of responses is roughly proportional to the peak time, Gibbon showed that a Poisson distribution for the accumulator requires a time-dependent variance in the decision and memory factors as well as in the internal clock. These additional sources will be seen to dominate overall variance in performances, emphasizing the important role of cognitive systems in time judgments. For such reasons, SET was considered a more general theory of cognitive process rather than merely a behavioral description of time (Gallistel, 1990; Oprisan & Buhusi, 2014).

1.1.3 (Non) Scalar Timing

As previously mentioned, psychophysical laws use continuous functions to describe the inner manifest of time (but not just retrained to the time domain), implying a unified view of perception along with the magnitude range. However, there has been empirical evidence reporting 'breaks' in these functions. For instance, maximum sensitivity, practice, and counting. Accordingly, it is reasonable to assume the existence of distinctive timing mechanisms according to the different time ranges. By reviewing estimated CVs from a wide variety of studies and paradigms, Gibbon also pointed out the potential violations of the scalar timing with possible 'jumps' of CVs between 1s to 1.5s (John Gibbon et al., 1997), implying a multi-timing mechanism in both animal and human timing. In spite of the great

success the scalar timing has enjoyed in both animal timing and human timing (Lorraine G. Allan, 1998; Lejeune & Wearden, 2006; J. Wearden, 2016), there had been an increasing amount of studies that report the violation of scalar timing.

The violations of scalar timing were reported more frequently in sub-second timing since sub-second intervals are mainly processed by automatic timing, which requires less attentional modulation, whereas supra-second durations are under the control of higher cognitive functions such as attention and working memory (Lewis & Miall, 2003). For instance, a sharp jump of CV was found around 50ms (equal to 20Hz of repetition rates) for the perception of inter-click interval (ICI) of click trains (Ungan & Yagcioglu, 2014). The inconstancy of CVs was also found between 1s to 2s by adopting multiple paradigms such as discrimination, production and, categorization. When intervals exceed 1.2s, performances are improved by explicit counting (Allman et al., 2016; Warren H. Meck & Benson, 2002). Other distinctive findings on the violations of the scalar timing included the increase of CVs across time (Poirier et al., 1969) in the millisecond and seconds ranges.

1.1.4 Linear versus Logarithmic scale of timeline

It was assumed by Weber that the inner perception linearly increases with the absolute magnitude, which resulted in the assumption of the linear internal coding. Alternatively, some other models, particularly according to Fecher's claim³, proposed that the internal representation should be logarithmically progressed to the external stimulus intensity. Concerning the coding efficiency, it is reasonable to accept the logarithmic encoding (thus compressed internal scale of the perceived items), which enables the minimum representation to encode a wider range of absolute magnitude (S. Dehaene et al., 2008; Sun et al., 2012; Varshney & Sun, 2013). Evidence that supports the logarithmic encoding has been shown, for example, in numerosity research, several recent studies have shown that the number sense is along the logarithmic scale (e.g., put 10 into the middle of 1 to 100), for the preschool kids and Amazonian indigene (S. Dehaene et al., 2008).

Attempts to directly reveal the subjective timeline had been made using a variety of psychophysical approaches, but distinguishing between the linear and logarithmic timing is largely constrained by the adopted experimental paradigms (J. Gibbon, 1999; Johnson et al., 2002; Matthews & Meck, 2014; Staddon & Higa, 1999). For example, in temporal bisection,

³ By regulating a constant inner sensitivity, Fechner regarded the constant k as an internal unit of sensory representation. Through basic mathematical integration (see full derivation in Baird, 1978), he obtained Fechner's law, in which a logarithmic transformation of stimulus magnitude was indicated.

the bisection point was often observed at the geometric mean (Russell M. Church & Deluty, 1977; Stubbs, 1976), which led to the earliest speculation that subjective timeline might be logarithmic: if time were linearly coded, the search for the midpoint on the subjective scale would be equally distant to both duration references ('Short' and 'Long'), which would lead to their arithmetic mean. However, Gibbon and colleagues provided an alternative explanation that the midpoint was calculated by taking the ratio of two durations, which also results in the geometric mean (L. G. Allan, 1998; L. G. Allan & Gibbon, 1991). Facing this situation, Gibbon and Church developed a 'time-left' procedure (J. Gibbon & Church, 1981), where subjects were given a choice for rewards between a short fixed interval and a long fixed interval with different levels of elapsed time (time left). The argument was: if time were linearly perceived, no preference would be shown between the elapsed time and the remaining time, and the obtained data confirmed this assumption. However, this paradigm was contested by several subsequent studies concerning alternative interpretations. For example, Dehaene (2001) argued that responses collected in the time-left procedure might be guided by the fastest rewards learned in the training phase, where the internal coding strategy showed little influence on the required choice. Moreover, when reducing the bias within the procedure, results tended to show more support of linear timing than logarithmic timing (Trujano & Orduña, 2015). To avoid the potential flaws in the time-left procedure, Yi (2009) adopted an indirect approach to examine the subjective timeline within the framework of signal detection theory (SDT) and found that logarithmic timing with fixed variability could fit the animal timing data better. However, when applying the SDT to humans, the non-linearity of temporal representation became less observable due to individual differences (Jozefowicz et al., 2018).

1.2 Context Effect in Interval Timing

Through psychophysical approaches, most of the human timing data acquired in the experimental environment highly retrained to the adopted experimental settings and paradigms (Simon Grondin, 2008, 2010; see Mioni et al., 2014). Additionally, non-temporal factors that define a presented interval can be surprisingly influential in collecting results. To name a few, whether an interval was 'filled' or 'empty', sensory modality, size, pitch, speed, emotionality, and so forth have been gained increasing attention, and the range of those factors are expanding (Brown, 1995; Gil & Droit-Volet, 2012; Matthews & Meck, 2014 for

more illustrations). Further, even with careful control for the aforementioned factors, the overall temporal context forming testing session (e.g., sample durations within or between blocks) and the most recent presented interval can both affect interval timing considerably. In the following section, I will introduce some of the 'classic' temporal distortions that highly depend on 'Context' factors.

1.2.1 Temporal context: 'Central Tendency' Effect

'Central tendency' effect, also referred to as *Regression Effect* or *Vierordt's law* (Gu & Meck, 2011; Lejeune & Wearden, 2009) states that a variety of subjective judgments, such as time, weight, brightness, area, size of angles, all show the same tendency to gravitate toward a mean magnitude (Hollingworth, 1910). It describes a systematic bias where, within a single test range, small tested magnitudes tend to be overestimated, whereas large tested magnitudes tend to be underestimated. This tendency in the temporal domain was first observed by Vierordt over 150 years ago (Vierordt, 1868), who conducted a duration reproduction task with intervals ranging from 0.5s to 6.5s. The data exhibited a classical regression effect with an 'indifferent point' - where response equal to the actual duration- somewhere within the range. The occurrence of the regression effect has been found in various procedures, such as order comparison and production (Bausenhart et al., 2014; Gu & Meck, 2011; Karaminis et al., 2016; Lejeune & Wearden, 2009). However, the strength of the effect is dependent on context factors, such as modality, ranges of intervals. For example, studies by Woodrow (H. Woodrow, 1930; Herbert Woodrow, 1933) used the production of individual intervals each day (0.2 to 30 s) showing the central tendency effect reduced and disappeared as the days of testing accumulated. In groups of people with special long-term training, musicians for instance, as the years of musical training increase, expert musicians showed overall smaller 'central tendency' than normal group in temporal reproductions (Aagten-Murphy et al., 2014); Also by adopting reproduction task, Cicchini and colleagues revealed that percussionists could perform veridically in both visual and auditory modality, while string musicians and non-musicians showed typical regression effect with different degree of tendencies in visual modality (Guido Marco Cicchini et al., 2012).

Context Factor 1 The first context factor investigated in this thesis is the temporal context comprising sample durations adopted in the testing environment. In Chapter 2 of the thesis, I hypothesized that temporal context plays a vital role in affecting different aspects of timing performances, such as accuracy and precision. As noted, the coefficient of variation (CV),

can be used as a metric to evaluate the goodness (the precision) of timing performance, a type of Weber's fraction in time perception. It has long been debated if the duration judgment follows the scalar property or not by using the observed CV. However, it has been largely neglected in the literature that the observed CV could also be modulated by the temporal context and decision uncertainty. In Chapter 2 I used the temporal reproduction paradigm to examine the variation of the CVs with two types of temporal contexts: a full-range mixed vs. a blocked context, separated in visual and auditory modalities.

1.2.2 The case of Ensemble Perception - incorporating knowledge about the overall temporal context

Encounter ensemble or even redundant information is inevitable in everyday perception. Walking on the street, we see lines of houses, bunches of trees, barely a single one. Our sensory system must have equipped an efficient mechanism in coping with such situations. In a seminal work of Ariely's, when asked participants to identify whether a probe object belongs to the previously represented group of objects, the response surprisingly and consistently lied at the mean object of the group (Ariely, 2001). Since then, research has shown that statistical properties of many features (mainly in the visual domain) can be automatically extracted, from basic features such as sizes and color (Chong & Treisman, 2005; Webster et al., 2014), to high cognitive levels processing from facial expressions, such as emotions, person identities and lifelikeness (de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007, 2012; Leib et al., 2016). Furthermore, perceptual averaging is not limited to a simultaneously presented group (or set) of items. Several studies have shown that perceptual averaging also takes place for sequentially presented stimuli, such as object weights and auditory frequencies (Curtis & Mullin, 1975; Piazza et al., 2013). Considering that the calculation capacity of such an averaging process is limited, it would be more effective to represent the main statistical information, such as mean and variance, of group items, instead of the full information of individual items. However, ensemble perception in the temporal domain is little investigated.

Context Factor 2 The second context factor investigated in this thesis focused on the general or 'ensemble' temporal environment in interval timing. In Chapter 3 of this thesis, I intend to apply this intuitive process - ensemble perception- to the time domain through extracting statistical properties, such as the mean in interval timing and intend to reveal the subjective timeline based on the collected statistical properties. As noted, one of the problems

of psychophysical approaches is that the inner perception is not always manifested by measured magnitudes (Eisler, 1976; Krueger, 1989), which inspired the idea to seek for indirect approach or intuitive process to reaching for the internal representation of sensory magnitudes. In Chapter 3, I will introduce one study that focuses on the question: 'whether the subjective timeline is linear or logarithmic?'

1.2.3 Modality Difference and Intensity effect

Auditory/Visual Differences in Interval Timing Temporal distortions can also occur when the stimulus of durations comprises different modalities, such as vision and audition (Lustig & Meck, 2011; T. Penney, 2003). In particular, when auditory and visual signals are intermixed within one session, participants tend to overestimate auditory signals and underestimate visual signals of equivalent duration. According to the 'Memory Mixing' account proposed by Penney et al (T. B. Penney et al., 2000), the auditory/visual difference of interval timing can seek for explanations within the framework of internal-clock model, in which an integrated memory (addressed as audio-vision memory) from both visual and auditory modality was formed due to the exposure of stimuli from both modalities. Thus, in the decision making (e.g., to classify a give duration in to 'short' or 'long' category), the memory distribution from each modality resulted in different groups of distribution due to different speed of the pacemaker, after the comparison to the audio-vision memory (Lorraine G. Allan, 1998; Taatgen & van Rijn, 2011). In some cases, however, the auditory/visual difference could be diminished. For example, participants can be 'released' from the interference effects of 'memory mixing' by providing appropriate feedback and/or by blocking trials in such a way as to discourage the formation of such distortions in temporal memory (Klapproth, 2009; Vatakis & Spence, 2006a, 2006b).

Intensity Effect The non-temporal characteristics from stimulus played a vital in influencing the duration perception, such as visual stimuli with larger magnitude such as size and luminance were judged longer, and auditory stimuli with larger intensity were tended to be judged longer (Proulx, 2010; Rammsayer & Verner, 2015; Xuan et al., 2007). In an early attempt, Goldstone and colleagues manipulated the intensity levels from visual and auditory stimuli in a duration comparison task and showed that differences between audition and vision in duration judgments were attenuated but not eliminated (Goldstone et al., 1978). Later, Matthews and colleagues have demonstrated that it was the stimulus contrast, rather than the absolute intensities, which affected the duration judgments (Matthews et al., 2011a).

In their work, they used different levels of intensities from both auditory and visual stimuli as well as the background intensity and found that the stimulus magnitude highly depended on the contrast between the stimulus intensity and background intensity (e.g., weak stimuli were judged to last (Matthews et al., 2011b) longer against a high-intensity background).

Context Factor 3 The third context factor investigated in this thesis is the intensity of non-temporal magnitudes in auditory and visual modalities. The auditory/visual differences in interval timing have been heavily investigated, yet the role of non-temporal magnitudes, such as the use of different visual sizes or the different levels of intensities in auditory stimuli, have been neglected in previous studies. Here, I hypothesized that non-temporal magnitudes from the stimulus features would affect the auditory/visual difference of duration judgments, particularly in the context where auditory and visual stimuli were intermixed in the same testing environment (e.g., same block). Specifically, through altering the relative relation between the intensity of stimuli and their background (high or low signal-noise-ratio of auditory stimuli with background noise), we should be able to manipulate the difference from modalities in duration judgments.

1.3 Perception as Bayesian Inference

Bayesian approaches have enjoyed a great deal of recent success in the application to model perceptual problems, mainly in the visual domain (Maloney & Mamassian, 2009; Rhodes, n.d.; Yanagisawa, 2015). The idea behind the Bayesian inference is to characterize general information as a probability distribution and to include knowledge about experienced events. In the temporal domain, as reviewed above, various types of time distortions have been found in the literature, which can be traced back 150 years to Vierordt's law (Lejeune & Wearden, 2009; Vierordt, 1868). However, quantitative predictions and mathematical formulation about time distortions have only been developed in the past decades, when Bayesian Inference was introduced for the explanation of contextual factors (Jazayeri & Shadlen, 2010; Frederike H. Petzschner & Glasauer, 2011; Frederike H. Petzschner et al., 2015; Shi & Burr, 2016; Shi et al., 2013). In this section, I will introduce the computational analysis of Bayesian inference and how the effect of temporal context can be captured based on Bayes' rule in interval timing.

1.3.1 Bayesian Formulation

The Bayesian inference optimizes the perceptual magnitude by taking advantage of two sources of information: a 'likelihood' function $P(S|D)$ - the likelihood of an observed measurement M given a particular stimulus D , and a prior probability $P(D)$ - the knowledge about the stimulus distribution before the observation. According to the Bayes' rule, the posterior distribution $P(D|S)$ can be derived:

$$P(D|S) \sim P(S|D) P(D) \quad (3)$$

In the temporal domain, the Bayesian approach thus provided explanations by using a quantitative prediction of the contribution of temporal context and the mechanisms that involve the subjective representation of duration. Because temporal measurements are static and noisy, integrating the prior knowledge of the statistical distribution of a series of stimulus durations can be beneficial for increasing the precision of duration estimates, although incorporating the prior may lead to sacrificing the accuracy. In this sense, contextual effects are statistically optimal and serve to minimize errors.

The case of Duration Reproduction In a seminal work in 2010, Jazayeri and Shadlen demonstrated a three-stage model using the Bayesian inference that takes into account the underlying distribution of samples to predict subjects' temporal reproductions (Jazayeri & Shadlen, 2010). In the first stage, the relationship between the sample interval and the measurement interval was characterized by measuring the noise, which was modeled as a Gaussian function centered at the mean of sample interval with standard deviation grows linearly along the mean (that is, the scalar timing); The second stage was the observing estimator where the measurement intervals were mapped to the estimation intervals, which would be the inputs for the 'deterministic mapping' model from the third stage for final production. The deterministic mapping functions are associated best with the Bayes least-squares (BLS). Explaining why behavior often conforms to Vierordt's Law is a major theoretical problem for contemporary models of time perception. An interesting fact was revealed from Jazayeri's model where the characteristics in temporal reproductions could be the result of an incorporation of prior experience, which might be correlated to contextual factors such as modality and temporal range, however, how prior knowledge could differ from such contextual factors and whether this model could be generalized to such factors remain unresolved. Additionally, a complete model of interval reproduction should measure motor noise directly. Instead, the proposed model made the simplifying assumption that it

was constant for all participants. Another improvement may be to model the data with likelihood functions and priors that are Gaussian with log-time, thereby more closely reflecting scalar timing. In chapter 2 of this thesis, I focus on investigations into two major factors: modality, and temporal range, of whether contextual effects could alter both the precision and accuracy of temporal reproductions.

1.3.2 Bayesian Integration and SET

Recently, the IP model has been further developed by Shi, Church and Meck (2013) by integrating the Bayesian inference. As has been illustrated in Figure 1 (right panel), the three stages in the IP model could be well-matched with the Bayesian inference process. The sensory likelihood is derived from the clock stage. The prior represents the durations stored in the reference memory, which is updated by current estimates (dashed black arrow). The posterior reflects the probability distribution of the current estimate, combining the clock reading and the influence of the reference memory (indicated by the dashed red arrow). In the decision stage, responses are made based on specific comparison rules. The goal of Bayesian inference is to minimize the loss function, whereas the comparator of the IP model uses a relative discrimination threshold.

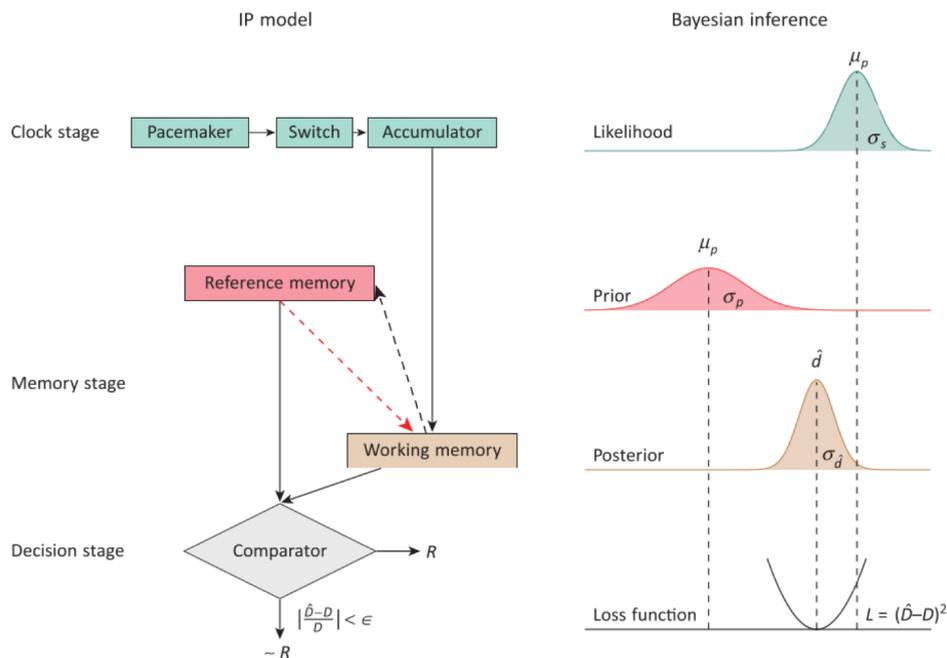


Figure 1. Information process model of interval timing and Bayesian inference of interval timing. Left panel: The three rows are in accordance with the three stages of the model: Clock, Memory and Decision. With the onset and offset of a switch, an accumulator receives and restores a certain number of pulses from a pacemaker(clock process). After reinforcement, the contents in the working memory

are transferred to the reference memory (memory process) for later comparison. A response occurs when a comparator yields a judgement that the discrepancy from two sources of temporal memory is smaller than the threshold. The right panel illustrates that three key components of Bayesian inference could be matched to the three stages of the IP model. The sensory likelihood is derived from the clock stage. The prior represents the durations stored in the reference memory, which is updated by current estimates (dashed black arrow). The posterior reflects the probability distribution of the current estimate, combining the clock reading and the influence of the reference memory (indicated by the dashed red arrow). In the decision stage, responses are made based on specific comparison rules. The optimization from a Bayesian inference is to minimize the loss function, whereas the comparator of the IP model uses a relative discrepancy in the memory load. (figure adapted from (Shi et al., 2013))

1.3.3 Bayesian Inference and Modality difference

As has been introduced in previous sections, the auditory/visual differences of interval timing can occur typically when participants experience both auditory and visual stimuli in the same testing session (T. B. Penney et al., 2000). Although the Information Process (IP) model of scalar timing provided a fairly good explanation to the modality effect, what factors quantitatively determine the level of contextual calibration still remain elusive. In line with the IP theory, the Bayesian inference could account of the ‘memory-mixing’ effect with assuming the ‘interval clock’ are integrated at a faster rate for auditory stimuli than for visual stimuli due to differential rates of opening and closing of the switch that allows pulses to flow from the pacemaker to the accumulator. As a consequence, the internal reference of the mean duration between the ‘short’ (S) and ‘long’ (L) anchor durations is larger for auditory stimuli than for visual stimuli. When the auditory and visual durations are mixed within the same memory distribution, assuming that auditory distributions and visual distributions are independent Gaussians, the internal reference of the mixed durations is a linear-weighted average of audition and Vision. Based on this mixed reference, the auditory and visual points of subjective equality (PSE) are shifted in opposite directions — as indicated by the filled squares and circles, respectively Moreover, the trade-off between precision and bias will depend on the magnitude of uncertainty and the selected cost function.

1.4 Open Issues and Research Questions

In line with traditional psychophysical methods, most of the approaches in modern timing research still follow a similar investigating process, that is, to investigate subjective perception through the relation between bias and precision of interval time. Temporal perception is highly susceptible to changes in experimental context and task and modern theoretical developments are providing increasingly sufficient pieces of evidence in cracking up the timing system. On these grounds, I intended to investigate three concrete factors that may affect interval timing and their potential indications on existing timing models.

The first study focused on one basic property in evaluating time performance — the precision of temporal estimation. According to the scalar timing, another form of Weber's law, the relative precision (measured by the Coefficient of Variance, CV) of temporal estimation should remain constant irrespective of the time duration. However, quite a few recent studies have reported the inconsistencies of this property, thus contesting the ubiquitous law. Although the validity of scalar property raised many debates during the past two decades, investigation concerning the role of temporal context in scalar property has not been carefully examined (Bizo et al., 2006; Simon Grondin, 2014). Here I proposed the violation of scalar property observed in experimental studies could be a side effect of random changes of tested durations, and further quantified such effect with the Bayesian inference model.

The second study focused on one basic question in psychophysics — the subjective timeline of human perception. I adopted summary statistics (also known as ensemble coding or ensemble perception) to investigate whether human participants can incorporate the temporal context and accurately extract the mean duration from a succession of temporal intervals. The results are in line with logarithmic scaling of time (in the range of sub-seconds to second), thus indicating a compressed subjective timeline in humans; little is known of ensemble perception in the temporal domain. Though investigating perception for ensemble intervals, the scale of the inner representation of timing was also expected to be revealed.

The third study was set to investigate two types of memory mixing effect - the modality mixing and non-temporal magnitude mixing. The modality mixing effect has been related to the auditory/visual difference - 'sounds are judged longer than lights' (Goldstone & Lhamon, 1974). Although the auditory/visual difference of interval timing can seek explanations within the framework of scalar timing theory (Goldstone & Lhamon, 1974; Lhamon &

Goldstone, 1974). Mixing effects of magnitudes from vision and audition, such as visual size, signal-noise-ratio, are still less understood. In the final study I will show that both types of mixing effects contribute to the estimation bias, and the magnitude mixing depends on the modality temporal precision.

Chapter II. The role of Temporal Context in Duration Reproductions: a Bayesian Explanation

2.1 Summary

A classic temporal effect was known as the *Vierordt's Law* (or the 'Central-Tendency Effect'), where estimated magnitudes tend to gravitate towards their mean. In a seminal work, Jazayeri and colleagues (2010) proposed a Bayesian estimator model that perfectly explained this effect in duration reproductions. However, whether the precision of our duration estimations - another important aspect of evaluating the goodness of our duration judgments, could also be influenced by the temporal context has not been discussed in previous studies. In this presented article, we tested the influence of the temporal context on duration reproductions in the range of milliseconds to seconds. For the same tested durations, we conducted two sessions with respect to the temporal contexts: the 'Mix' session - all sample intervals were presented in the same test block, and the 'Block (ed)' session - sample intervals were divided into three conditions ('short', 'intermediate', 'long') and tested separately in different blocks. We found that the relative precision - measured using the Coefficient of Variance (CV) showed different tendencies along the tested durations between the 'Mix' and 'Block (ed)' session, demonstrating the influences from the temporal context. Moreover, we proposed a two-stage Bayesian estimator model to incorporate such contextual effects. Our results showed that our timing behavior is highly dependent on the temporal context composed from the testing environments, as well as the stimulus modality (visual versus auditory). And the model simulations also provided explanations as to how the contextual factors and modality influences could draw influences to our internal estimations in different stages.

2.2 Reference

This work was carried out under the supervision of Zhuanghua Shi, and in cooperation with Fredrik Allenmark; Z.S. conceived the idea, Y.R. and Z.S. designed the research, Y.R. collected and analyzed the data, Y.R., Z.S. and H.J.M discussed the results and wrote the paper.

Yue Ren, Fredrik Allenmark, Hermann J Müller, Zhuanghua Shi (2021). Variation in the 'coefficient of variation': rethinking the violation of the scalar property in time-duration judgments. *Acta Psychologica* 214:103263.



Variation in the “coefficient of variation”: Rethinking the violation of the scalar property in time-duration judgments

Yue Ren, Fredrik Allenmark, Hermann J. Müller, Zhuanghua Shi *

General and Experimental Psychology, Psychology Department, LMU Munich, Germany

ARTICLE INFO

Keywords:

Coefficient of variation (CV)
Weber fraction
Duration reproduction
Bayesian model

ABSTRACT

The coefficient of variation (CV), also known as relative standard deviation, has been used to measure the constancy of the Weber fraction, a key signature of efficient neural coding in time perception. It has long been debated whether or not duration judgments follow Weber’s law, with arguments based on examinations of the CV. However, what has been largely ignored in this debate is that the observed CVs may be modulated by temporal context and decision uncertainty, thus questioning conclusions based on this measure. Here, we used a temporal reproduction paradigm to examine the variation of the CV with two types of temporal context: full-range mixed vs. sub-range blocked intervals, separately for intervals presented in the visual and auditory modalities. We found a strong contextual modulation of both interval-duration reproductions and the observed CVs. We then applied a two-stage Bayesian model to predict those variations. Without assuming a violation of the constancy of the Weber fraction, our model successfully predicted the central-tendency effect and the variation in the CV. Our findings and modeling results indicate that both the accuracy and precision of our timing behavior are highly dependent on the temporal context and decision uncertainty. And, critically, they advise caution with using variations of the CV to reject the constancy of the Weber fraction of duration estimation.

1. Introduction

Maintaining both high accuracy and precision of timing performance, particularly in the range of milliseconds to seconds, is fundamental to our basic functioning and survival (Buhusi & Meck, 2005; Meck, 1983). Most timing theories explicitly incorporate Weber scaling in their models (Church et al., 1994): the standard deviation of interval estimation is proportional to the absolute time interval – a characteristic also referred to as the scalar property. For example, the information-processing model of “scalar timing theory” (STT) (Gibbon et al., 1984; Gibbon & Church, 1990) adopts the framework of the classical internal-clock model (Treisman, 1963), which assumes a pacemaker-accumulator that linearly records timed durations. In the STT, the scalar property arises mainly from the variability in the memory transformation of the accumulated ticks (Gibbon, 1991). The scalar property has been confirmed in many animal studies (Gibbon et al., 1997) and in human time perception (Wearden & Lejeune, 2008), and recently has been shown to be an emergent property in artificial perceptron neurons (Buhusi & Oprisan, 2013).

One way of testing the scalar property is to use the estimated

standard deviation (*SD*) and mean (*M*) from each duration timed to construct a coefficient of variance ($CV=SD/M$). The scalar property requires the CV to be constant across the tested range of time samples. Reviewing studies of human timing, Wearden and Lejeune (2008) found that many studies, varying from time reproduction to verbal estimation, confirmed the scalar property. However, violations of the scalar property have been observed when the durations timed were ultrashort or with tasks varying in difficulty or with extensive training (Bizo et al., 2006; Grondin & Killeen, 2009; Matthews & Grondin, 2012). For example, when the range of to-be-estimated durations was rather broad (from 68 ms to 16 min), the observed CVs decreased as duration increased (Lewis & Miall, 2009). Also, it is known that the variability is higher for very short durations (<100 ms) relative to long durations, due to the sensory limits and temporal summation (Scharnowski et al., 2007). Violations of the scalar property manifest mainly in a change of the CV across different time ranges, where the CV is determined by two parameters, *SD* and *M*. Since both parameters are susceptible to bias in certain contexts, errors may creep in the estimation of the CV. For example, the range of to-be-tested durations can heavily influence duration estimation, which is known as the central-tendency effect

* Corresponding author.

E-mail address: shi@lmu.de (Z. Shi).

(Bausenhart et al., 2014; Gu et al., 2016; Gu & Meck, 2011; Jazayeri & Shadlen, 2010; Lejeune & Wearden, 2009; Shi, Church, & Meck, 2013): short durations are often overestimated and long durations underestimated. Thus, the mean estimate for the short/long durations would be larger/smaller than expected, resulting in the variation of the CV across different durations. Similarly, sensory limits in the short-range and decision noise, in general, could affect the observed SDs. As a consequence, the variation of the CVs may well be attributable to context modulation and/or noise perturbation, rather than reflect true violations of the scalar property.

Accordingly, the present study aimed to investigate whether the variation of the CV could be explained solely by contextual modulation and sensory and decision noise, without resorting to “violations” of the scalar property. We adopted the duration-reproduction paradigm, which has been used in previous studies for demonstrating violations of the scalar property (e.g., Lewis & Miall, 2009). We hypothesized that a large range of test durations would impact the variation of the CV more than a small range, given that the central-tendency effect would be stronger for extreme durations. As the central-tendency effect can be predicted quantitatively by the Bayesian inference framework (Jazayeri & Shadlen, 2010; Petschner et al., 2015; Shi & Burr, 2016; Shi, Church, & Meck, 2013), we also modeled the reproduced duration using Bayesian models. Importantly, we incorporated the scalar property in the model, that is: we assume there is *no* violation of the scalar property. If the model with the scalar property can predict the variation of the CV, relying on the variation of the CV alone would be *insufficient* to falsify the scalar property.

In more detail, we asked participants to reproduce the same physical intervals presented under two different conditions. In the whole-range (“mixed”) condition, intervals from 300 ms to 16 s were randomly intermixed, while in the other, sub-range (“blocked”) conditions, intervals were separately tested within sub-second, second, and supra-second (sub-)ranges. In addition, we tested reproduction in both the auditory and visual modalities, given that subjective durations are known to differ between visual and auditory signals (Ganzenmüller et al., 2012; Matthews & Meck, 2014; Shi, Ganzenmüller, & Müller, 2013; Wearden, 2006; Wearden et al., 1998): temporal precision is higher for the auditory system. Finding essentially similar result patterns would strengthen the generalizability of any conclusions. We hypothesized that the CV would show greater variation in the “mixed”, whole-range versus the “blocked”, sub-range conditions, and in vision as compared to audition. And we expected the Bayesian model with the scalar property would predict those variations in CVs.

2. Method

2.1. Participants

A total of 52 volunteers (13 participants for each experimental session), aged 21–33 years (27 females), were recruited from the subject pool of LMU Munich Psychology Department. This number was based on the sample sizes in previous duration-reproduction studies (Cicchini et al., 2012; Jazayeri & Shadlen, 2010; Lewis & Miall, 2009), which ranged from 5 to 14 participants. Our participants had all normal or corrected-to-normal vision, normal hearing, and no somatosensory disorders. Participants provided written informed consent in accordance with the Declaration of Helsinki prior to the experiment and received 9 Euro/h for their participation.

2.2. Apparatus and stimuli

The study was conducted in a dimly lit, sound-attenuated laboratory cabin. Visual stimuli were displayed on a 21-inch CRT monitor with a refresh rate of 100 Hz and a resolution of 800 × 600 pixels. In visual sessions, targets were grey squares (8° × 8° of visual angle), with two levels of luminance: grey (17.5 cd/m²) and white (42.7 cd/m²),

presented on a black screen background (1.60 cd/m²). In Experiments 3 and 4 (auditory sessions), a natural water-flow sound (60 dB, measured at the sound source) was chosen for auditory presentation. Compared to simple sine waves, the water-flow sound potentially reduces fatigue and discomfort during longer stimulus presentations (e.g., 16 s). Auditory stimuli were delivered via two loudspeakers placed on both sides of the computer screen (with a separation of 40 cm). The experimental presentations were generated in Matlab (version 2015a) and with the Psychtoolbox-3 toolbox.

2.3. Experimental procedure

We adopted a between-subject design considering two experimental factors: Temporal Context (“Mix” versus “Block(ed)”) and Modality (“Vision” vs. “Audition”). Accordingly, four separate sessions were conducted with separate groups of participants, hereafter referred to as: “Vis/Mix”, “Vis/Block(ed)”, “Aud/Mix”, “Aud/Block(ed)”. We employed a temporal reproduction task (Lewis & Miall, 2009), in which participants first received a stimulus (a white square or a sound) for a given duration, and then were asked to reproduce that duration by pressing a response key for as long as they had perceived the duration.

2.3.1. The visual sessions (“Vis./Mix” and “Vis./Block(ed)”)

Each trial started with the word prompt “Presentation” shown for 300 ms, indicating the initial presentation of the experiment. This cue was followed by a grey square presented at the center of the display, to which participants were instructed to press the left arrow key with the left index finger when they were ready to start the presentation. Pressing of the key triggered a color change of the square from grey to white. Participants had been told that the duration of the white square was the “target duration” that they had to remember and to reproduce. The duration of the white square was selected from nine intervals (separated equally on the logarithmic scale): 0.30, 0.49, 0.81, 1.33, 2.19, 3.60, 5.92, 9.73, and 16.00 s. At the end of the interval, the color of the square changed (automatically) back to the initial grey, upon which participants had to release the left arrow key. That is, participants were required to hold the key throughout the presentation of the white square and only release it at the end of its duration, signaled by its reversion to a grey square. We kept this production procedure the same as the previous study (Lewis & Miall, 2009) for the purpose of cross-study comparison. The reproduction phase was separated from the target presentation by a 250-ms blank screen, upon which the word “Reproduction” appeared at the screen center for 300 ms. Immediately after this verbal signal, a grey square appeared, prompting participants to start reproducing the given target duration by pressing the right arrow key with their right index finger; participants proceeded to the reproduction at their own pace. Their keypress, again, triggered the change of the square’s color from grey to white. Participants were instructed to keep pressing the key for as long as they had perceived the target duration. The key release triggered a color change back to grey. The next trial started following a random blank interval varying from 500 to 1000 ms.

The same procedure was applied to both visual sessions, the only difference being the exposure to different temporal contexts. In the “Vis./Mix” condition, nine intervals were randomly mixed within each block, whereas in the “Vis./Block(ed)” condition, they were divided into three sub-groups according to their temporal range: the “Short” (0.30, 0.49, 0.81 s), “Intermediate” (1.33, 2.19, 3.60 s), and “Long” groups (5.92, 9.73, 16.00 s). Each experimental session consisted of 15 blocks of 18 trials each.

2.3.2. The auditory sessions (“Aud./Mix” and “Aud./Block(ed)”)

The same paradigm was adopted for auditory sessions, in which auditory excerpts of a natural water-flow sound were used for duration presentation. On each trial, after the visual cues “Presentation” and, respectively, “Reproduction” (indicating the initiation of the respective trial phase) for 300 ms, a fixation cross was presented in the center of the

display until the end of the phase. In the presentation phase, participants were asked to press the left arrow key with the left index finger when they were ready to start the trial. Pressing of the key triggered the water-flow sound for a given duration. Again, participants were required to hold down the key throughout the presentation of the water-flow sound and only release it once the sound stopped. In the reproduction phase, participants were instructed to self-initiate reproduction by pressing the right arrow key with the right index finger, which triggered the water-flow sound. Participants were instructed to keep pressing the key for as long as they had perceived the target duration.

Participants underwent three blocks of training (9 trials per block) prior to performing the experiment proper, which took some 10–15 min to complete. During training, participants' reproduction errors that exceeded 50% of the sample duration were followed by a “warning” message, “Too Short” or “Too Long”, respectively. Participants were told to perform the task avoiding any form of counting. No feedback was given during the formal experiment. To avoid contamination of different temporal contexts, each participant took part in only one experiment in this study, which lasted about 1.5 h. Participants were free to take a break between blocks when needed to prevent loss of concentration and alertness.

2.4. Bayesian modeling

Here, we propose a two-stage Bayesian Estimator to model performance in the temporal reproduction task. Fig. 1 illustrates the computation processes of the model; the computational steps are as follows.

2.4.1. Stage I: duration coding

Following classical psychophysics (Fechner, 1860), we assume the internal duration coding inherits the scalar property from Weber's law: the just noticeable difference is proportional to the absolute magnitude. Accordingly, we introduced a logarithmic transformation of a given sample interval D to the internally measured time:

$$S = \log(D) + \epsilon, \quad (1)$$

where D is the sample duration on the linear scale and S the internal representation of measured duration on the logarithmic scale. The

random variable ϵ represents normally distributed internal-measurement noise.

Because sensory input is noisy, duration estimates can be improved by taking into account the prior probability of encountering a particular duration. In general, when an ideal observer follows Bayesian integration for the perceived duration, where both the prior and the likelihood are independent Gaussians, the optimal internal estimate μ_{X_i} for a given interval is essentially a weighted sum of the interval measure S_i and the mean of the (biased) prior (Cicchini et al., 2012; Jazayeri & Shadlen, 2010; Petzschner et al., 2015; Shi, Church, & Meck, 2013):

$$\mu_{X_i} = (1 - \omega_p)S_i + \omega_p(\mu_p + \Delta_1), \quad (2)$$

and its variance is

$$\sigma_{X_i}^2 = \frac{\sigma_s^2 \sigma_p^2}{\sigma_p^2 + \sigma_s^2}, \quad (3)$$

where the weight $\omega_p = \frac{1/\sigma_p^2}{1/\sigma_p^2 + 1/\sigma_s^2}$ is proportional to the inverse of the sum of the variances; σ_s^2 is the variance of the sensory measurement, which is constant in log-scale representation for a given modality for each participant (i.e., the scalar property on the linear scale); and σ_p^2 is the variance of the prior. Additionally, we consider a shift term (Δ_1) to incorporate a general bias on the mean of the prior distribution.

2.4.2. Stage II: duration reproduction

Duration estimates are transformed back to a linear scale; accordingly, the variance $\sigma_{X_i}^2$ is transformed into the variance of a log-normal distribution:

$$\sigma_{X_i}^2 = \left| e^{\sigma_{X_i}^2} - 1 \right| e^{2\mu_{X_i} + \sigma_{X_i}^2}. \quad (4)$$

In modeling this phase, we also take into account an additional source of variability resulting from the response uncertainty, which has been considered in previous studies (Bizo et al., 2006; Getty, 1975). We assume this duration-independent variability σ_r remains the same across all tested durations for a given participant, which forms the variance:

$$\sigma_i = \sqrt{\sigma_{X_i}^2 + \sigma_r^2}. \quad (5)$$

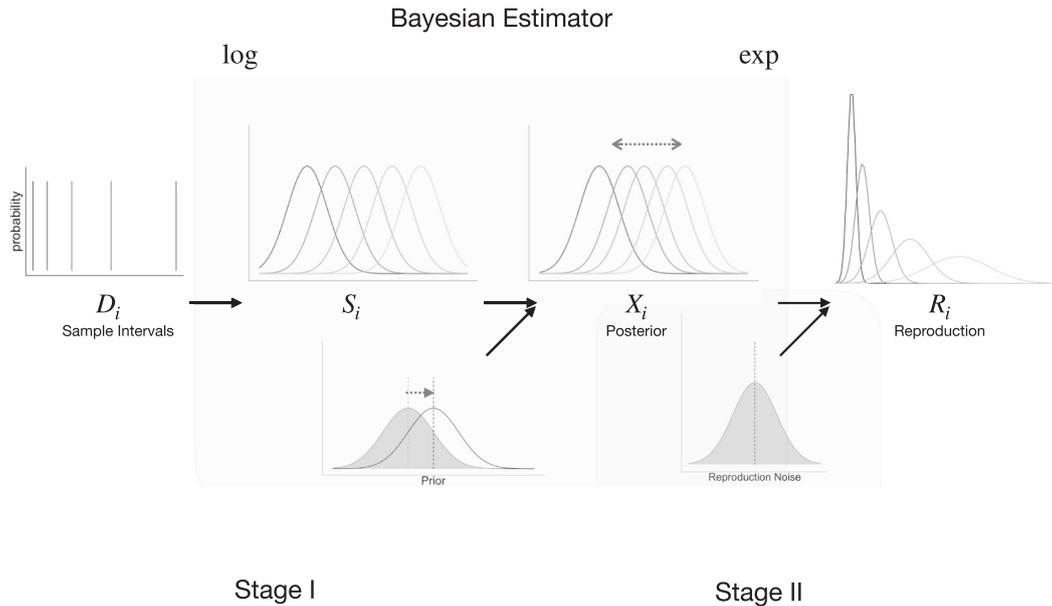


Fig. 1. Two-stage Bayesian-Estimator model for temporal duration reproduction. In the first, “presentation” phase, physical time is encoded through a logarithmic transformation, according to Weber-Fechner's law. The sensory input is then integrated with the prior, which is developed based on the past sample durations encountered. Both the prior and sensory likelihood are assumed to be independent Gaussians. A normally distributed decision noise, which is independent of the sensory magnitudes, contributes to the “reproduction” in the second stage.

We also assume that an additional bias may be introduced at the reproduction stage. Thus, the reproduced duration is:

$$R_i = e^{\frac{\mu_{x_i} + \Delta_2 + \sigma_{x_i}^2}{2}} \quad (6)$$

where Δ_2 is the reproduction stage bias. And the observed CV_i of each interval can be calculated as:

$$CV_i = \frac{\sigma_i}{R_i} \quad (7)$$

In total this model has five parameters: the standard deviations of the sensory noise and the prior (σ_s and σ_p), the two shift terms (Δ_1 and Δ_2) and the stimulus-independent response uncertainty (σ_r). For fitting the data from the mixed conditions the model had only these five parameters. However, for the blocked conditions, we expected there to be contextual influences based on the range of durations within a block (Teghtsoonian & Teghtsoonian, 1978). In order to explore which parameters would be influenced the most by this context, we compared models with each parameter fixed to the same value in each block to models where the parameters could differ between the “short”, “medium” and “long” duration blocks (see Appendix A for details). Thus, in total 32 models were compared. The model fitting was performed in two steps: first, the model predictions for the logarithmic scale mean μ_{x_i} as a function of the log-transformed sample intervals ($\log(D)$) was fitted, assuming a normally distributed likelihood with a constant mean (this was effectively fitting a straight line, in the mixed condition, or three line segments, in the blocked condition). This provided the starting values of the σ_s and σ_p parameters for the full model fit, which was a maximum likelihood fit to the distribution of reproduced durations, assuming that these were normally distributed with the mean given by Eq. (6) and the standard deviation given by Eq. (5) (assuming a normal distribution was an approximation, used in order to simplify the calculations, since the model actually predicted that the distribution should be the convolution of a normal and a log-normal distribution). The optimizations for the maximum likelihood fitting were performed using the “optim” function in R with the “L-BFGS_B” algorithm. If an optimization failed to converge it was attempted again after adding small random numbers to the starting values, and this was repeated until it converged.

3. Results

To exclude duration reproductions likely reflecting lapses of attention or accidental responses, we adopted an outlier criterion based on the interquartile range (IQR): for each participant, at each given

duration, reproduced durations exceeding three times the IQR were omitted from further analysis. In addition, one participant from the “Vis/Block” condition and two from the “Aud/Mix” condition were excluded, due to their outliers exceeding 20% of the total trials.

3.1. Mean reproduction errors and CV

We measured the *Relative Reproduction Error (RRE)* by calculating the difference between the actual response (R_i) and the sample duration (D_i), normalized by the sample duration ($RRE = \frac{R_i - D_i}{D_i}$). This metric provides a measure of the degree of the estimation bias, which is comparable across different durations. Fig. 2 shows both the mean RRE (colored dots) and the model fitting (colored lines) for Experiments 1 to 4, separately for the auditory and visual reproductions. By visual inspection, the RREs exhibit different patterns with respect to the modality of interval presentation. The mean reproductions show larger biases for the visual presentation as compared to the auditory presentation. In addition, the mean RREs show overestimations for both the “short” and “intermediate” blocks (the first six durations ranging from 300 to 811 ms) but underestimation for the “long” blocks, evidencing a clear central-tendency effect. Moreover, the central-tendency effects likely occurred within the test range - a grand central-tendency effect for the “Mixed” conditions, and three separate effects for the “Blocked” conditions. To confirm the differences in RREs among conditions, we further run ANOVAs on the mean absolute RREs, which only revealed a main effect of Modality, $F(1, 45) = 9.10, p < .01, \eta_p^2 = 0.17, BF_{incl} = 13.01$. The main effect of Condition (“Blocked” vs. “Mixed”), $F(1, 45) = 1.67, p = .20, \eta_p^2 = 0.04, BF_{incl} = 1.01$, and the interaction between Modality and Condition, $F(1, 45) = 1.40, p = .24, \eta_p^2 = 0.03, BF_{incl} = 0.33$, were non-significant, indicative of the main difference in RREs being that between audition and vision.

The mean CVs are plotted in Fig. 3A as a function of the sample interval, separately for experimental conditions. Similar to mean reproductions, the CVs, too, were influenced by Temporal Context and Modality. Under the “Vis/Mix” condition, the CVs appeared to largely decrease as the sample duration increased, which is consistent with previous reports (Lewis & Miall, 2009). In addition, higher precision of temporal estimation in the auditory modality was evidenced by smaller CV values under both (“Mix(ed)” and “Block(ed)”) Context conditions. To quantify the variations of CVs across time, for each participant, we fitted the decreasing trend (hereafter referred to as “CV slopes”) using simple linear regression: $CV = a + b \cdot \log(\text{Duration})$. The reason for using the linear regression instead of more complex forms of curve fitting was two-fold: (i) we aimed to test whether the CV remained constant across the four experimental conditions; (ii) and the results

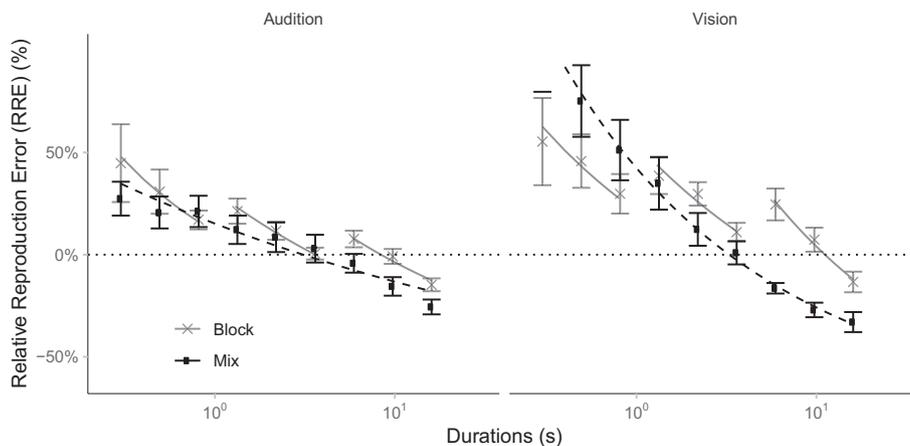


Fig. 2. Relative Reproduction Errors (RREs) (dots and crosses) and RREs predicted by the two-stage model (lines), as a function of the sample interval. The left panel depicts the auditory sessions, the right panel the visual sessions. The black dots represent reproductions from “Mixed” conditions and grey crosses from “Blocked” conditions. The dashed lines represent the prediction from “Mixed” condition and solid lines from the “Blocked” conditions.

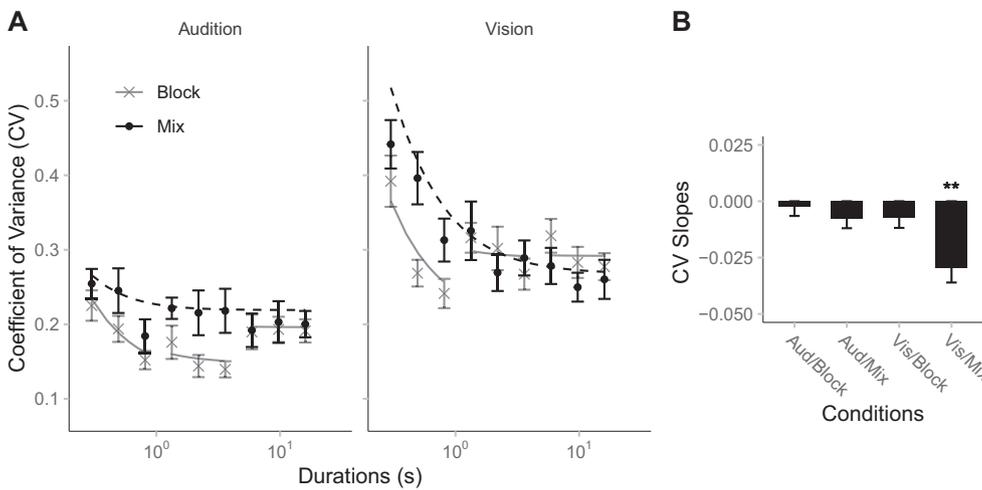


Fig. 3. (A) Mean CVs (dots and crosses) as a function of the sample duration, separately for the auditory (left panel) and visual (right panel) sessions. The black dots represent reproductions from “Mixed” conditions and grey crosses from “Blocked” conditions. The dashed lines represent the model predictions for the “Mixed” condition and solid lines for the “Blocked” conditions. (B) Mean CV slopes as a linear function of logarithmic duration in four experimental conditions. Error bars indicate one standard error. The slope was obtained by estimating parameter b of the linear function $CV = a + b \log(\text{Duration})$.

would be comparable to previous reports (e.g., Lewis & Miall, 2009). Note, though, this analysis does not include the possible “jumps” that the Blocked conditions introduced (see Fig. 3, abrupt changes of the CV across the blocked ranges). We will return to this issue in the next subsection “Model predictions”, which provides more specific psychological grounded model predictions and explanations. The mean CV slopes for four experimental conditions are depicted in Fig. 3B.

The estimated slopes were then submitted to an ANOVA with the (between-subject) factors of Modality (Vision vs. Audition) and Temporal context (“Mixed” vs. “Blocked”), which revealed both main effects to be significant: Modality, $F(1, 45) = 6.84, p = .01, \eta_p^2 = 0.13, BF_{incl} = 3.72$; Context, $F(1, 45) = 7.36, p < .01, \eta_p^2 = 0.14, BF_{incl} = 5.26$; the interaction was non-significant: $F(1, 45) = 2.74, p = .11, \eta_p^2 = 0.06, BF_{incl} = 0.96$. Thus, the variation in the CVs is mainly influenced by the two factors: the decrease was larger with visual than with auditory intervals and larger for whole-range mixed versus sub-range blocked interval durations. Notably, however, the CV slope was largest in the “Vis/Mix” condition. Testing the CV slopes against the null-hypothesis of constant CVs across the range of sample durations (i.e., CV slope = 0) for each experiment revealed there to be a significant (decreasing) trend only in the “Vis/Mix” condition, with a slope of -0.03 from the simple linear regression with logarithmically scaled durations, $t(12) = 4.50, p < .001$. For the other three conditions, by contrast, the (numerically decreasing) slopes did not differ reliably from 0. In the next subsection, we look into these modulations of the modality and temporal context from the perspective of cognitive and computational models.

3.2. Model predictions

The proposed Bayesian-Estimator model for predicting reproduction durations and the CVs has five free parameters: the variance of the prior (σ_p^2), the deviation of the mean of the prior from the physical mean (Δ_1), the variance of the likelihood (σ_s^2), the reproduction stage bias (Δ_2), and the variance of the decision noise (σ_r^2). The Bayesian observer model described above yields a prediction that minimizes the expected loss associated with the coefficient of variation and erroneous estimates in duration reproduction. The model comparison revealed that the best model (in terms of the Bayesian Information Criterion, BIC) allowed the reproduction bias (shift term) Δ_2 and the standard deviation of the prior σ_p , but not any of the other three parameters, to differ among blocks in the “Blocked” conditions. That is, these parameters differ from the “Mixed” conditions in that separate σ_p and Δ_2 parameters were used for “short”, “medium” and “long” duration blocks, resulting in nine parameters in total. The predicted reproduction errors of the best model are shown as lines (dashed lines for Mix(ed) and solid lines for “Block(ed)” conditions) in Fig. 2, and the model-predicted CVs as curves in

Fig. 3A. And the estimates of the model parameters for the four experimental conditions are summarized in Table 1.

In our model framework, we first adopted two parameters (σ_p and Δ_1) to account for the variability and shift in the prior distribution. The best model showed that the shifts of the prior Δ_1 did not differ significantly from 0 in any of the four conditions (one-sample t -test compared to 0: $p = .71, BF_{10} = 0.29$ in the “Vis/Mix” condition; $p = .84, BF_{10} = 0.33$ in the “Aud/Mix” condition; $p = .92, BF_{10} = 0.29$ in the “Aud/Block” condition; $p = .10, BF_{10} = 0.84$ in the “Vis/Block” condition). Moreover, the variability parameter σ_p was not significantly different between the visual and auditory modalities in the “Mixed” condition (two-sample t -test: $p = .81, BF_{10} = 0.38$). A further 2×3 mixed ANOVA with the between-subject factor (“Modality”) and the within-subject factor (i.e., the Temporal Range) for the “Blocked” conditions also failed to reveal any significant difference in σ_p between vision and audition, $F(1, 23) = 0.04, p = .85, \eta_p^2 = 0.002, BF_{incl} = 0.36$, or Modality \times Temporal Range interaction $F(1.53, 35.26) = 0.64, p = .49, \eta_p^2 = 0.03, BF_{incl} = 0.28$. Only the main effect of the Temporal Range turned out significant, $F(1.53, 35.26) = 4.09, p = .04, \eta_p^2 = 0.15, BF_{incl} = 2.80$ (degree of freedoms adjusted by Greenhouse-Geisser sphericity correction, same for the other tests): the variability σ_p decreased as the duration range of the “Blocked” condition increased. The Bayes factors associated with the modality differences can be interpreted in terms of strong evidence of “no difference” in Δ_1 and σ_p between the auditory and visual modalities, arguing in favor of an amodal/supramodal representation of the prior. Interestingly, in the “Mixed” conditions, the weight w_p was higher with visual than with auditory interval presentation (Welch Two Sample t -test: $p = .01, BF_{10} = 4.28$), which is consistent with literature that audition has general higher temporal resolution than vision. However, the weights were not different in the “Blocked” conditions (all $ps \geq .1, BF_{incl} < 0.74$), in part likely due to the block-wise variation of the temporal ranges.

Second, we assumed no violation of the scalar property (i.e., Weber’s law) during the sensory measurement. In line with this, a 2×2 mixed ANOVA on σ_s with the factors Modality (Vision, Auditory) and Condition (“Mixed”, “Block”) revealed no significant effects (Modality: $F(1, 45) = 0.10, p = .76, \eta_p^2 = 0.002, BF_{incl} = 0.29$; Condition: $F(1, 45) = 0.94, p = .04, \eta_p^2 = 0.02, BF_{incl} = 0.42$; interaction, $F(1, 45) = 0.28, p = .60, \eta_p^2 = 0.01, BF_{incl} = 0.40$). One key ingredient for capturing the variation of the CV is the noise parameter σ_r , which plays a bigger role with the short as compared to the long durations (see Eq. (5) in the Bayesian Modeling section). As shown in Table 1, the mean σ_r is larger for the visual modality relative to the auditory modality, and it is the largest in the “Vis/Mix” condition. A further 2×2 ANOVA with Modality (Vision, Auditory) and Condition (“Mixed”, “Block”) as between-subject factors revealed the main effect of Modality to be significant, $F(1, 45) = 11.39,$

Table 1
Model parameters of the Bayesian predictions.

Parameters	Vision				Audition			
	Mix(ed)	Block(ed)			Mix(ed)	Block(ed)		
		Short	Intermediate	Long		Short	Intermediate	Long
ω_p	0.26 ± 0.17	0.22 ± 0.18	0.24 ± 0.15	0.35 ± 0.14	0.12 ± 0.07	0.19 ± 0.23	0.20 ± 0.13	0.21 ± 0.16
σ_p	0.73 ± 0.70	1.42 ± 1.27	0.95 ± 0.98	0.58 ± 0.70	0.68 ± 0.23	1.38 ± 1.19	0.80 ± 0.80	0.94 ± 0.97
Δ_1	0.01 ± 0.15	0.18 ± 3.58			-0.02 ± 0.15	-2.20 ± 4.76		
σ_s	0.34 ± 0.14	0.39 ± 0.41			0.24 ± 0.09	0.41 ± 0.67		
σ_r	0.28 ± 0.23	0.13 ± 0.11			0.05 ± 0.06	0.09 ± 0.09		
Δ_2	0.04 ± 0.16	0.32 ± 0.71	-0.06 ± 0.83	-0.17 ± 0.77	0.00 ± 0.15	0.98 ± 2.27	0.49 ± 0.97	0.17 ± 1.00

$p < .01$, $\eta_p^2 = 0.20$, $BF_{incl} = 16.04$, while there was no overall difference between the “Mixed” and “Blocked” conditions, $F(1,45) = 1.72$, $p = .20$, $\eta_p^2 = 0.04$, $BF_{incl} = 0.58$. The Modality × Condition interaction was significant, $F(1,45) = 5.20$, $p = .03$, $\eta_p^2 = 0.10$, $BF_{incl} = 1.43$. The interaction was attributable mainly to the “Mix(ed)” conditions, with σ_r being significantly larger for the visual (“Vis/Mix”) than for the auditory modality (“Aud/Mix”), $t(22) = 3$, $p < .01$. This pattern suggests that a relatively high uncertainty (i.e., larger σ_r) is associated with the visual modality in general, independently of the presented durations: high uncertainty (e.g., in the “Vis/Mix” condition) introduces a large duration-independent noise term in the reproduction.

More interestingly, in the “Blocked” conditions, the shift in the reproduction Δ_2 (i.e., reproduction bias) appears to decrease as the range of durations increases. A 3×2 mixed ANOVA on Δ_2 with the factors Modality (Vision, Auditory) and Temporal Range (“Small”, “Intermediate”, “Large”) revealed a main effect on the within-subject factor Temporal Range, $F(1.07,24.57) = 6.98$, $p = .01$, $\eta_p^2 = 0.23$, $BF_{incl} = 19.53$, but no interaction, $F(1.07,24.57) = 0.43$, $p = .53$, $\eta_p^2 = 0.02$, $BF_{incl} = 0.24$, or main effect of Modality, $F(1,23) = 1.32$, $p = .26$, $\eta_p^2 = 0.05$,

$BF_{incl} = 0.72$. This captures the “jumps” in the reproduction errors and the CVs in the “Blocked” conditions. The jumps of the CVs, captured by Δ_2 , is likely attributable to the sequential (Cicchini et al., 2018; Fischer & Whitney, 2014), block-by-block variation of the range of durations. A short-range block was most likely preceded by the long- and intermediate-range block, and vice versa. Thus, the tendency of the reproduction was partially carried over across blocks. Interestingly, the best model suggests that the across-block carry-over effect arises in the second stage of the reproduction, rather than the first stage of the duration encoding.

To visualize the goodness of the fit of the best model for individual data, we plotted the predicted RREs versus the observed RREs in Fig. 4. As can be seen, the predicted individual RREs lie mostly on the diagonal line. A correlation analysis revealed high correlations across all four experimental conditions ($r^2 \geq .77$). This was also true for the reproduction variability, measured by the standard deviation of the reproduction ($r^2 \geq .88$, see Appendix B for the plot). In summary, the prediction of the best model is in good agreement with the empirical data we observed.

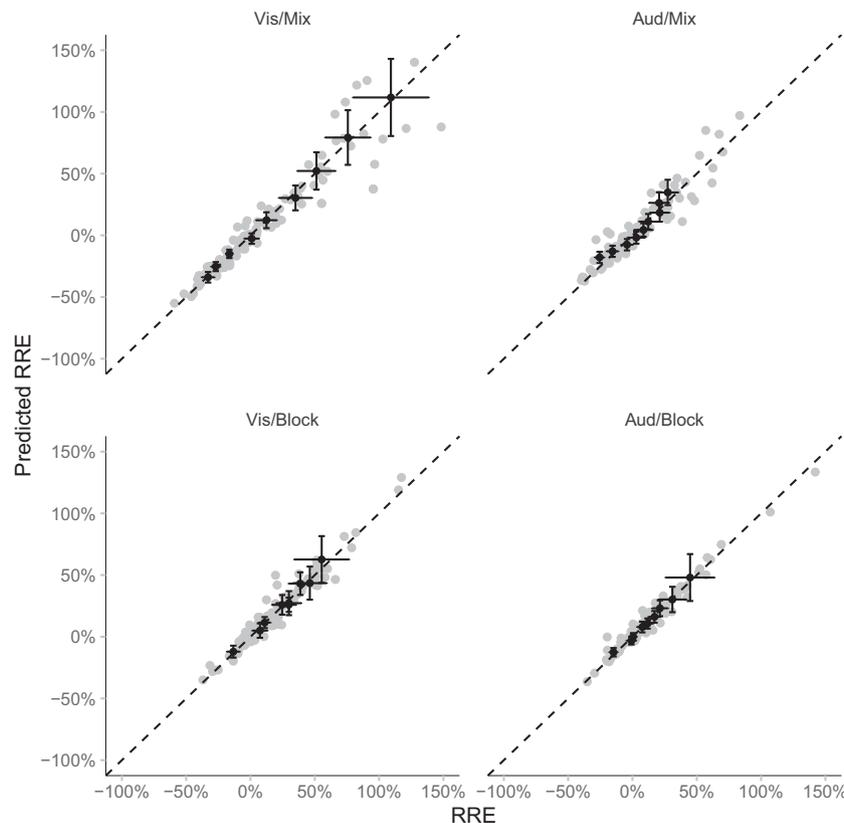


Fig. 4. Predicted Relative Reproduction Errors (RREs) vs. the observed RREs for each observer in all conditions (grey dots), with the mean (\pm SE) model prediction for each combination of the test durations and the experimental conditions against the corresponding, observed mean (\pm SE) RREs. The diagonal line shows the ideal model fit (where the predicted RREs match the observed RREs).

4. Discussion

It has been long debated in the literature whether or not the scalar property is violated in duration perception. The “golden” method for testing for violations of the scalar property has been to use the coefficient of variation (CV), which is defined by the observed standard deviation over the observed mean. Yet, little consideration has been given to the fact that the two parameters may be influenced by the experimental context. The present study was designed to examine for such influences by using a duration-reproduction task, asking participants to reproduce a duration selected from a wide range of intervals (from 300 ms to 16 s) in two types of temporal context: a full-range mixed and sub-range blocked condition, and from both the visual and auditory modalities. We observed the classical central-tendency effect in all four conditions, as well as the variation of the CVs. Importantly, without assuming violation of the scalar property, our two-stage Bayesian model successfully predicted the central-tendency effect and the decrease of the CVs along the timeline. Thus, our findings suggest that the observed CV is not strictly tied to the internal scalar property; and using the CV alone to reject the scalar property is not warranted by the behavioral findings, given that both contextual modulation and decision uncertainty play critical roles in time estimation.

4.1. Variation of CVs in behavioral studies

To fully evaluate the goodness of human timing performance, one has to consider both aspects of temporal judgments: the mean estimate and its precision. While the relation between these two should exhibit the scalar property characteristic, there are findings at variance with this (Meck et al., 1984; Wearden, 1992; Wearden et al., 1997). Some of the reported violations – such as a larger estimation bias with ultrashort intervals or lower precision with more difficult tasks (Lejeune & Wearden, 2006; Wearden & Lejeune, 2008) – can be easily accommodated by adding additional constraints (e.g., sensory limits). However, when it comes to systematic violations of scalar timing, such as the continuous decrease of the CVs across a very wide duration range from 68 ms to 16 min (Lewis & Miall, 2009), the roles played by context factors (in particular, temporal context and presentation modality) must be taken into consideration. Unfortunately, though, context modulation has been largely neglected in the debate of the variation of CVs; and instead, researchers attempted to tweak parameters of their respective timing models. For instance, Gibbon (1991) argued that the variation of the CV arises when the Poisson variance of the pacemaker and the ratio comparison between the current and remembered durations are allowed to change across durations. Memory-trace models, such as the “multiple time scales” (MTS) model (Staddon & Higa, 1999), argue that the scalar property is closely related to Jost’s logarithmic law of memory forgetting (Ebbinghaus, 1885; Jost, 1897; White, 2001). Slightly tweaking the forgetting slope would predict the observed decreasing CV (Staddon & Higa, 1999). However, even with the adjusted parameters, these models cannot explain why different CVs may be observed for the same physical duration depending on different contexts (in the present study: e.g., for the same visual duration in the full-range vs. the short sub-range condition). In the current study, we first replicated previous findings (Lewis & Miall, 2009), demonstrating a similar decreasing trend of the CVs as duration increased in the visual full-range condition; we then expanded the investigation to two different types of temporal context and two modalities. We observed a reduction of the systematic violation when the sample durations were presented in blocked sub-ranges and when the temporal intervals were delivered via the (higher-precision) auditory modality. These results confirm that both the accuracy and precision of timing performance are dependent on the context factors.

4.2. Bayesian integration with contexts

According to Bayesian theory, perception emerges from the

probabilistic inference. The fundamental problem encountered by the brain is to cope with uncertainty from the environment. To minimize uncertainty, the brain needs to make maximal use of the available information, such as knowledge about previously experienced events and the present sensory inputs. The uncertainty can be optimally reduced (to a minimum) when this information is integrated according to its reliability (Ernst & Di Luca, 2011; Taubert et al., 2016). The Bayesian-Estimator model proposed in the current study makes two adjustments in evaluating the sources of uncertainty arising from both stages of the task (duration production and reproduction), according to different temporal contexts: First, based on the fact that subjective duration can differ between different modalities (e.g., Wearden et al., 1998) and temporal context, we assume that the prior itself can be biased. Thus, we implemented a parameter, Δ_1 , to capture this feature. Second, we consider additional biases (Δ_2) that might occur during the reproduction in Stage II, in particular, carry-over of the response tendency from previous into the current trial blocks. Third, we assume time reproduction is corrupted by some duration-independence uncertainty factor, which is captured by the parameter σ_r . This uncertainty is relatively small and can be neglected when durations are at the super-second level. However, with durations in the sub-second range, this factor has to be taken into account in the model.

Among 32 models we compared, the best model assumes the variability of the prior (σ_p) and the reproduction bias (Δ_2) is influenced by the temporal context (i.e., three separate, narrow sub-ranges in the blocked conditions, as compared to one broader range in the full-range mixed condition). Varying the σ_p allows the model to capture the uncertainty of the “Blocked” temporal context, while the reproduction bias (Δ_2) captures the jumps between blocks. For example, the mean reproduction from both the “Short” and “Intermediate” duration blocks exhibited overall overestimations under “Block(ed)” conditions (see Fig. 2). It is possible that these overall overestimations were introduced by the preceding blocks, most likely the “longer” block. This kind of carry-over effect has been found in perceptual judgments, such as serial dependence (Fischer & Whitney, 2014; S. Glasauer & Shi, 2018; Stefan Glasauer, 2019), as well as in temporal reproduction (Wiener et al., 2014). For example, examining the influence of observers’ previous “long” responses on current bisection performance (regardless of later duration presented), Wiener et al. (2014) found a strong tendency for observers to carry over responding “long”, thus dissociating response carry-over from perceptual bias in the bisection task. Also, this response carry-over was more marked for the visual as compared to the auditory modality. In our “Blocked” conditions, we found a similar reproduction carry-over effect captured by the reproduction bias (Δ_2), which was largest for the “short” range, and smallest for the “long” range. Interestingly, the estimated values of the Δ_1 parameter in the “Mixed” conditions (“Vix/Mix” and “Aud/Mix”) were close to unbiased (i.e., 0), with little variability across participants, whereas the estimates varied greatly across individuals in the “Blocked” conditions (the standard error was almost tenfold for the “Blocked” as compared to the “Mixed” conditions). This suggests that participants could establish a relatively unbiased prior given a stable temporal context (the durations were randomly sampled from one distribution), while the priors may change dramatically when the environment changes (witness the large variation with the blockwise presentation). Our model thus incorporates this volatility change in determining the perceptual bias and the response carry-over effect in temporal reproduction.

The duration-independent variability that we introduced in the model – parameter σ_r – was assumed to be independent of duration magnitude. Without assuming any violation of the scalar property, this key parameter captures the variation of the CV, which is consistent with the variation of CVs reported in the literature (Lewis & Miall, 2009). In our model, we considered two sources of variability: the scalar property (the square root of which should linearly increase with the mean) and σ_r . The estimates of σ_r from our model agree with the behavioral observations of the CVs: across the four conditions, the estimated value of σ_r was

largest in the “Vis/Mix” condition (see Table 1), indicative of the largest contribution of decision uncertainty to the variance of reproductions in this condition. Importantly, while we have referred to σ_r as decision uncertainty, our key prediction of larger CVs for the shortest duration is not dependent on whether σ_r solely reflects the variance introduced during the decision of when to stop the reproduction or whether there are also contributions from earlier processes, such as “sensory-onset” variance. Rather, our key prediction only requires that σ_r is independent of duration magnitude.

4.3. Modality difference and the central tendency

Cicchini et al. (2012) demonstrated a robust audiovisual difference in duration reproduction: participants’ responses to visual stimuli tended to gravitate toward the mean of the sample durations (the central-tendency effect) but remained veridical with auditory stimuli (the latter was the case even though participants differed widely in the level of musical expertise). In the current study, we showed that a central-tendency effect in duration reproduction exists in both presentation modalities, though it is more pronounced in vision than in audition, indicated by the relative RREs - thus confirming the audition-vision difference in time judgments (Cicchini et al., 2012).

Interestingly, though, according to our best model, the prior representation of the range of the test durations does not differ much between the audition and vision: there was no significant difference in the two parameters representing the prior knowledge - that is, σ_p (variance of the prior distribution) and Δ_1 (the shift term of the prior) - between the visual and auditory modalities. This suggests that the internal presentation of the temporal context is amodal, consistent with previous findings (Zhang & Zhou, 2017). Our Bayesian model provides a framework of where the modality difference may arise in each phase of the production-reproduction task. In the production phase, when the duration is transferred to short-term memory, the representation of the temporal context (prior knowledge) appears to be little influenced by the presentation modality. Accordingly, the prior knowledge representation appears largely modality-independent. At the same time, the reproduction carry-over bias Δ_2 did show a decreasing trend across the range of the “Blocked” conditions, reflecting block-wise bias.

4.4. Variation of the CV and sub-second timing

In a meta-analysis of animal studies Gibbon et al. (1997) discerned two “jumps” in the CVs at approximately 0.1 and 1.5 s, which they took to suggest that there might be different timing mechanisms for different time ranges. Moreover, the CV has been reported to be particularly large for durations below 100 ms (see a review, Wearden & Lejeune, 2008), for which duration judgments also exhibit a large overestimation (Chen et al., 2016). The large CV for ultrashort visual durations has been argued to be influenced by temporal summation (Gorea, 2015; Scharnowski et al., 2007), that is: below 100 ms, perceived duration depends heavily on light intensity. This would be associated with rather high uncertainty in the estimation of short visual time intervals. In our study, the overall reproduction error for the 300-ms interval reached 102%. However, the “jumps” in our study were observed only in the “Blocked” (and not the “Mixed”) conditions, reflecting a carry-over effect induced by the block-wise variation of the duration ranges.

Of note, we replicated the previous report of the CV changing continuously along the timeline, rather than exhibiting abrupt jumps (Lewis & Miall, 2009). The continuous change of the CV is likely driven by a “hidden” (not directly observable) factor that had not been revealed in early studies. According to our model, which does capture the continuous decrease of the CV, this pattern arises because, while the sensory uncertainty scales with sample duration according to the scalar

property, the decision uncertainty itself is independent of the length of the to-be-judged duration. As a result, the contribution of the decision uncertainty to the total uncertainty increases the CV more for short durations compared to longer durations.

It should be noted that the variation of CV does not solely depend on the temporal context and non-decision noise. It has been reported that extensive learning can enhance temporal discrimination, indicated by a decrease of the Weber fraction (i.e., CV) over the course of the training (Karmarkar & Buonomano, 2003; Matthews & Grondin, 2012). In fact, perceptual learning can boost perceptual discrimination, generally, and not just in the time domain (e.g., Schwartz et al., 2002; Shams & Seitz, 2008). There is evidence indicating that perceptual learning involves a re-tuning of decision templates over the course of training (Li et al., 2004). Interestingly, though, there have been several reports that the enhanced discrimination brought about by learning is rather target-specific. For instance, Karmarkar and Buonomano (2003) showed that training on a 100- or 200-ms interval did enhance the temporal discrimination of the respective (target) interval, without generalizing to untrained intervals. Similarly, in animal temporal-reproduction study, Bizo et al. (2006) found a U-shaped Weber fraction which they attributed to different rates of reinforcement training: a high rate of reinforcement for intermediate durations, relative to extremely short or long durations, led to enhance temporal discrimination (evidenced by a reduced Weber fraction) for the intermediate range of durations. From the perspective of the optimal model framework, perceptual or reinforcement learning of particular durations would reduce the variability of sensory measure σ_s for the learned durations, which may lead to unequal measurement uncertainty σ_s in the logarithmic timeline. This kind of “real” violation in the scalar property can also be reflected in other models we compared (see Appendix A: Model comparison) without assuming the constancy of σ_s . However, in the present study, the best model doesn’t need this additional variation assumption.

In summary, the subjective time is susceptible to various contextual modulations, as is the coefficient of variation (CV). In light of the present results, the “golden” method to test for violation of the scalar property has to be reconsidered: we successfully modeled the observed variation of the CV without assuming any violations of the scalar property. Both the temporal context and decision uncertainty are contributing to the variation of the CV, with the latter (decision uncertainty) playing a critical role in accounting for the large CV for sub-second durations.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.actpsy.2021.103263>.

CRedit authorship contribution statement

Y.R. and Z.S. conceived the experimental design. Y.R. collected the data. Y.R., F.A., and Z.S. carried out statistical analysis and Bayesian modeling. Y.R., F.A., Z.S., H.J.M. drafted the manuscript. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Acknowledgment

This work was supported by the German Research Foundation (DFG) grants SH166/3-2, awarded to ZS.

The study was conducted with the understanding and written, informed consent from each participant, with the approval of the Ethics Board of the LMU Faculty of Pedagogics and Psychology (23.04.2018).

Data accessibility

Data can be accessed on the Github platform: https://github.com/msenselab/variation_in_cv.

Appendix A. Model comparison

In the blocked conditions, we expected that there might be some influence of the different temporal contexts in the “short”, “medium”, and “long” duration blocks. In order to investigate what form of influence temporal context would take, we performed a model comparison in which, for each of the five model parameters, we used either a single value of the parameter for all blocks or different values for the “short”, “medium”, and “long” blocks. The five parameters we compared in this way were: the standard deviation of the sensory measurement σ_s , the standard deviation of the prior σ_p , the two shift parameters Δ_1 and Δ_2 , as well as the standard deviation related to response uncertainty. Combinatorially, these five factors yielded in a total of $2^5 = 32$ different possible models. We fitted each of these models to the data of individual participants and evaluated them in terms of the average Bayesian Information Criterion (BIC) across participants. Table A1 lists the parameters that took on different values for the “short”, “medium”, and “long” duration blocks for the five models with the lowest BIC values (averaged across participants). The best model allowed the reproduction bias Δ_2 and the standard deviation of the prior σ_p , but not the other three parameters, to differ among blocks.

Table A1

The five models with the lowest BIC.

Separate σ_s	Separate σ_p	Separate Δ_1	Separate σ_r	Separate Δ_2	BIC
No	Yes	No	No	Yes	490
Yes	No	No	No	Yes	492
No	Yes	No	No	No	494
Yes	Yes	No	No	No	495
Yes	No	Yes	No	No	497

In addition to finding the best model (with the lowest BIC), we further evaluated for each parameter whether the model performed better with or without the property of allowing that parameter to vary among blocks, by comparing the BIC of the best models with that property to the BIC of the best model without it (ΔBIC_{\min} in Table A2) as well as the difference in average BIC across all models with the property and those without it ($\Delta\text{BIC}_{\text{mean}}$ in Table A2).

Table A2

For each factor, this table presents the difference in average BIC across all models with each property and those without it ($\Delta\text{BIC}_{\text{mean}}$), as well as the difference in BIC of the best model (with the lowest BIC) with a property and the best model without it (ΔBIC_{\min}). Negative values support models with the property over models without it.

	$\Delta\text{BIC}_{\text{mean}}$	ΔBIC_{\min}
Separate σ_s	-4.8	2.0
Separate σ_p	-6.2	-2.0
Separate Δ_1	-2.8	6.3
Separate σ_r	5.5	6.4
Separate Δ_2	-5.4	-3.7

This analysis supported models with a different reproduction bias parameter in different blocks and a separate σ_p parameter in each block. For the σ_r parameter, the analysis supported models with a single parameter across blocks. For σ_s and Δ_1 , the results were ambiguous: models with a single σ_s parameter and a single Δ_1 parameter were supported in terms of ΔBIC_{\min} (since these were properties of the best model); in terms of $\Delta\text{BIC}_{\text{mean}}$, by contrast, models in which these parameters could differ between blocks performed better.

Appendix B

To identify how well the model predicts the variability of the duration reproduction at the individual-participant level, we plotted the predicted standard deviations (SDs) of the duration reproduction from the best model vs. the observed reproduction variability (SDs) in Fig. S1, separately for the four experimental conditions. A correlation analysis revealed the predicted and observed SDs to be highly correlated ($r^2 = .924, .879, .956, .95$ for the “Vis/Mix”, “Aud/Mix”, “Vis/Block”, “Aud/Block” conditions respectively), indicating that the model prediction is in good agreement with the observed data.

References

- Bausenhart, K. M., Dyjas, O., & Ulrich, R. (2014). Temporal reproductions are influenced by an internal reference: Explaining the Vierordt effect. *Acta Psychologica*, *147*, 60–67.
- Bizo, L. A., Chu, J. Y. M., Sanabria, F., & Killeen, P. R. (2006). The failure of Weber’s law in time perception and production. *Behavioural Processes*, *71*(2–3), 201–210.
- Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews. Neuroscience*, *6*(10), 755–765.
- Buhusi, C. V., & Oprisan, S. A. (2013). Time-scale invariance as an emergent property in a perceptron with realistic, noisy neurons. *Behavioural Processes*. <https://doi.org/10.1016/j.beproc.2013.02.015>
- Chen, L., Bao, Y., & Wittmann, M. (2016). *Sub- and supra-second timing: Brain*. Frontiers Media SA: learning and development.
- Church, R. M., Meck, W. H., & Gibbon, J. (1994). Application of scalar timing theory to individual trials. *J. Exp. Psychol. Anim. Behav. Process.*, *20*(2), 135–155.
- Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(3), 1056–1060.
- Cicchini, G. M., Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence. *Proc. Biol. Sci. R. Soc.*, *285*(1890). <https://doi.org/10.1098/rspb.2018.1722>
- Ebbinghaus, H. (1885). *Über Das Gedächtnis: Untersuchungen Zur Experimentellen Psychologie*. LLC: Creative Media Partners.
- Ernst, M. O., & Di Luca, M. (2011). Multisensory perception: From integration to remapping. In *Sensory cue integration* (pp. 224–250). <https://doi.org/10.1093/acprof:oso/9780195387247.003.0012>
- Fechner, G. T. (1860). *Elemente der Psychophysik*. Breitkopf und Härtel: Vol I and II.
- Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743.
- Genzenmüller, S., Shi, Z., & Müller, H. J. (2012). Duration reproduction with sensory feedback delay: Differential involvement of perception and action time. *Frontiers in Integrative Neuroscience*, *6*(October), 1–11.

- Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. In *Vol. 18(1). Perception & psychophysics* (pp. 1–8). <https://doi.org/10.3758/bf03199358>
- Gibbon, J. (1991). Origins of scalar timing. *Learning and Motivation*, 22(1–2), 3–38.
- Gibbon, J., & Church, R. M. (1990). Representation of time. *Cognition*, 37(1–2), 23–54.
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, 423, 52–77.
- Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. (1997). Toward a neurobiology of temporal cognition: Advances and challenges. *Current Opinion in Neurobiology*, 7(2), 170–184.
- Glasauer, S. (2019). Chapter 1 - Sequential Bayesian updating as a model for human perception. In S. Ramat, & A. G. Shaikh (Eds.), *Vol. 249. Progress in brain research* (pp. 3–18). Elsevier.
- Glasauer, S., & Shi, Z. (2018). 150 years of research on Vierordt's law-Fechner's fault? *bioRxiv*. <https://doi.org/10.1101/450726v1.abstract>
- Gorea, A. (2015). A refresher of the original Bloch's law paper (Bloch, July 1885). *I-Perception*, 6(4), Article 2041669515593043.
- Grondin, S., & Killeen, P. R. (2009). Tracking time with song and count: Different Weber functions for musicians and nonmusicians. *Attention, Perception & Psychophysics*, 71(7), 1649–1654.
- Gu, B. M., Jurkowski, A. J., Shi, Z., & Meck, W. H. (2016). Bayesian optimization of interval timing and biases in temporal memory as a function of temporal context, feedback, and dopamine levels in young, aged and *Timing & Time Perception*, 4(4), 315–342. https://brill.com/view/journals/time/4/4/article-p315_1.xml.
- Gu, B. M., & Meck, W. H. (2011). New perspectives on Vierordt's law: Memory-mixing in ordinal temporal comparison tasks. In *Multidisciplinary aspects of time and time perception* (pp. 67–78). https://doi.org/10.1007/978-3-642-21478-3_6
- Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, 13(8), 1020–1026.
- Jost, A. (1897). *Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen* (L. Voss).
- Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning & Memory*, 10(2), 141–147.
- Lejeune, H., & Wearden, J. H. (2006). Scalar properties in animal timing: Conformity and violations. *Quarterly Journal of Experimental Psychology*, 59(11), 1875–1908.
- Lejeune, H., & Wearden, J. H. (2009). Vierordt's. The Experimental Study of the Time Sense (1868) and its legacy. In *European journal of cognitive psychology* (Vol. 21, Issue 6, pp. 941–960). <https://doi.org/10.1080/09541440802453006>.
- Lewis, P. A., & Miall, R. C. (2009). The precision of temporal judgement: Milliseconds, many minutes, and beyond. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.*, 364(1525), 1897–1905.
- Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision "template" for position discrimination. *Nature Neuroscience*, 7(2), 178–183.
- Matthews, W. J., & Grondin, S. (2012). On the replication of Kristofferson's (1980) quantal timing for duration discrimination: Some learning but no quanta and not much of a Weber constant. *Attention, Perception & Psychophysics*, 74(5), 1056–1072.
- Matthews, W. J., & Meck, W. H. (2014). Time perception: The bad news and the good. *Wiley Interdiscip. Rev.*, 5(4), 429–446.
- Meck, W. H. (1983). Selective adjustment of the speed of internal clock and memory processes. *J. Exp. Psychol. Anim. Behav. Process.*, 9(2), 171–201.
- Meck, W. H., Komeily-Zadeh, F. N., & Church, R. M. (1984). Two-step acquisition: Modification of an internal clock's criterion. *J. Exp. Psychol. Anim. Behav. Process.*, 10(3), 297–306.
- Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, 1–9.
- Scharnowski, F., Hermens, F., & Herzog, M. H. (2007). Bloch's law and the dynamics of feature fusion. *Vision Research*, 47(18), 2444–2452.
- Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: A functional MRI study of visual texture discrimination. *Proceedings of the National Academy of Sciences*. <https://www.pnas.org/content/99/26/17137.short>.
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, 12(11), 411–417.
- Shi, Z., & Burr, D. (2016). Predictive coding of multisensory timing. *Current Opinion in Behavioral Sciences*, 8, 200–206.
- Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences*, 17(11), 556–564.
- Shi, Z., Ganzenmüller, S., & Müller, H. J. (2013). Reducing bias in auditory duration reproduction by integrating the reproduced signal. *PLoS One*, 8(4), Article e62065.
- Staddon, J. E., & Higa, J. J. (1999). Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior*, 71(2), 215–251.
- Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, 6, 32239.
- Teghtsoonian, R., & Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics*, 24(4), 305–314.
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychol. Monogr. Gen. Appl.*, 77(13), 1–31.
- Wearden, J. H. (1992). Temporal generalization in humans. In *Journal of experimental psychology: Animal behavior processes* (Vol. 18(2), pp. 134–144). <https://doi.org/10.1037//0097-7403.18.2.134>
- Wearden, J. H. (2006). When do auditory/visual differences in duration judgments occur? *The Quarterly Journal of Experimental Psychology*, 59(10), 1709–1724.
- Wearden, J. H., Denovan, L., Fakhri, M., & Haworth, R. (1997). Scalar timing in temporal generalization in humans with longer stimulus durations. *J. Exp. Psychol. Anim. Behav. Process.*, 23(4), 502–511.
- Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why "sounds are judged longer than lights": Application of a model of the internal clock in humans. *Q. J. Exp. Psychol. Sect. B*, 51(2), 97–120.
- Wearden, J. H., & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations. *Quarterly Journal of Experimental Psychology*, 61(4), 569–587.
- White, K. G. (2001). Forgetting functions. *Animal Learning & Behavior*, 29(3), 193–207.
- Wiener, M., Thompson, J. C., & Coslett, H. B. (2014). Continuous carryover of temporal context dissociates response bias from perceptual influence for duration. *PLoS One*, 9(6), Article e100803.
- Zhang, H., & Zhou, X. (2017). Supramodal representation of temporal priors calibrates interval timing. *Journal of Neurophysiology*, 118, 1244–1256. [jn.01061.2015](https://doi.org/10.1152/jn.01061.2015).

Chapter III. The Internal Subjective Timeline: Linear or Logarithmic?

3.1 Summary

It has long been proposed that we may adopt a more efficient encoding strategy in our perceptual system in coping with the large range of objects' magnitude. Numbers, for instance, have been demonstrated to be logarithmically perceived by humans. In this presented work, we discussed the question of whether our subjective (or interval) timeline is linear or logarithmic. In contrast to most of the previous studies, which mainly operated within the classical psychophysical tasks, we adopted an intuitive approach - ensemble perception to investigate the potential timeline in millisecond-range timing. Ensemble perception, or statistical property, refers to the idea that humans could accurately and rapidly extract the mean from a group of similar items, such as size and numbers in the visual domain. We asked participants to estimate the mean (or average) from a train of consecutively presented duration composed from either visual or auditory stimuli, respectively in two sessions. We found that not only participants could accurately measure the mean of a train of durations, the mean was also consistently lied around the geometric average of the sample values. Hence, we provided behavioral evidence that the interval timeline might be logarithmic, not linear in the timerange of millioseconds.

3.2 Author Contributions

This work was carried out under the supervision of Zhuanghua Shi, and in cooperation with Fredrik Allenmark; Y.R. and Z.S. conceived the idea, designed the research, Y.R. collected and analyzed the data. Y.R and F.A modeled the data. Y.R., F.A., H.J.M and Z.S. discussed the results and wrote the paper.

Yue Ren, Fredrik Allenmark, Hermann J Müller, Zhuanghua Shi (2020). Logarithmic encoding of ensemble time intervals. *Scientific Reports* 10: 18174.



OPEN

Logarithmic encoding of ensemble time intervals

Yue Ren, Fredrik Allenmark, Hermann J. Müller & Zhuanghua Shi✉

Although time perception is based on the internal representation of time, whether the subjective timeline is scaled linearly or logarithmically remains an open issue. Evidence from previous research is mixed: while the classical internal-clock model assumes a linear scale with scalar variability, there is evidence that logarithmic timing provides a better fit to behavioral data. A major challenge for investigating the nature of the internal scale is that the retrieval process required for time judgments may involve a remapping of the subjective time back to the objective scale, complicating any direct interpretation of behavioral findings. Here, we used a novel approach, requiring rapid intuitive ‘ensemble’ averaging of a whole set of time intervals, to probe the subjective timeline. Specifically, observers’ task was to average a series of successively presented, auditory or visual, intervals in the time range 300–1300 ms. Importantly, the intervals were taken from three sets of durations, which were distributed such that the arithmetic mean (from the linear scale) and the geometric mean (from the logarithmic scale) were clearly distinguishable. Consistently across the three sets and the two presentation modalities, our results revealed subjective averaging to be close to the geometric mean, indicative of a logarithmic timeline underlying time perception.

What is the mental scale of time? Although this is one of the most fundamental issues in timing research that has long been posed, it remains only poorly understood. The classical internal-clock model implicitly assumes linear coding of time: a central pacemaker generates ticks and an accumulator collects the ticks in a process of linear summation^{1,2}. However, the neuronal plausibility of such a coding scheme has been called into doubt: large time intervals would require an accumulator with (near-)unlimited capacity³, making it very costly to implement such a mechanism neuronally^{4,5}. Given this, alternative timing models have been proposed that use oscillatory patterns or neuronal trajectories to encode temporal information^{6–9}. For example, the striatal beat-frequency model^{6,9,10} assumes that time intervals are encoded in the oscillatory firing patterns of cortical neurons, with the length of an interval being discernible, for time judgments, by the similarity of an oscillatory pattern with patterns stored in memory. Neuronal trajectory models, on the other hand, use intrinsic neuronal patterns as markers for timing. However, owing to the ‘arbitrary’ nature of neuronal patterns, encoded intervals cannot easily be used for simple arithmetic computations, such as the summation or subtraction of two intervals. Accordingly, these models have been criticized for lacking computational accessibility¹¹. Recently, a neural integration model^{12–14} adopted stochastic drift diffusion as the temporal integrator which, similar to the classic internal-clock model, starts the accumulation at the onset of an interval and increases until the integrator reaches a decision threshold. To avoid the ‘unlimited-capacity’ problem encountered by the internal-clock model, the neural integration model assumes that the ramping activities reach a fixed decision barrier, though with different drift rates—in particular, a lower rate for longer intervals. However, this proposal encounters a conceptual problem: the length of the interval would need to be known at the start of the accumulation. Thus, while a variety of timing models have been proposed, there is no agreement on how time intervals are actually encoded.

There have been many attempts, using a variety of psychophysical approaches, to directly uncover the subjective timeline that underlies time judgments. However, distinguishing between linear and logarithmic timing turned out to be constrained by the experimental paradigms adopted^{15–21}. In temporal bisection tasks, for instance, a given probe interval is compared to two, short and long, standard intervals, and observers have to judge whether the probe interval is closer to one or the other. The bisection point—that is, the point that is subjectively equally distant to the short and long time references—was often found to be close to the geometric mean^{22,23}. Such observations led to the earliest speculation that the subjective timeline might be logarithmic in nature: if time were coded linearly, the midpoint on the subjective scale should be equidistant from both (the short and long) references, yielding their arithmetic mean. By contrast, with logarithmic coding of time, the midpoint between both references (on the logarithmic scale) would be their geometric mean, as is frequently observed. However, Gibbon and colleagues offered an alternative explanation for why the bisection point may

General and Experimental Psychology, Psychology Department, LMU Munich, 80802 Munich, Germany. ✉email: strogway@psy.lmu.de

turn out close to the geometric mean, namely: rather than being diagnostic of the internal coding of time, the midpoint relates to the comparison between the ratios of the elapsed time T with respect to the *Short* and *Long* reference durations, respectively; accordingly, the subjective midpoint is the time T for which the ratios *Short*/ T and T /*Long* are equal, which also yields the geometric mean^{24,25}. Based on a meta-analysis of 148 experiments using the temporal bisection task across 18 independent studies, Kopec and Brody concluded that the bisection point is influenced by a number of factors, including the short-long spread (i.e., the *Long/Short* ratio), probe context, and even observers' age. For instance, for short-long spreads less than 2, the bisection points were close to the geometric mean of the short and long standards, but they shifted toward the arithmetic mean when the spread increased. In addition, the bisection points can be biased by the probe context, such as the spacing of the probe durations presented^{15,17,26}. Thus, approaches relying on simple duration comparison have limited utility to uncover the internal timeline.

The timeline issue became more complicated when it was discovered that time judgments are greatly impacted by temporal context. One prime example is the central-tendency effect^{27,28}: instead of being veridical, observed time judgments are often assimilated towards the center of the sampled durations (i.e., short durations are over- and long durations under-estimated). This makes a direct interpretation of the timeline difficult, if not impossible. On a Bayesian interpretation of the central-tendency effect, the perceived duration is a weighted average of the sensory measure and prior knowledge of the sampled durations, where their respective weights are commensurate to their reliability^{29,30}. There is one point within the range of time estimation where time judgments are accurate: the point close to the mean of the sampled durations (i.e., prior), which is referred to as 'indifference point'²⁷. Varying the ranges of the sampled durations, Jones and McAuley³¹ examined whether the indifference point would be closer to the geometric or the arithmetic mean of the test intervals. The results turned out rather mixed. It should be noted, though, that the mean of the prior is dynamically updated across trials by integrating previous sampled intervals into the prior—which is why it may not provide the best anchor for probing the internal timeline.

Probing the internal timeline becomes even more challenging if we consider that the observer's response to a time interval may not directly reflect the *internal* representation, but rather a decoded outcome. For example, an external interval might be encoded and stored (in memory) in a compressed, logarithmic format internally. When that interval is retrieved, it may first have to be decoded (i.e., transformed from logarithmic to linear space) in working memory before any further comparison can be made. The involvement of decoding processes would complicate drawing direct inferences from empirical data. However, it may be possible to escape such complications by examining basic 'intuitions' of interval timing, which may bypass complex decoding processes. One fundamental perceptual intuition we use all the time is 'ensemble perception'. Ensemble perception refers to the notion that our sensory systems can rapidly extract statistical (summary) properties from a set of similar items, such as their sum or mean magnitude. For example, Dehaene et al.³² used an individual number-space mapping task to compare Mundurucu, an Amazonian indigenous culture with a reduced number lexicon, to US American educated participants. They found that the Mundurucu group, across all ages, mapped symbolic and nonsymbolic numbers onto a logarithmic scale, whereas educated western adults used linear mapping of numbers onto space—favoring the idea that the initial intuition of number is logarithmic³². Moreover, kindergarten and pre-school children also exhibit a non-linear representation of numbers close to logarithmic compression (e.g., they place the number 10 near the midpoint of the 1–100 scale)³³. This nonlinearity then becomes less prominent as the years of schooling increase^{34–36}. That is, the sophisticated mapping knowledge associated with the development of 'mathematical competency' comes to supersede the basic intuitive logarithmic mapping, bringing about a transition from logarithmic to linear numerical estimation³⁷. However, rather than being unlearned, the innate, logarithmic scaling of number may in fact remain available (which can be shown under certain experimental conditions) and compete with the semantic knowledge of numeric value acquired during school education.

Our perceptual intuition works very fast. For example, we quickly form an idea about the average size of apples from just taking a glimpse at the apple tree. In a seminal study by Ariel³⁸, participants, when asked to identify whether a presented object belonged to a group of similar items, tended to automatically respond with the mean size. Intuitive averaging has been demonstrated for various features in the visual domain³⁹, from primary ensembles such as object size^{40,41} and color⁴², to high-level ensembles such as facial expression and likeliness^{43–46}. Rather than being confined to the (inherently 'parallel') visual domain, ensemble perception has also been demonstrated for sequentially presented items, such as auditory frequency, tone loudness, and weight^{47–50}. In a cross-modal temporal integration study, Chen et al.⁵¹ showed that the average interval of a train of auditory intervals can quickly capture a subsequently presented visual interval, influencing visual motion perception.

In brief, our perceptual systems can automatically extract overall statistical properties using very basic intuitions to cope with sensory information overload and the limited capacity of working memory. Thus, given that ensemble perception operates at a fast and low-level stage of processing (possibly bypassing many high-level cognitive decoding processes), using ensemble perception as a tool to test time perception may provide us with new insights into the internal representation of time intervals.

On this background, we designed an interval duration-averaging task in which observers were asked to compare the average duration of a set of intervals to a standard interval. We hypothesized that if the underlying interval representation is linear, the intuitive average should reflect the arithmetic mean (AM) of the sample intervals. Conversely, if intervals are logarithmically encoded internally and intuitive averaging operates on that level (i.e., without remapping individual intervals from logarithmic to linear scale), we would expect the readout of the intuitive average at the intervals' geometric mean (GM). This is based on the fact that the exponential transform of the average of the log-encoded intervals is the geometric mean. Note, though, that the subjective averaged duration may be subject to general bias and sequence (e.g., time-order error^{52,53}) effects, as has often been observed in studies of time estimation⁵⁴. For this reason, we considered it wiser to compare response patterns across multiple sets of intervals to the patterns predicted, respectively, from the AM and the GM, rather than

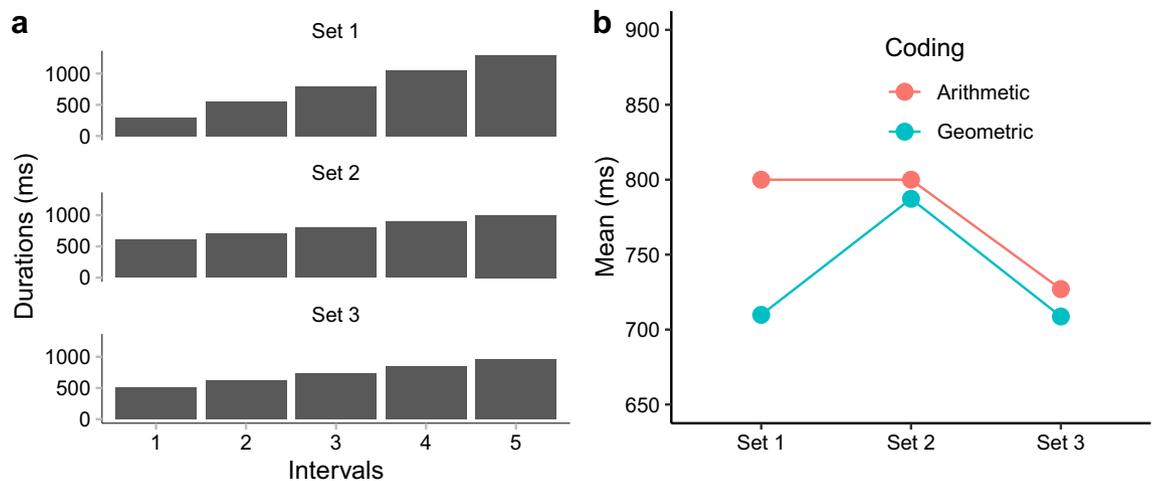


Figure 1. Illustration of three sets of intervals used in the study. **(a)** Three sets of intervals each of five intervals (Set 1: 300, 550, 800, 1050, 1300 ms; Set 2: 600, 700, 800, 900, 1000 ms; Set 3: 500, 610, 730, 840, 950 ms). The presentation order of the five intervals was randomized within each trial. **(b)** Predictions of ensemble averaging based on two hypothesized coding schemes: Linear Coding and, respectively, Logarithmic Coding. Sets 1 and 2 have the same arithmetic mean of 800 ms, which is larger than the arithmetic mean of the group 3 (727 ms). Sets 1 and 3 have the same geometric mean of 710 ms, which is smaller than the geometric mean of set 1 (787 ms).

comparing the subjective averaged duration directly to either the AM or the GM of the intervals. Accordingly, we carefully chose three sets of intervals, for which one set would yield a different average to the other sets according to each individual account (see Fig. 1). Each set contained five intervals—Set 1: 300, 550, 800, 1050, 1300 ms; Set 2: 600, 700, 800, 900, 1000 ms; and Set 3: 500, 610, 730, 840, 950 ms. Accordingly, Sets 1 and 2 have the same arithmetic mean (800 ms), which is larger than the arithmetic mean of Set 3 (727 ms). And Sets 1 and 3 have the same geometric mean (710 ms), which is shorter than the geometric mean of Set 2 (787 ms). The rationale was that, given the assumptions of linear and logarithmic representations make distinct predictions for the three sets, we may be able to infer the internal representation by observing the behavioral outcome based on the predictions.

Subjective durations are known to differ between visual and auditory signals^{5,55,56}, as our auditory system has higher temporal precision than the visual system. Often, sounds are judged longer than lights^{55,57}, where the difference is particularly marked when visual and auditory durations are presented intermixed in the same testing session⁵⁸. It has been suggested that time processing may be distributed in different modalities⁵⁹, and the internal pacemaker ‘ticks’ faster for the auditory than the visual modality⁵⁵. Accordingly, the processing strategies may potentially differ between the two modalities. Thus, in order to establish whether the internal representation of time is modality-independent, we tested both modalities using the same set of intervals in separate experiments.

Methods

Ethics statement. The methods and experimental protocols were approved by the Ethics Board of the Faculty of Pedagogy and Psychology at LMU Munich, Germany, and are in accordance with the Declaration of Helsinki 2008.

Participants. A total of 32 participants from the LMU Psychology community took part in the study, 1 of whom were excluded from further analyses due to lower-than-chance-level performance (i.e., temporal estimates exceeded 150% of the given duration). 16 participants were included in Experiment 1 (8 females, mean age of 22.2), and 15 participants were included in Experiment 2 (8 females, mean age of 26.4). Prior to the experiment, participants gave written informed consent and were paid for their participation of 8 Euros per hour. All reported a normal (or corrected-to-normal) vision, normal hearing, and no somatosensory disorders.

Stimuli. The experiments were conducted in a sound-isolated cabin, with dim incandescent background lighting. Participants sat approximately 60 cm from a display screen, a 21-inch CRT monitor (refresh rate 100 Hz; screen resolution 800 × 600 pixels). In Experiment 1, auditory stimuli (i.e., intervals) were delivered via two loudspeakers positioned just below the monitor, with a left-to-right separation of 40 cm. Brief auditory beeps (10 ms, 60 dB; frequency of 2500 or 3000 Hz, respectively) were presented to mark the beginning and end of the auditory intervals. In Experiment 2, the intervals were demarcated visually, namely, by presenting brief (10-ms) flashes of a gray disk (5° of visual angle in diameter, 21.4 cd/m²) in center of the display monitor against black screen background (1.6 cd/m²).

As for the length of the (five) successively presented intervals on a given trial, there were three sets: Set 1: 300, 550, 800, 1050, 1300 ms; Set 2: 600, 700, 800, 900, 1000 ms; and Set 3: 500, 610, 730, 840, 950 ms. These sets were constructed such that Sets 1 and 2 had the same arithmetic mean (800 ms), which is larger than the arithmetic mean of Set 3 (727 ms). And Sets 1 and 3 have the same geometric mean (710 ms), which is shorter

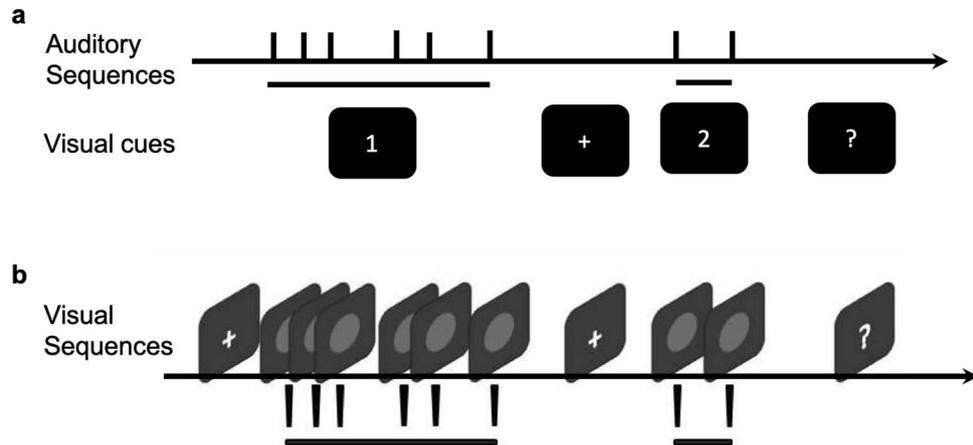


Figure 2. Schematic illustration of a trial in Experiments 1 and 2. **(a)** In Experiment 1, an auditory sequence of five intervals demarcated by six short (10-ms) auditory beeps of a particular frequency (either 2500 or 3000 Hz) was first presented together with a visual cue ‘1’. After a short gap with visual cue ‘+’, the second, comparison interval was demarcated by two beeps of a different frequency (either 3000 or 2500 Hz). A question mark prompts participants to respond if the mean interval of the first was longer or shorter than the second. **(b)** The temporal structure was essentially the same in Experiment 2 as in Experiment 1, except that the intervals were marked by a brief flash of a grey disk in the monitor center. Given that the task required a visual comparison, the two interval presentation phases were separated by a fixation cross.

than the geometric mean of Set 2 (787 ms). Of note, the order of the five intervals (of the presented set) was randomized on each trial.

Procedure. Two separate experiments were conducted, testing auditory (Experiment 1) and visual stimuli (Experiment 2), respectively. Each trial consisted of two presentation phases: successive presentation of five intervals, followed by the presentation of a single comparison interval. Participants’ task was to indicate, via a keypress response, whether the comparison interval was shorter or longer than the average of the five successive intervals. The response could be given without stress on speed.

In Experiment 1 (auditory intervals), trials started with a fixation cross presented for 500 ms, followed by a succession of five intervals demarcated by six 10-ms auditory beeps. Along with the onset of the auditory stimuli, a ‘1’ was presented on display monitor, telling participants that this was the first phase of the comparison task. The series of intervals was followed by a blank gap (randomly ranging between 800 and 1200 ms), with a fixation sign ‘+’ on the screen (indicating the transition to the comparison phase 2). After the gap, a single comparison duration demarcated by two brief beeps (10 ms) was presented, together with a ‘2’, indicating phase two of the comparison. Following another random blank gap (of 800–1200 ms), a question mark (?) appeared in the center of the screen, prompting participants to report whether the average interval of the first five (successive) intervals was longer or shorter than the second, comparison interval (Fig. 2a). Participants issued their response via the left or right arrow keys (on the keyboard in front of them) using their two index fingers, corresponding to either ‘shorter’ or ‘longer’ judgments. To make the two parts 1 and 2 of the interval presentation clearly distinguishable, two different frequencies (2500 and 3000 Hz) were randomly assigned to the first and, respectively, the second set of auditory interval markers.

Experiment 2 (visual intervals) was essentially the same as Experiment 1, except that the intervals were delivered via the visual modality and were demarcated by brief (10-ms) flashes of gray disks in the screen center (see Fig. 2b). Also, the visual cue signals used to indicate the two interval presentation phases (‘1’, ‘2’) in the ‘auditory’ Experiment 1 were omitted, to ensure participants’ undivided attention to the judgment-relevant intervals.

In order to obtain, in an efficient manner, reliable estimates of both the point of subjective equality (PSE) and the just noticeable difference (JND) of the psychometric function of the interval comparison, we employed the updated maximum-likelihood (UML) adaptive procedure from the UML toolbox for Matlab⁶⁰. This toolbox permits multiple parameters of the psychometric function, including the threshold, slope, and lapse rate (i.e., the probability of an incorrect response, which is independent of stimulus interval) to be estimated simultaneously. We chose the logistic function as the basic psychometric function and set the initial comparison interval to 500 ms. The UML adaptive procedure then used the method of maximum-likelihood estimation to determine the next comparison interval based on the participant’s responses to minimize the expected variance (i.e., uncertainty) in the parameter space of the psychometric function. In addition, after each response, the UML updated the posterior distributions of the psychometric parameters (see Fig. 3b for an example), from which the PSE and JND can be estimated (for the detailed procedure, see Shen et al.⁶⁰). To mitigate habituation and expectation effects, we presented the sequences of comparison intervals for the three different sets randomly intermixed across trials, concurrently tracking the three separate adaptive procedures.

Prior to the testing session, participants were given verbal instructions and then familiarized with the task in a practice block of 30 trials (10 comparison trials for each set). Of note, upon receiving the instruction, most

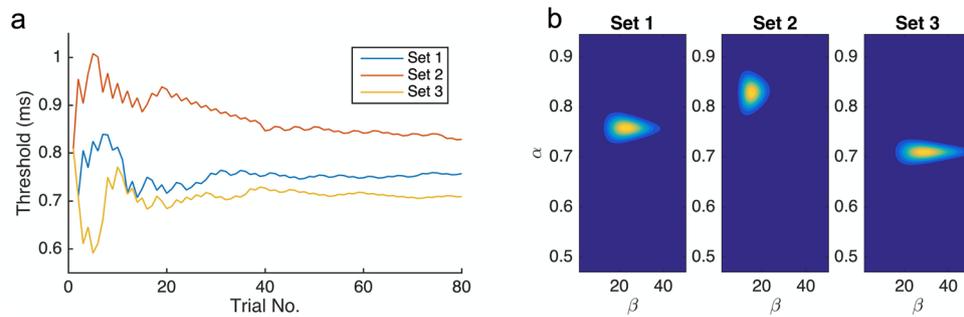


Figure 3. (a) Trial-wise update of the threshold estimate (α) for the three different interval sets in Experiment 1, for one typical participant. (b) The posterior parameter distributions of the threshold (α) and slope (β) based on the logistic function $p = 1/(1 + e^{-(x-\alpha)\cdot\beta})$, separately for the three sets (240 trials in total) for the same participant.

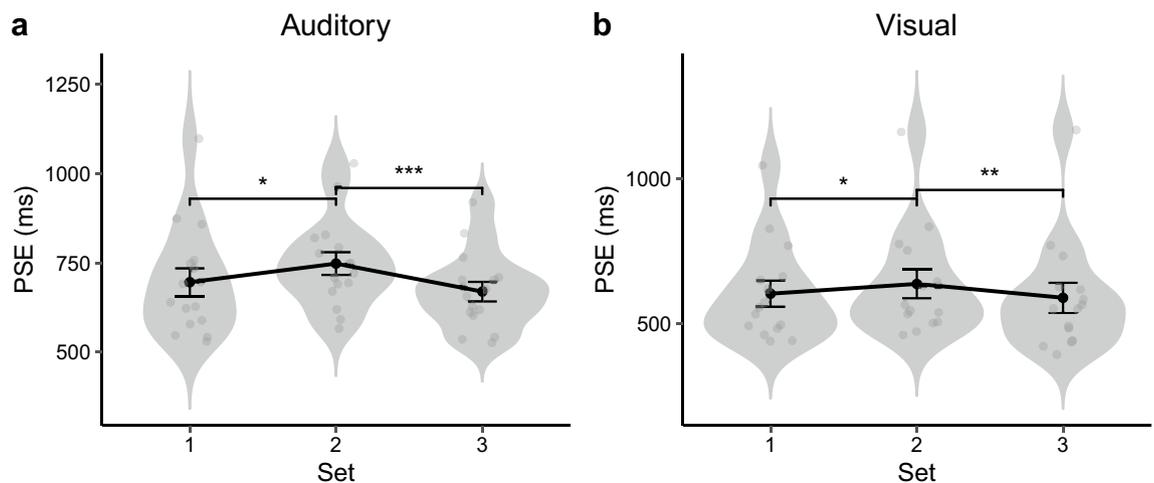


Figure 4. Violin plot of the distribution of individual subjective mean intervals (gray dots) of three tested sets, with the grand mean PSE (and associated standard error) overlaid on the respective set, separately for Experiment 1 (a) and Experiment 2 (b). *denotes $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

participants spontaneously voiced concern about the difficulty of the judgment they were asked to make. However, after performing just a few trials of the training block, they all expressed confidence that the task was easily doable after all, and they all went on to complete the experiment successfully. In the formal testing session, each of the three sets was tested 80 times, yielding a total of 240 trials per experiment. The whole experiment took some 60 min to complete.

Statistical analysis. All statistical tests were conducted using repeated-measures ANOVAs—with additional Bayes-Factor analyses (using JASP software) to comply with the more stringent criteria required for acceptance of the null hypothesis^{61,62}. All Bayes factors reported for ANOVA main effects are “inclusion” Bayes factors calculated across matched models. Inclusion Bayes factors compare models with a particular predictor to models that exclude that predictor, providing a measure of the extent to which the data support inclusion of a factor in the model. The Holm–Bonferroni method and Bayes factor have been applied for the post-hoc analysis.

Results

Figure 3 depicts the UML estimation for one typical participant: the threshold (α) and the slope (β) parameters of the logistic function $p = 1/(1 + e^{-(x-\alpha)\cdot\beta})$. By visual inspection, the thresholds reached stable levels within 80 trials of dynamic updating (Fig. 3a), and the posterior distributions (Fig. 3b) indicate the two parameters were converged in all three sets.

Figure 4 depicts the mean thresholds (PSEs), averaged across participants, for the three sets of intervals, separately for the auditory Experiment 1 and the visual Experiment 2. In both experiments, the estimated averages from the three sets showed a similar pattern, with the mean of Set 2 being larger than the means of both Set 1 and Set 3. Repeated-measures ANOVAs, conducted separately for both experiments, revealed the Set (main) effect to be significant both for Experiment 1, $F(2, 30) = 10.1$, $p < 0.001$, $\eta_g^2 = 0.064$, $BF_{incl} = 58.64$, and for Experiment 2, $F(2, 28) = 8.97$, $p < 0.001$, $\eta_g^2 = 0.013$, $BF_{incl} = 30.34$. Post-hoc Bonferroni-corrected comparisons confirmed

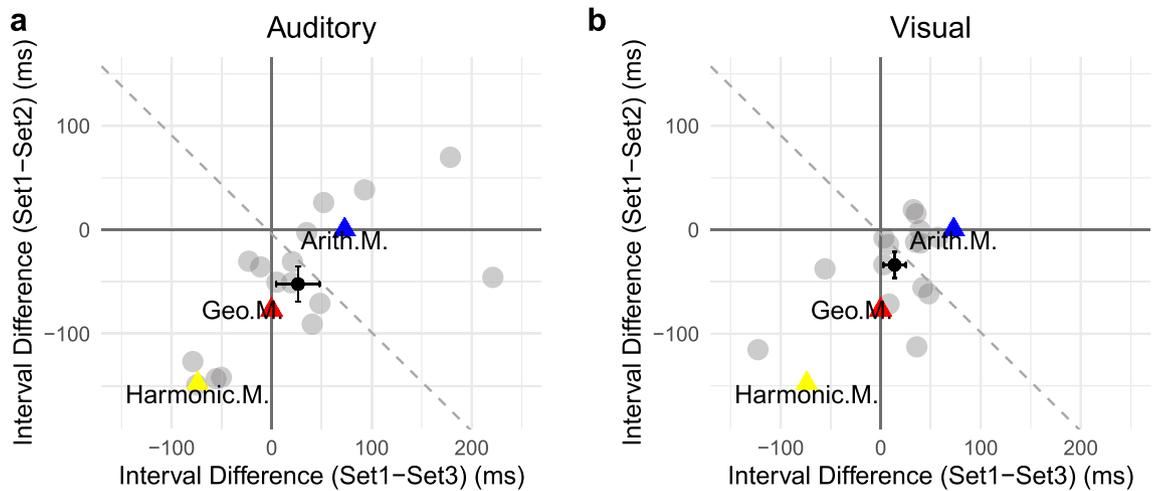


Figure 5. Difference in PSEs between Sets 1 and 2 plotted against the difference between Sets 1 and 3 for all individuals (gray dots) in Experiments 1 (a) and 2 (b). The dark triangles represent the ideal locations of arithmetic averaging (Arith.M) and geometric averaging (Geo.M). The black dots, with the standard-error bars, depict the mean differences across all participants. The dashed lines represent the line of reflection between the ‘geometric’ and ‘arithmetic’ ideal locations.

the Set effect to be mainly due to the mean being highest with Set 2. In more detail, for the auditory experiment (Fig. 4a), the mean of Set 2 was larger than the means of Set 1 [$t(15) = 3.14, p = 0.013, BF_{10} = 7.63$] and Set 3 [$t(15) = 5.12, p < 0.001, BF_{10} = 234$], with no significant difference between the latter ($t(15) = 1.26, p = 0.23, BF_{10} = 0.5$). The result pattern was similar for the visual experiment (Fig. 4b), with Set 2 generating a larger mean than both Set 1 ($t(14) = 3.13, p = 0.015, BF_{10} = 7.1$) and Set 3 ($t(14) = 4.04, p < 0.01, BF_{10} = 32.49$), with no difference between the latter ($t(14) = 1.15, p = 0.80, BF_{10} = 0.46$). This pattern of PSEs (Set 2 > Set 1 = Set 3) is consistent with one of our predictions, namely, that the main averaging process for rendering perceptual summary statistics is based on the geometric mean, in both the visual and the auditory modality.

To obtain a better picture of individual response patterns and assess whether they are more in line with one or the other predicted pattern illustrated in Fig. 1b, we calculated the PSE differences between Sets 1 and 2 and between Sets 1 and 3 as two indicators. Figure 5 depicts the difference between Sets 1 and 2 over the difference between Sets 1 and 3, for each participant. The ideal differences between the respective arithmetic means and the respective geometric means are located on the orthogonal axes (triangle points). By visual inspection, individuals (gray dots) differ considerably: while many are closer to the geometric than to the arithmetic mean, some show the opposite pattern. We used the line of reflection between the ‘arithmetic’ and ‘geometric’ points to separate participants into two groups: geometric- and arithmetic-oriented groups. Eleven (out of 16) participants exhibited a pattern oriented towards the geometric mean in Experiment 1, and nine (out of 15) in Experiment 2. Thus, geometric-oriented individuals outnumbered arithmetic-oriented individuals (7:3 ratio). Consistent with the above PSE analysis, the grand mean differences (dark dots in Fig. 5) and their associated standard errors are located within the geometric-oriented region.

Of note, however, while the mean patterns across three sets are in line with the prediction of geometric interval averaging (see the pattern illustrated in Fig. 1b) for both experiments, the absolute PSEs were shorter in the visual than in the auditory conditions. Further tests confirmed that, in the ‘auditory’ Experiment 1, the mean PSEs did not differ significantly from their correspondent physical geometric means (one-sample Bayesian *t*-test pooled across the three sets), $t(47) = 1.70, p = 0.097, BF_{10} = 0.587$, but they were significant smaller than the physical arithmetic means, $t(47) = 3.87, p < 0.001, BF_{10} = 76.5$. In the ‘visual’ Experiment 2, the mean PSEs for all three interval sets were significantly smaller than both the physical geometric mean [$t(44) = 4.74, p < 0.001, BF_{10} = 924.1$] and the arithmetic mean [$t(44) = 6.23, p < 0.001, BF_{10} > 1000$]. Additionally, the estimated mean durations were overall shorter for the visual (Experiment 2) versus the auditory intervals (Experiment 1), $t(91) = 2.97, p < 0.01, BF_{10} = 9.64$. This modality effect is consistent with previous reports that auditory intervals are often perceived as longer than physically equivalent visual intervals^{55,63}.

Another key parameter providing an indicator of an observer’s temporal sensitivity (resolution) is given by the just noticeable difference (JND), defined as the interval difference between the 50%- and 75%-thresholds estimated from the psychometric function. Figure 6 depicts the JNDs obtained in Experiments 1 and 2, separately for the three sets of intervals. Repeated-measures ANOVAs, with Set as the main factor, failed to reveal any differences among the three sets, for either experiment [Experiment 1: $F(2, 30) = 1.05, p = 0.36, BF_{incl} = 0.325$; Experiment 2: $F(2, 28) = 0.166, p = 0.85, BF_{incl} = 0.156$]. Comparison across Experiments 1 and 2, however, revealed the JNDs to be significantly smaller for auditory than for visual interval averaging, $t(91) = 2.95, p < 0.01, BF_{10} = 9.08$. That is, temporal resolution was higher for the auditory than for the visual modality, consistent with the literature⁶⁴.

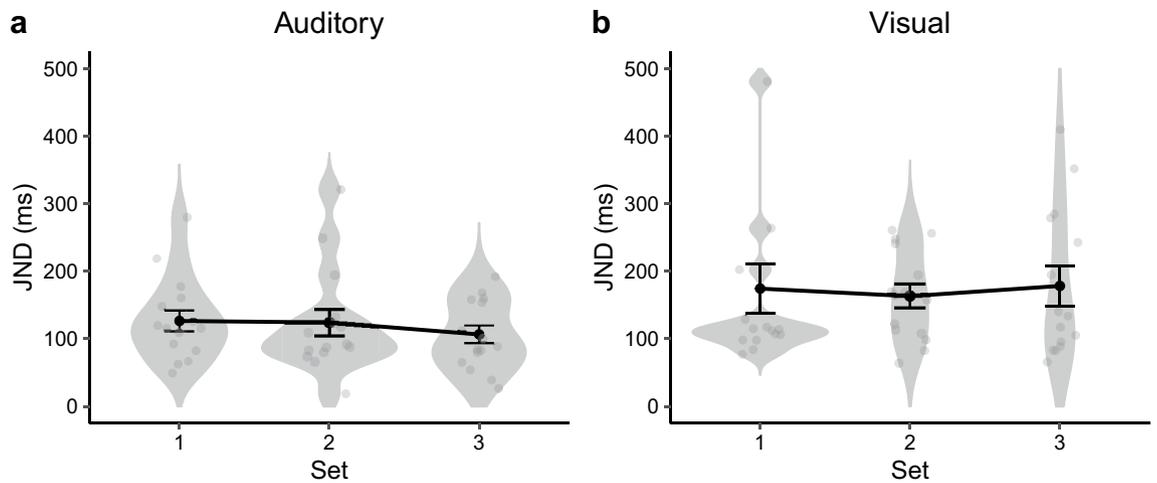


Figure 6. Violin plot of the distribution of individual JNDs (gray dots) of three tested sets, with the mean JND (and associated standard error) overlaid on the respective set, separately for Experiment 1 (a) and Experiment 2 (b).

Thus, taken together, evaluation of both the mean and sensitivity of the participants' interval estimates demonstrated not only that ensemble coding in the temporal domain is accurate and consistent, but also that the geometric mean is used as the predominant averaging scheme for performing the task.

Model simulations. Although our results favor the geometric averaging scheme, one might argue that participants adopt alternative schemes to simple, equally weighted, arithmetic or geometric averaging. For instance, the weight of an interval in the averaging process might be influenced by the length or/and the position of that interval in the sequence. For example, a long interval might engage more attention than a short interval, and weights are assigned to intervals according to their lengths. Alternatively, short intervals might be assigned higher weights. This would be in line with an animal study⁶⁵, in which pigeons received reinforcement after varying delay intervals. The pigeons assigned greater weight to short delays, as reflected by an inverse relationship between delay and efficacy of reinforcement. In case each interval is weighted precisely relative to its inverse (reciprocal), the result would be harmonic averaging, that is: the reciprocal of the arithmetic mean of the reciprocals of the presented ensemble intervals (i.e., $M_h = \left(\sum_{i=1}^n \frac{1}{x_i}\right)^{-1}$). A daily example of the harmonic mean is that when one drives from A to B at a speed of 90 km/h and returns with 45 km/h, the average speed is the harmonic mean of 60 km/h, not the arithmetic or the geometric mean.

To further examine how closely the perceived ensemble means, reflected by the PSEs, match what would be expected if participants had been performing different types of averaging (arithmetic, geometric, weighted, and harmonic), as well as to explore the effect of an underestimation bias that we observed for the visual modality, we compared and contrasted four model simulations. All four models assume that each interval was corrupted by noise, where the noise scales with interval length according to the scalar property¹.

In more detail, the arithmetic-, weighted-, and harmonic-mean models all assume that each perceived interval is corrupted by normally distributed noise which follows the scalar property:

$$T_i \sim N(\mu_i, \mu_i w_f),$$

where T_i is the perceived duration of interval i , μ_i is its physical duration, and w_f is the Weber scaling. In contrast, the geometric-averaging model assumes that the internal representation of each interval is encoded on a logarithmic timeline, and all intervals are equally affected by the noise, which implicitly incorporates the scalar property:

$$\log(T_i) \sim N(\log(\mu_i), \sigma_t),$$

where σ_t is the standard deviation of the noise.

Given that the perceived duration is subject to various types of contextual modulation (such as the central-tendency bias^{28–30}) and modality differences⁵⁵, individual perceived intervals might be biased. To simplify the simulation, we assume a general bias in ensemble averaging, which follows the normal distribution:

$$B \sim N(\mu_b, \sigma).$$

Accordingly, the arithmetic (M_A) and harmonic (M_H) average of the five intervals in our experiments are given by:

$$M_A = \sum_{i=1}^5 T_i / 5 + B,$$

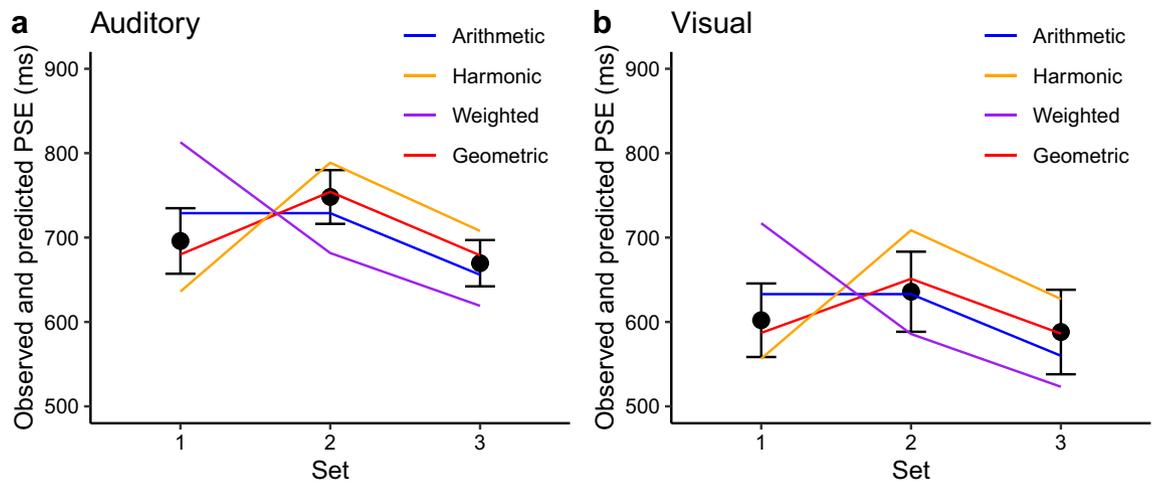


Figure 7. Predicted and observed PSE's for Experiment 1 (a) and Experiment 2 (b). The filled circles show the observed PSE's (i.e. the grand mean PSE's, which are also shown in Fig. 4, and the error bars represent the associated standard errors); the lines represent the predictions of the four models described in the text.

$$M_H = 5 / \left(\sum_1^5 1/T_i \right) + B.$$

In the weighted-mean model, the intervals are weighted by their relative duration within the set, and the weighted intervals are subject to normally distributed noise and averaged, with a general bias added to the average:

$$M_W = \sum_1^5 (w_i T_i) / 5 + B,$$

where the weight $w_i = \mu_i / \sum_1^5 \mu_i$.

The geometric-mean model assumes that the presented intervals are first averaged on a logarithmic scale, and corrupted independently by noise and the general bias, while the ensemble average is then back-transformed into the linear scale for 'responding':

$$M_G = e^{\frac{\sum_1^5 \log(T_i) / 5 + B}{1}}.$$

It should be noted that the comparison intervals could also be corrupted by noise. In addition, trial-to-trial variation of the comparison intervals may introduce the central-tendency bias^{28–30}. However, the central-tendency bias does not shift the mean PSE, which is the measure we focused on here. Thus, for the sake of simplicity, we omit the variation of the comparison intervals in the simulation. Evaluation of each of the above models was based on 100,000 rounds of simulation for each interval set (per model). For the arithmetic, geometric, and weighted means, the noise parameters (w_f and σ) make no difference to the average prediction, given that, over a large number of simulations, the influence of noise on the linear interval averaging would be zero (i.e., the mean of the noise distribution). Therefore, the predictions for these models are based on a noise-free model version (i.e., the noise parameters were set to zero), with the bias parameter (μ_b) chosen to minimize the sum of square distances between the model predictions and the average PSE's from each experiment. For the harmonic mean, owing to the non-linear transformation, the noise does make a difference to the average prediction and the best parameters, which minimize the sum of squared errors (i.e. the sum of squared differences between the model predictions and the observed PSE's), was determined by grid search, i.e. by evaluating the model for all combinations of parameters on a grid covering the range of the most plausible values for each parameter and finding the combination that minimized the error on that grid.

Among the four models, the model using the geometric mean provides the closest fit to the (pattern of the) average PSEs observed in both experiments (see Fig. 7). By visual inspection, across the three interval sets, the pattern of the average PSEs is the closest to that predicted by the geometric mean, which makes the same predictions for Sets 1 and 3. Note, though, that the PSE observed for Set 1 slightly differs from that for Set 3, by being shifted somewhat in the direction of the prediction based on the arithmetic mean (i.e., shifted towards the PSE for Set 2). The harmonic-mean model predicts that the PSE to be smaller for Set 1 as compared to Set 3, which was, however, not the case in either experiment. On the weighted-mean model, the PSE was expected to be the largest for Set 1, which differs even more from the observed PSE.

Furthermore, as is also clear by visual inspection, there was a greater bias in the direction of shorter durations in the visual compared to the auditory experiment (witness the lower PSEs in Fig. 7b compared to Fig. 7a), which was reflected in a difference in the bias parameter (μ_b). The value of the bias parameter associated with

the best fit of the geometric mean model was -0.04 for Experiment 1 (auditory) and -0.20 for Experiment 2 (visual), which correspond to a shortening by 4% in the auditory and by 18% in the visual experiment. For the arithmetic and weighted-mean models, both bias parameters reflect a larger degree of shortening compared to the geometric-mean model, while the bias parameters of the harmonic-mean model were somewhat smaller compared to the bias parameters of the geometric mean model.

Discussion

The aim of the present study was to reveal the internal encoding of subjective time by examining intuitive ensemble averaging in the time domain. The underlying idea was that ensemble summary statistics are computed at a low level of temporal processing, bypassing high-level cognitive decoding strategies. Accordingly, ensemble averaging of time intervals may directly reflect the fundamental internal representation of time. Thus, if the internal representation of the timeline is logarithmic, basic averaging should be close to the geometric mean (see Footnote 1); alternatively, if time intervals are encoded linearly, ensemble averaging should be close to the arithmetic mean. We tested these predictions by comparing and contrasting ensemble averaging for three sets of time intervals characterized by differential patterns of the geometric and arithmetic means (see Fig. 1b). Critically, the pattern of ensemble averages we observed most closely matched that of the geometric mean (rather than those of the arithmetic, weighted, or, respectively, harmonic means), and this was the case with both auditory (Experiment 1) and visual intervals (Experiment 2) (see results of modeling simulation in Fig. 7). Although some 30% of the participants appeared to prefer arithmetic averaging, the majority showed a pattern consistent with geometric averaging. These findings thus lend support to our central hypothesis: regardless of the sensory modality, intuitive ensemble averaging of time intervals (at least in the 300- to 1300-ms range) is based on logarithmically coded time, that is: the subjective timeline is logarithmically scaled.

Unlike ensemble averaging of visual properties (such as telling the mean size or mean facial expression of simultaneously presented objects), there is a pragmatic issue of how we can average (across time) in the temporal domain—in Wearden and Jones's¹⁶ words: 'can people do this at all?' (p. 1295). Wearden and Jones¹⁶ asked participants to average three consecutively presented durations and compare their mean to that of the subsequently comparison duration. They found that participants were indeed able to extract the (arithmetic) mean; moreover, the estimated means remained indifferent to variations in the spacing of the sample durations. In the current study, by adopting the averaging task for multiple temporal intervals (> 3), we resolved the problem encountered by the temporal bisection task, namely: it cannot be ruled out that finding of the bisection point to be nearest the geometric mean is the outcome of a ratio comparison^{24,25}, rather than reflecting the internal timeline (see "Introduction").

Specifically, we hypothesized that temporal ensemble perception may be indicative of a fast and intuitive process likely involving two stages: transformation, either linearly or nonlinearly, of the sample durations onto a subjective scale^{66–68} and storage in short-term (or working) memory (STM); followed by estimation of the average of the multiple intervals on the subjective scale and then remapping from the subjective to the objective scale. One might assume that the most efficient form of encoding would be linear, avoiding the need for nonlinear transformation. But this is at variance with our finding that, across the three sets of intervals, the averaging judgments followed the pattern predicted by logarithmic encoding (for both visual and auditory intervals). The use of logarithmic encoding may be owing to the limited capacity of STM: uncompressed intervals require more space ('bits') to store, as compared to logarithmically compressed intervals. The brain appears to have chosen the latter for efficient STM storage in the first stage. However, nonlinear, logarithmic encoding in stage 1 could give rise to a computational cost for the averaging process in stage 2: averaging intervals on the objective, external scale would require the individual encoded intervals to be first transformed back from the subjective to the objective scale, which, due to being computationally expensive, would reduce processing speed. By contrast, arithmetic averaging on the subjective scale would be computationally efficient, as it requires only one step of remapping—of the subjective averaged interval onto the objective scale. Intuitive ensemble processing of time appears to have opted for the latter, ensuring computational efficiency. Thus, given the subjective scale is logarithmic, intuitive averaging would yield the geometric mean.

It could, of course, be argued that participants may adopt alternative weighting schemes to simple (equally weighted) arithmetic or geometric averaging. For example, the weight of an interval in the averaging process might be influenced by the length of that interval or/and the position of that interval within the sequence. Thus, for example, a long interval might engage more attention than a short interval, and weights are assigned to the intervals according to their lengths. Alternatively, greater weight might be assigned to shorter intervals, consistent with animal studies. For instance, Killen⁶⁵, in a study with pigeons, found that trials with short-delay reinforcement (with food tokens) had higher impact than trials with long-delay reinforcement, biasing the animals to respond earlier than the arithmetic and geometric mean interval, but close to the harmonic mean. We simulated such alternative averaging strategies—finding that the prediction of geometric averaging was still superior to those of arithmetic, weighted, and, respectively, harmonic averaging: none of the three alternative averaging schemes could explain the patterns we observed in Experiments 1 and 2 better than the geometric averaging. Thus, we are confident that intuitive ensemble averaging is best predicted by the geometric mean. Of course, it would be possible to think of various other, complex weighting schemes that we did not explore in our modeling. However, based on Occam's razor, our observed data patterns favor the simple geometric averaging account.

Logarithmic representation of stimulus intensity, such as of loudness or weight, has been proposed by Fechner over one and a half centuries ago⁶⁹, based on the fact that the JND is proportionate to stimulus intensity (Weber's law). It has been shown that, for the same amount of information (quantized levels), the logarithmic scale provides the minimal expected relative error that optimizes communication efficiency, given that neural storage of sensory or magnitude information is capacity-limited⁷⁰. Accordingly, logarithmic timing would provide a good

solution for coping with limited STM capacity to represent longer intervals. However, as argued by Gallistel⁷¹, logarithmic encoding makes valid computations problematic: “Unless recourse is had to look-up tables, there is no way to implement addition and subtraction, because the addition and subtraction of logarithmic magnitudes corresponds to the multiplication and division of the quantities they refer to” (p. 8). We propose that the ensuing computational complexity pushed intuitive ensemble averaging onto the internal, subjective scale—rather than the external, objective scale, which would have required multiple nonlinear transformations. Thus, our results join the increasing body of studies suggesting that, like other magnitudes^{72,73}, time is represented internally on a logarithmic scale and intuitive averaging processes are likely bypassing higher-level cognitive computations. Higher-level computations based on the external, objective scale can be acquired through educational training, and this is linked to mathematical competency^{37,72,74}. Such high-level computations are likely to become involved (at least to some extent) in magnitude estimation, which would explain why investigations of interval averaging have produced rather mixed results^{15,16,31}. Even in the present study, the patterns exhibited by some of the participants could not be explained by purely geometric encoding, which may well be attributable to the involvement of such higher processes. Interestingly, a recent study reported that, under dual-task conditions with an attention-demanding secondary task taxing visual working memory, the mapping of number onto space changed from linear to logarithmic⁷⁵. This provides convergent support for our proposal of an intuitive averaging process that operates with a minimum of cognitive resources.

Another interesting finding of the present study concerns the overall underestimation of the (objective) mean interval duration, which was evident for all three sets of intervals and for both modalities (though it was more marked with visual intervals). This general underestimation is consistent with the subjective ‘shortening effect’: a source of bias reducing individual durations in memory^{76,77}. The underestimation was less pronounced in the auditory (than the visual) modality, consistent with the classic ‘modality effect’ of auditory events being judged as longer than visual events. The dominant account of this is that temporal information is processed with higher resolution in the auditory than in the visual domain^{30,55,58,78}. Given the underestimation bias, our analysis approach was to focus on the global pattern of observed ensemble averages across multiple interval sets, rather than examining whether the estimated average for each individual set was closer to the arithmetic or the geometric mean. We did obtain a consistent pattern across all three sets and for both modalities, underpinned by strong statistical power. We therefore take participants’ performance to genuinely reflect an intuitive process of temporal ensemble averaging, where the average lies close to the geometric mean.

Another noteworthy finding was that the JNDs were larger in the visual than in the auditory modality (Fig. 6), indicative of higher uncertainty, or more random guessing, in ensemble averaging in the visual domain. As random guessing would corrupt the effect we aimed to observe^{79–81}, this factor would have obscured the underlying pattern more in the visual than in the auditory modality. To check for such a potential impact of random responses on temporal averaging, we fitted additional psychometric functions to the original response data from our *visual* experiment. These fits used the logistic psychometric function with and without a lapse-rate parameter, as well as a mixed model—of both temporal responses, modeled by a gamma distribution, and non-temporal responses, modeled by an exponential distribution—proposed by Laude et al.⁸¹, and finally a model with the non-temporal component from the model of Laude et al. combined with the logistic psychometric function. We found that the model of Laude et al. did not improve the quality of the fit sufficiently to justify the extra parameters, as evaluated using the Akaike Information Criterion (AIC), and adding a lapse rate improved the AIC only slightly (average AIC: logistic with no lapse rate: 99.1, gamma with non-temporal responses: 102, logistic with non-temporal responses: 99.3, and logistic with lapse rate: 97.9). Importantly, the overall pattern of the PSEs remained the same when the PSEs were estimated from a psychometric function with a lapse rate parameter (set 1: 591 ms; set 2: 629 ms; set 3: 578 ms): the PSE remained significantly larger for Set 2 compared to Set 1 ($t(14) = 2.56$, $p = 0.02$) and for Set 2 compared to Set 3 ($t(14) = 2.84$, $p = 0.01$), without a significant difference between Sets 1 and 3 ($t(14) = 0.76$, $p = 0.46$). Thus, the pattern we observed is rather robust (it does not appear to have been affected substantially by random guessing), favoring geometric averaging not only in the auditory but also in the visual modality.

In summary, the present study provides behavioral evidence supporting a logarithmic representation of subjective time, and that intuitive ensemble averaging is based on the geometric mean. Even though the validity of behavioral studies is being increasingly acknowledged, achieving a full understanding of human timing requires a concerted research effort from both the psychophysical and neural perspectives. Accordingly, future investigations (perhaps informed by our work) would be required to reveal the—likely logarithmic—neural representation of the inner timeline.

Data availability

The data and codes for all experiments are available at: <https://github.com/msenselab/Ensemble.OpenCodes>.

Received: 3 January 2020; Accepted: 7 October 2020

Published online: 23 October 2020

References

- Gibbon, J. Scalar expectancy theory and Weber’s law in animal timing. *Psychol. Rev.* **84**, 279–325 (1977).
- Church, R. M. Properties of the internal clock. *Ann. N. Y. Acad. Sci.* **423**, 566–582 (1984).
- Buhusi, C. V. & Meck, W. H. What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* **6**, 755–765 (2005).
- Eagleman, D. M. & Pariyadath, V. Is subjective duration a signature of coding efficiency?. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 1841–1851 (2009).
- Matthews, W. J. & Meck, W. H. Time perception: The bad news and the good. *Wiley Interdiscip. Rev.* **5**, 429–446 (2014).

6. Matell, M. S. & Meck, W. H. Cortico-striatal circuits and interval timing: Coincidence detection of oscillatory processes. *Cogn. Brain Res.* **21**, 139–170 (2004).
7. Matell, M. S., Meck, W. H. & Nicolelis, M. A. L. Interval timing and the encoding of signal duration by ensembles of cortical and striatal neurons. *Behav. Neurosci.* **117**, 760–773 (2003).
8. Buonomano, D. V. & Karmarkar, U. R. How do we tell time? *Neuroscientist* **8**, 42–51 (2002).
9. Gu, B. M., van Rijn, H. & Meck, W. H. Oscillatory multiplexing of neural population codes for interval timing and working memory. *Neurosci. Biobehav. Rev.* **48**, 160–185 (2015).
10. Oprisan, S. A. & Buhusi, C. V. Modeling pharmacological clock and memory patterns of interval timing in a striatal beat-frequency model with realistic, noisy neurons. *Front. Integr. Neurosci.* **5**, 52 (2011).
11. Wilkes, J. T. & Gallistel, C. R. Information theory, memory, prediction, and timing in associative learning. In *Computational Models of Brain and Behavior* (ed. Moustafa, A. A.). <https://doi.org/10.1002/9781119159193.ch35> (2017).
12. Simen, P., Balci, F., de Souza, L., Cohen, J. D. & Holmes, P. A model of interval timing by neural integration. *J. Neurosci.* **31**, 9238–9253 (2011).
13. Balci, F. & Simen, P. Decision processes in temporal discrimination. *Acta Psychol.* **149**, 157–168 (2014).
14. Simen, P., Vlasov, K. & Papadakis, S. Scale (in)variance in a unified diffusion model of decision making and timing. *Psychol. Rev.* **123**, 151–181 (2016).
15. Wearden, J. H. & Ferrara, A. Stimulus spacing effects in temporal bisection by humans. *Q. J. Exp. Psychol. B* **48**, 289–310 (1995).
16. Wearden, J. H. & Jones, L. A. Is the growth of subjective time in humans a linear or nonlinear function of real time? *Q. J. Exp. Psychol.* **60**, 1289–1302 (2007).
17. Brown, G. D. A., McCormack, T., Smith, M. & Stewart, N. Identification and bisection of temporal durations and tone frequencies: Common models for temporal and nontemporal stimuli. *J. Exp. Psychol. Hum. Percept. Perform.* **31**, 919–938 (2005).
18. Yi, L. Do rats represent time logarithmically or linearly? *Behav. Process.* **81**, 274–279 (2009).
19. Gibbon, J. & Church, R. M. Time left: Linear versus logarithmic subjective time. *J. Exp. Psychol.* **7**, 87–108 (1981).
20. Jozefowicz, J., Gaudichon, C., Mekkas, F. & Machado, A. Log versus linear timing in human temporal bisection: A signal detection theory study. *J. Exp. Psychol. Anim. Learn. Cogn.* **44**, 396–408 (2018).
21. Kopec, C. D. & Brody, C. D. Human performance on the temporal bisection task. *Brain Cogn.* **74**, 262–272 (2010).
22. Church, R. M. & Deluty, M. Z. Bisection of temporal intervals. *J. Exp. Psychol. Anim. Behav. Process.* **3**, 216–228 (1977).
23. Stubbs, D. A. Scaling of stimulus duration by pigeons. *J. Exp. Anal. Behav.* **26**, 15–25 (1976).
24. Allan, L. G. & Gibbon, J. Human bisection at the geometric mean. *Learn. Motiv.* **22**, 39–58 (1991).
25. Allan, L. G. The influence of the scalar timing model on human timing research. *Behav. Process.* **44**, 101–117 (1998).
26. Penney, T. B., Brown, G. D. A. & Wong, J. K. L. Stimulus spacing effects in duration perception are larger for auditory stimuli: Data and a model. *Acta Psychol.* **147**, 97–104 (2014).
27. Lejeune, H. & Wearden, J. H. Vierordt's the experimental study of the time sense (1868) and its legacy. *Eur. J. Cogn. Psychol.* **21**, 941–960 (2009).
28. Hollingworth, H. L. The central tendency of judgment. *J. Philos. Psychol. Sci. Methods* **7**, 461–469 (1910).
29. Jazayeri, M. & Shadlen, M. N. Temporal context calibrates interval timing. *Nat. Neurosci.* **13**, 1020–1026 (2010).
30. Shi, Z., Church, R. M. & Meck, W. H. Bayesian optimization of time perception. *Trends Cogn. Sci.* **17**, 556–564 (2013).
31. Jones, M. R. & McAuley, J. D. Time judgments in global temporal contexts. *Percept. Psychophys.* **67**, 398–417 (2005).
32. Dehaene, S., Izard, V., Spelke, E. & Pica, P. Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science* **320**, 1217–1220 (2008).
33. Siegler, R. S. & Booth, J. L. Development of numerical estimation in young children. *Child Dev.* **75**, 428–444 (2004).
34. Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S. & Zorzi, M. Numerical estimation in preschoolers. *Dev. Psychol.* **46**, 545–551 (2010).
35. Booth, J. L. & Siegler, R. S. Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* **42**, 189–201 (2006).
36. Barth, H. C. & Paladino, A. M. The development of numerical estimation: Evidence against a representational shift. *Dev. Sci.* **14**, 125–135 (2011).
37. Libertus, M. E., Feigenson, L. & Halberda, J. Preschool acuity of the approximate number system correlates with school math ability. *Dev. Sci.* **14**, 1292–1300 (2011).
38. Ariely, D. Seeing sets: Representation by statistical properties. *Psychol. Sci.* **12**, 157–162 (2001).
39. Whitney, D. & Yamanashi Leib, A. Ensemble perception. *Annu. Rev. Psychol.* **69**, 105–129 (2018).
40. Chong, S. C. & Treisman, A. Representation of statistical properties. *Vis. Res.* **43**, 393–404 (2003).
41. Chong, S. C. & Treisman, A. Statistical processing: Computing the average size in perceptual groups. *Vis. Res.* **45**, 891–900 (2005).
42. Webster, J., Kay, P. & Webster, M. A. Perceiving the average hue of color arrays. *J. Opt. Soc. Am. A* **31**, A283 (2014).
43. Haberman, J. & Whitney, D. Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In *Oxford Series in Visual Cognition. From Perception to Consciousness: Searching with Anne Treisman* (Eds Wolfe, J. & Robertson, L.) 339–349. <https://doi.org/10.1093/acprof:osobl/9780199734337.003.0030> (Oxford University Press, 2012).
44. Kramer, R. S. S., Ritchie, K. L. & Burton, A. M. Viewers extract the mean from images of the same person: A route to face learning. *J. Vis.* **15**, 1–9 (2015).
45. Leib, A. Y., Kosovicheva, A. & Whitney, D. Fast ensemble representations for abstract visual impressions. *Nat. Commun.* **7**, 13186 (2016).
46. Haberman, J. & Whitney, D. Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* **17**, 751–753 (2007).
47. Curtis, D. W. & Mullin, L. C. Judgments of average magnitude: Analyses in terms of the functional measurement and two-stage models. *Percept. Psychophys.* **18**, 299–308 (1975).
48. Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A. & Whitney, D. Humans use summary statistics to perceive auditory sequences. *Psychol. Sci.* **24**, 1389–1397 (2013).
49. Anderson, N. H. Application of a weighted average model to a psychophysical averaging task. *Psychon. Sci.* **8**, 227–228 (1967).
50. Schweickert, R., Han, H. J., Yamaguchi, M. & Fortin, C. Estimating averages from distributions of tone durations. *Atten. Percept. Psychophys.* **76**, 605–620 (2014).
51. Chen, L., Zhou, X., Müller, H. J. & Shi, Z. What you see depends on what you hear: Temporal averaging and crossmodal integration. *J. Exp. Psychol. Gen.* **147**, 1851–1864 (2018).
52. Hellström, Å. The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychol. Bull.* **97**, 35–61 (1985).
53. Le Dantec, C. et al. ERPs associated with visual duration discriminations in prefrontal and parietal cortex. *Acta Psychol.* **125**, 85–98 (2007).
54. Shi, Z., Ganzenmüller, S. & Müller, H. J. Reducing bias in auditory duration reproduction by integrating the reproduced signal. *PLoS ONE* **8**, e62065 (2013).
55. Wearden, J. H., Edwards, H., Fakhri, M. & Percival, A. Why “Sounds Are Judged Longer Than Lights”: Application of a model of the internal clock in humans? *Q. J. Exp. Psychol. Sect. B* **51**, 97–120 (1998).
56. Wearden, J. H. When do auditory/visual differences in duration judgments occur? *Q. J. Exp. Psychol.* **59**, 1709–1724 (2006).
57. Goldstone, S. & Lhamon, W. T. Studies of auditory-visual differences in human time judgment. 1. Sounds are judged longer than lights. *Percept. Mot. Skills* **39**, 63–82 (1974).

58. Penney, T. B., Gibbon, J. & Meck, W. H. Differential effects of auditory and visual signals on clock speed and temporal memory. *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 1770–1787 (2000).
59. Ivry, R. B. & Schlerf, J. E. Dedicated and intrinsic models of time perception. *Trends Cogn. Sci.* **12**, 273–280 (2008).
60. Shen, Y., Dai, W. & Richards, V. M. A MATLAB toolbox for the efficient estimation of the psychometric function using the updated maximum-likelihood adaptive procedure. *Behav. Res. Methods* **47**, 13–26 (2015).
61. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
62. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. & Iverson, G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16**, 225–237 (2009).
63. Ganzenmüller, S., Shi, Z. & Müller, H. J. Duration reproduction with sensory feedback delay: Differential involvement of perception and action time. *Front. Integr. Neurosci.* **6**, 1–11 (2012).
64. Shipley, T. Auditory flutter-driving of visual flicker. *Science* **145**, 1328–1330 (1964).
65. Killeen, P. On the measurement of reinforcement frequency in the study of preference. *J. Exp. Anal. Behav.* **11**, 263–269 (1968).
66. Gibbon, J. The structure of subjective time: How time flies. *Psychol. Learn. Motiv.* [https://doi.org/10.1016/s0079-7421\(08\)60017-1](https://doi.org/10.1016/s0079-7421(08)60017-1) (1986).
67. Johnson, K. O., Hsiao, S. S. & Yoshioka, T. Neural coding and the basic law of psychophysics. *Neuroscientist* **8**, 111–121 (2002).
68. Taatgen, N. A., van Rijn, H. & Anderson, J. An integrated theory of prospective time interval estimation: The role of cognition, attention, and learning. *Psychol. Rev.* **114**, 577–598 (2007).
69. Fechner, G. T. *Elemente der Psychophysik*, Vol. I and II (Breitkopf and Härtel, Leipzig, 1860).
70. Sun, J. Z., Wang, G. I., Goyal, V. K. & Varshney, L. R. A framework for Bayesian optimality of psychophysical laws. *J. Math. Psychol.* **56**, 495–501 (2012).
71. Gallistel, C. R. Mental magnitudes. In *Space, Time, and Number in the Brain: Searching for the Foundations of Mathematical Thought* (eds Dehaene, S. & Brannon, E. M.) 3–12 (Elsevier, Amsterdam, 2011).
72. Dehaene, S. Subtracting pigeons: Logarithmic or linear?. *Psychol. Sci.* **12**, 244–246 (2001).
73. Roberts, W. A. Evidence that pigeons represent both time and number on a logarithmic scale. *Behav. Proc.* **72**, 207–214 (2006).
74. Anobile, G. *et al.* Spontaneous perception of numerosity in pre-school children. *Proc. Biol. Sci.* **286**, 20191245 (2019).
75. Anobile, G., Cicchini, G. M. & Burr, D. C. Linear mapping of numbers onto space requires attention. *Cognition* **122**, 454–459 (2012).
76. Meck, W. H. Selective adjustment of the speed of internal clock and memory processes. *J. Exp. Psychol. Anim. Behav. Process.* **9**, 171–201 (1983).
77. Spetch, M. L. & Wilkie, D. M. Subjective shortening: A model of pigeons' memory for event duration. *J. Exp. Psychol. Anim. Behav. Process.* **9**, 14–30 (1983).
78. Gu, B. M., Cheng, R. K., Yin, B. & Meck, W. H. Quinpirole-induced sensitization to noisy/sparse periodic input: Temporal synchronization as a component of obsessive-compulsive disorder. *Neuroscience* **179**, 143–150 (2011).
79. Daniels, C. W. & Sanabria, F. Interval timing under a behavioral microscope: Dissociating motivational and timing processes in fixed-interval performance. *Learn. Behav.* **45**, 29–48 (2017).
80. Lejeune, H. & Wearden, J. H. The comparative psychology of fixed-interval responding: Some quantitative analyses. *Learn. Motiv.* **22**, 84–111 (1991).
81. Laude, J. R., Daniels, C. W., Wade, J. C. & Zentall, T. R. I can time with a little help from my friends: effect of social enrichment on timing processes in Pigeons (*Columba livia*). *Anim. Cogn.* **19**, 1205–1213 (2016).

Author contributions

Y.R. and Z.S. conceived the study and analyzed the data, F.A. did the modeling and simulation. Y.R., F.A., H.J.M. and Z.S. drafted and revised the manuscript. Y.R. prepared Fig. 2.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Research Foundation (DFG) Grants SH166/3-2, awarded to ZS.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Chapter IV. Influences of stimulus modality and non-temporal magnitudes in duration judgment

Abstract

When auditory and visual signals were presented in the same experimental sessions, auditory signals were often judged to last longer than visual signals for the physically equivalent duration. Previous research has suggested that such auditory-visual difference can be incorporated within the scope of Scalar Timing Theory (STT), in which the speed of pacemaker differs between visual and auditory modalities. Thus when compared to a ‘mixture’ of standard representations from vision and audition, each duration judgment differs according to the stimulus modality. However, very few studies have addressed the role of non-temporal characteristics, such as the size of visual signals, which may affect the modality difference between vision and audition. Here we investigated the effects of stimulus modality and stimulus magnitude in duration judgment, specifically when auditory and visual signals were mixed within the same session. Our results showed the classic visual-auditory effect, when the auditory and visual durations were tested within the same session, but not in separate sessions. More interestingly, we found a second type of the memory mixing in duration judgment. That is, non-temporal magnitudes (e.g., visual size) influence both the tested duration, but also the estimation of the other durations from the other modality within the same session. We also showed the second type of the memory mixing depends on the modality precision. Our findings suggest that the memory mixing is not limited to modality, but also for the non-temporal magnitudes, which should be considered in the duration estimation.

Introduction

In daily life, we encounter abundant continuous temporal cues from surroundings that help us locate ourselves in time, and in most cases, such cues are not only from a single modality but multiple sensory domains, either simultaneously or sequentially. For example, to achieve perfect timing with a dance partner, one has to coordinate his/her movement with the rhythm of the music and his/her partner's movement. We rely on multiple sources of temporal information for reliable time judgments, but multiple sources of temporal information often do not agree with each other. One such interesting phenomenon in time judgment has been observed repeatedly since the 1960s is the audio-visual difference where "sounds are judged longer than lights" (Goldstone et al., 1978; Goldstone & Lhamon, 1974). When presented to auditory stimuli or visual stimuli with the physically equivalent duration, participants tended to judge the visual stimuli to last longer. This modality difference in duration judgment is particularly strong when visual and auditory durations are intermixed in the same testing sessions (T. B. Penney et al., 2000a; Wearden et al., 1998).

To date, the mechanism underlying the auditory/visual difference is still under debate. Many previous attempts on this issue have sought explanations within the framework of Gibbon's Scalar Expectancy Theory (SET) (Gibbon, 1977; Gibbon et al., 1984). The SET proposes a central internal-clock mechanism in which temporal information about each timed event is abstracted, encoded, and acted upon (Gibbon et al., 1984; Treisman, 1963). The central internal-clock consists of three components: a switch that controls the onset and offset of timing events, a pacemaker that emits pulses at a certain pace, and a counter that accumulates those pulses, which further would be stored in the memory as temporal units. One common explanation that accounts for the modality difference is that the internal pacemaker ticks faster for the auditory than the visual modality (Droit-Volet & Wearden, 2002; Penton-Voak et al., 1996; Wearden et al., 1999), which yields more pulses from auditory than visual durations in the accumulator, thus longer duration in the auditory than in the visual modality. An alternative explanation focused on attentional allocation, in which processing temporal information from the auditory modality is more efficient than in the visual modality (Lejeune, 1998; T. Penney, 2003).

Specifically, the attention-controlled switch in the auditory modality is less easily influenced by other attentional tasks, compared to the visual modality that hard to remain closed for the accumulation of pacemaker pulses (in other words, less ‘loss’ of the ticks for auditory duration), which results in more counts of pulses thus longer duration in the auditory than the visual modality.

While the majority of previous studies focus on modality differences within the theoretical developments of timing mechanisms, the interaction between the modality and non-temporal stimulus features (e.g., stimulus intensity) have been largely neglected. The non-temporal properties from stimulus characteristics played a vital role in influencing the duration perception. For example, visual duration with larger magnitude, such as size and luminance, as compared to the same duration with smaller magnitude, are likely to be judged longer. Similarly, auditory duration with high intensity, compared to the same duration with low intensity, tends to be judged longer (Proulx, 2010; T. H. Rammsayer & Verner, 2015; Xuan et al., 2007). In an early attempt, Goldstone and colleagues manipulated the intensity levels from visual and auditory signals in a duration comparison task and showed that the audio-visual modality difference in duration judgments can be attenuated, but not eliminated using stronger intensity of visual stimuli (Goldstone et al., 1978). Recently, Matthew and colleagues have demonstrated that it is the stimulus contrast, rather than the absolute intensities, that affects the duration judgments (Matthews et al., 2011). In their work, they used different levels of intensities from both auditory and visual stimuli as well as the background intensity and found that the stimulus magnitude highly depended on the contrast between the stimulus intensity and background intensity (e.g., weak stimuli were judged to last longer against a high-intensity background).

Given that both the non-temporal intensity and the modality have great impacts on duration judgments, it is unclear how these two factors interact with each other in a mixing context on perceived time. In this background, we aimed to investigate this by randomly mixing auditory and visual durations in the same session, and manipulated non-temporal magnitudes within each modality. We aimed to test two alternative hypotheses: The manipulation of the non-temporal magnitude only influences duration judgment in the same modality. Alternatively, the

manipulation of the non-temporal magnitude alters the general temporal context, subsequently influencing the time percept in the other modality. In addition, the temporal precision of duration estimations may vary from different modalities, the influence from one modality to the other might be different from the other way around. Specifically, we designed three experiments. In Experiment 1, we focused on the mixing effect between the auditory and visual durations with manipulation of the non-temporal intensity in the auditory modality. That is, two levels of auditory signal-to-noise (SNR) were selected for auditory presentations in two sessions. In Experiment 2, we further test the mixing effect of the auditory and visual duration judgments by manipulating the visual sizes. In Experiment 3, we conducted the same experiment as in Experiment 2, but tested the auditory and visual durations in separate sessions. Experiment 3 served as a baseline for Experiment 2. In addition, to test the generalizability of the mixing effect, we applied two different duration paradigms in the study - the duration estimation in Experiments 1 and 3, and the duration bisection task in Experiment 3.

Experiment 1.

Methods

Participants

11 participants were recruited from the subject pool of Ludwig-Maximilians University in Experiment 1. All 11 participants (8 females, mean age of 26.9 years) had normal hearing and normal or corrected-to-normal vision and gave written informed consent, which was conducted in accordance with the declaration of Helsinki 2008 and received 9 Euro/hour for their participation. The study was approved by the Ethics Board of the LMU Faculty of Pedagogics and Psychology.

Apparatus and stimuli

The experiments were performed in a sound-attenuated, dimly lit (0.76 cd/m^2) laboratory cabin. Auditory stimulus was a sinusoidal 500 Hz tone (60 dBA), which was embedded in a continuous

pink noise background (60 dB). Visual stimulus was a gray square patch (75 cd/m^2) on a black background (17.5 cd/m^2). The audio stimuli were pure 500 Hz tone, with two levels of presentations: 0 dB SNR (hereafter session ‘High’) and -10 dB SNR (hereafter session ‘Low’), with the background white noise of 60 dB. Five durations within millisecond range (200, 283, 400, 566, and 800 ms) were selected as target durations.

Design and procedure

We adopted the classic bisection paradigm in this experiment. At the beginning of each session, participants were presented with the short (S) - 100 ms and long (L) - 900 ms anchors for three times from each modality in order (visual stimuli first). Each trial commenced with a fixation cross presented on the center of the screen. After a random pause between 300 ms to 600 ms, either an auditory or a visual signal was presented and lasted for a target duration, which was randomly selected from the five sample durations. After the signal was off for a random interval between 200 ms to 500 ms, a question mark was presented in the middle of the monitor requesting a response from participants. Participants were instructed to categorize the perceived duration into either ‘short’ or ‘long’, according to the anchors presented at the beginning of each session. Responses were required using both of their index fingers. If ‘short’ was the answer, participants were asked to press the left arrow key using their left index finger, and vice versa. Once the response was given, the screen was refreshed. After a 1-second pause, the next trial began.

Each participant took part in two sessions with a 10-minute break in between. The experiment was either started with the ‘High’ or ‘Low’ session, and the order was counterbalanced among participants. Each session consisted of 7 blocks with 20 trials of each. The experiment took approximately 60 minutes to complete.

Results and discussion

In Experiment 1, auditory signals were modulated by two levels of signal-noise ratios (SNRs) of auditory signals: 0 dB and -10 dB. Accordingly, four psychometric curves (see left panel of

Figure 4) were acquired with respect to two levels of modalities ('Auditory' and 'Visual') and SNRs ('High' and 'Low'). Within each session, the psychometric curves from the auditory modality lie on the left side of the ones from the visual modality, indicating longer judgments of auditory than visual signals. Interestingly, when comparing data in session 'SNR Low' (dashed lines) to 'SNR High' (solid lines), the psychometric curves from both modalities showed a leftwards shift (i.e., PSEs become smaller), implying increasing of perceived 'Long' judgments.

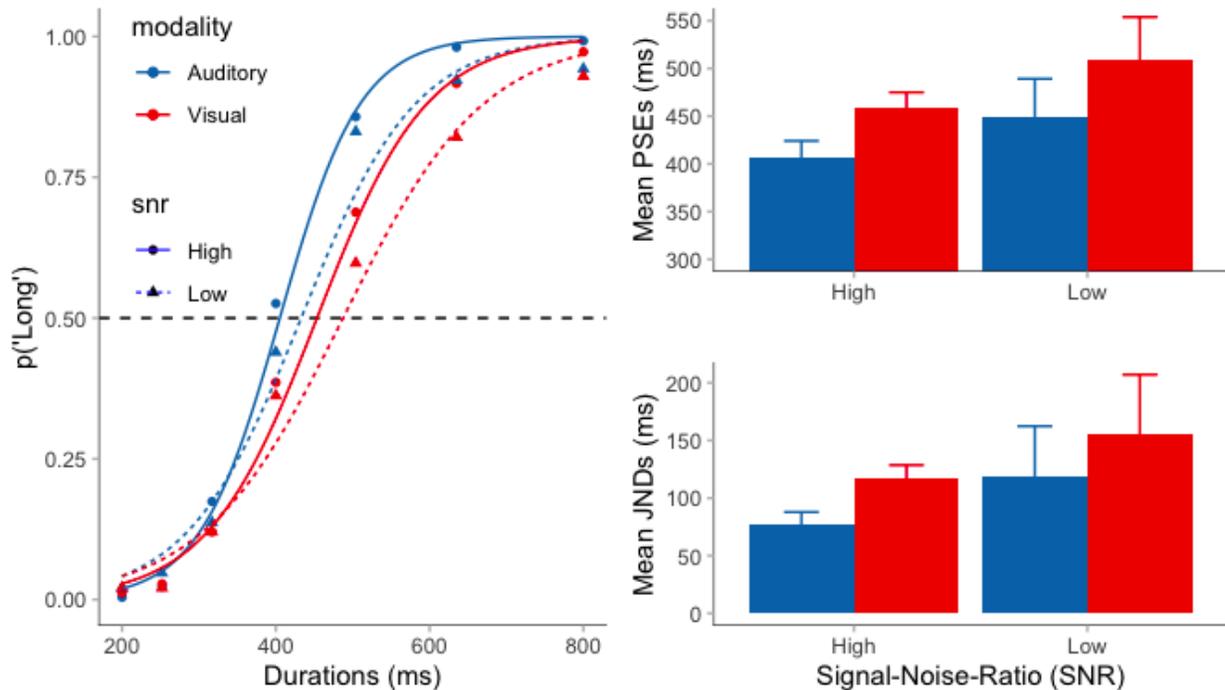


Figure 1. left panel: psychometric functions (mean proportion of 'long' responses plotted against sample duration) from four conditions in Experiment 3. Blue and red lines represent data from auditory and visual modality, respectively. Solid and dashed lines represented data from session 'High' and session 'Low', respectively; right panels: mean PSEs (upper panel) and mean JNDs (lower panel) from the psychometric functions, in four conditions.

We then extracted the *point of subjective equality (PSE)*, an index indicates the mean judgment threshold and the *just noticeable difference (JND)*, an index that indicates that sensitivity of performance (smaller value indicate high sensitivity) from the psychometric functions and

submitted both PSEs and JNDs to the ANOVAs for further analyses. First, a two-way repeated ANOVA on PSEs confirmed the modality effect, $F(1, 10) = 9.41, p = .01, \eta_g^2 = 0.07$. Again, here we replicated the classical memory mixing effect (Gu & Meck, 2011; T. B. Penney et al., 2000a), and corroborated the findings in Experiment 1. However, there was neither significant difference of session, $F(1, 10) = 0.76, p = .40, \eta_g^2 = 0.05$, nor significant interaction, $F(1, 10) = 0.01, p = .91, \eta_g^2 = 0.0005$. To achieve a better index of whether this null result between sessions reflects robust evidence or just a lack of sensitivity (i.e., too few participants), we conducted Bayesian repeated-measures ANOVA based on five different models according to the condition factors, resulting in a Bayes factor equal to 0.364, while data were grouped according to the high and low levels of SNR. Such a low Bayes factor provides some evidence to support that the results obtained in this condition reflect a genuine lack of influence of the two levels of SNR in the temporal bisection tasks. This result was different from what we found in Experiment 1. In Experiment 1, when the size of the visual stimulus was changed in different sessions, we found a general bias in the whole session. Here, however, we showed the SNR manipulation had little influence on the whole session. One possible account is that the auditory modality has a relatively high temporal resolution, which is less perturbed by the non-temporal features, such as the SNRs. In other words, the SNR manipulation here had little influence on the auditory durations, as such the high and low sessions have similar priors (given that the visual intensity did not change).

Further analysis of the temporal precision JND confirms this temporal precision account. A two-way repeated ANOVAs on JNDs showed significantly larger values in the visual modality, $F(1, 10) = 14.17, p = .004, \eta_g^2 = 0.03$, yet no significant effect of session, $F(1, 10) = 0.78, p = .40, \eta_g^2 = 0.04$, nor of Modality x Session interaction, $F(1, 10) = 0.014, p = .91, \eta_g^2 < 0.001$. The auditory modality has better temporal precision, and the SNR manipulation did not change the temporal precision.

In summary, Experiment 1 focused on the role of two levels of SNRs of auditory signals in a temporal bisection task, particularly in the testing context where both auditory and visual stimuli were experienced by participants in the same test session. The change of SNRs between sessions does not show much difference, although within each session auditory duration was perceived longer than the visual duration with the same physical length. Given that the overall loudness of auditory stimuli from both sessions remained at the same level (60 dB), different SNRs would result in the difference of absolute intensities of the target signal that ‘buried’ in the background noise (‘High’ SNR resulted in ‘louder’ target signals), which would presumably lead to different duration estimations between sessions. Yet, the finding suggests that the high temporal precision of the auditory modality is relatively robust and resistant to the perturbation of the background noise in duration estimation.

Experiment 2

Methods

Participants

14 participants (7 females, mean age of 24.9 years), were recruited from the subject pool of Ludwig-Maximilians University. All participants had normal hearing and normal or corrected-to-normal vision. Participants provided written informed consent in accordance with the Declaration of Helsinki prior to the experiment and received 9 Euro/hour for their participation. The study was approved by the Ethics Board of the LMU Faculty of Pedagogics and Psychology.

Apparatus and stimuli

The experiment was conducted in the same sound-isolated and dimly lit (0.76 cd/m^2) experimental cabin as in Experiment 1. The same equipment was used for generating auditory and visual signals. The visual stimulus was presented on a 21' CRT monitor with a screen

resolution of 800×600 pixels and a refresh rate of 90 Hz. Two sizes of visual patches were used: a small patch (subtended $3^\circ \times 3^\circ$) and a big visual patch (subtended $10^\circ \times 10^\circ$) at a viewing distance of 60 cm.

Design and Procedure

Experiment 2 aimed to investigate whether duration estimation can be influenced by the size of visual stimuli. Participants received visual and auditory durations (both ranging from 200 ms to 800 ms, but randomly interleaved across trials) in one test session and were asked to estimate the duration and marked the perceived duration onto a line-scale (see Figure 2). There were two sessions in Experiment 2: one session with the large visual patch, and one session with the small visual patch to deliver the visual duration. The loudness of the auditory stimuli remained the same for both sessions. The test durations were selected from the following five durations: 200, 283, 400, 566 and 800 ms, which were randomly mixed across trials.

At the beginning of each block, one short (100 ms) and one long (900 ms) intervals from both modalities, were presented twice to remind participants the short and long anchoring durations. They were told that the short was 100 ms and the long was 900 ms. The formal testing trial started with a fixation cross presented on the center of the screen. After a random interval of 300-600 ms, the fixation disappeared and either an auditory or a visual stimulus was presented for a given duration. After the offset of the target duration for a random 200-500 ms, a ruler with marks of 0, 500 ms, 1000 ms, and 1500 ms was shown on the lower half of the screen, on which a cursor can be moved by the use of the mouse (see Figure 2). The initial position of the cursor was randomly located between 0 and 1500 ms. Participants were asked to move the cursor to a location that represented the perceived duration and to click to confirm their response. No feedback was given after each response. After the response the display was cleared for one second, and the next trial began.

The order of two sessions were counterbalanced among participants. Each session consisted of 9 blocks with each of 40 trials of each. The experiment lasted approximately 90 minutes to complete and participants were asked to take a 15-minute break between sessions.

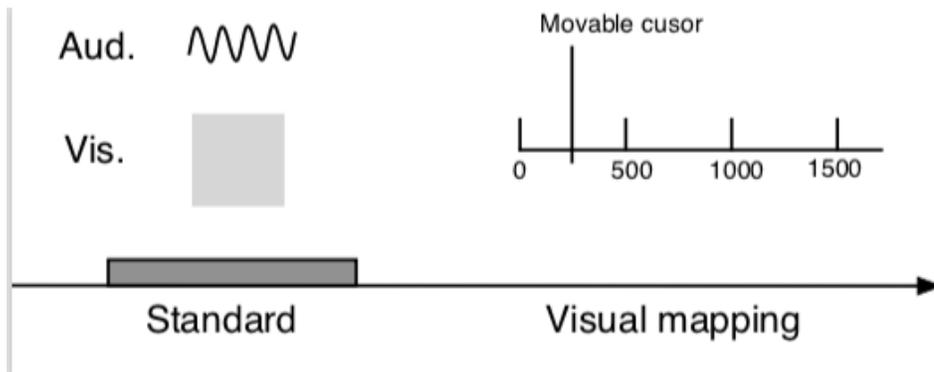


Figure 2. Schematic illustration of the experimental paradigm of duration estimation in Experiment 1. A given duration (ranging between 200 to 800 ms) was delivered via auditory or visual stimulus. Participants were asked to judge the perceived duration and to map the duration to a line-scale using the movable cursor. The initial cursor was located at a random position between 0 ms to 1500 ms.

Results and discussion

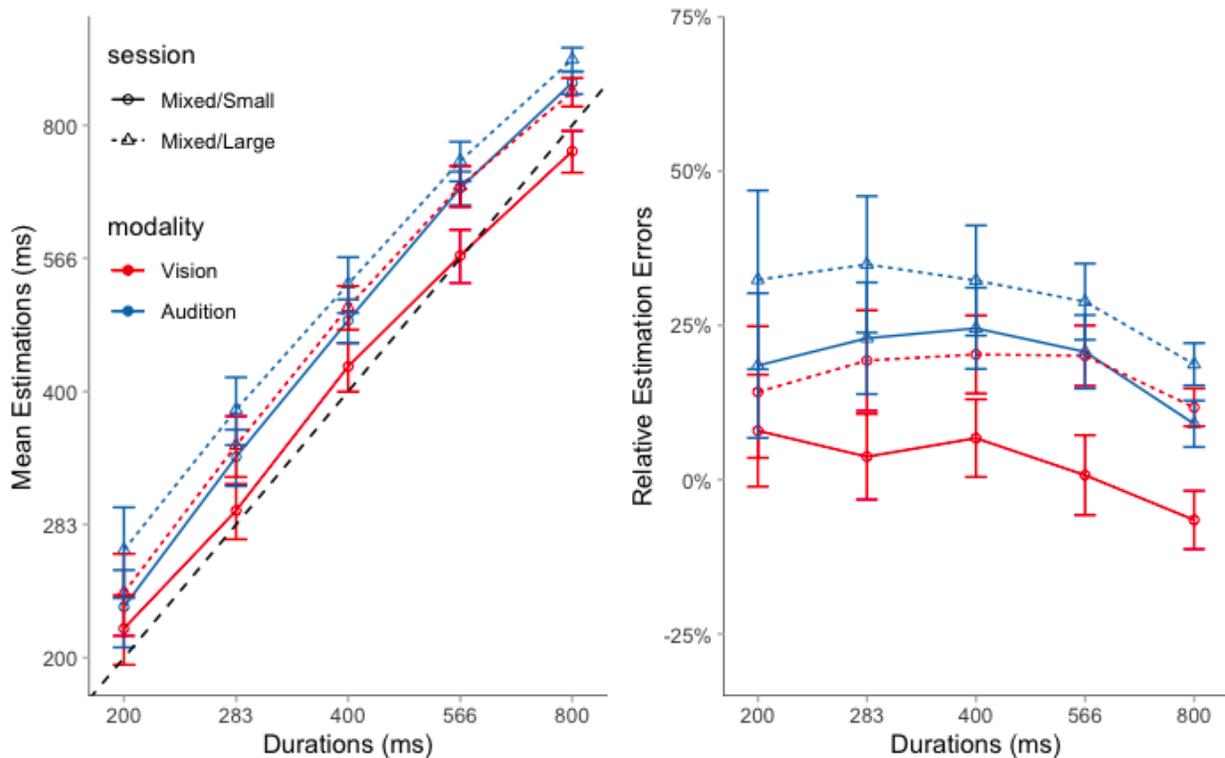


Figure 3. Mean estimations (left panel) and relative estimation errors (right panel) as a function of the test duration, separated for auditory and visual modality and two sessions in Experiment 1. Blue and red lines represent data from auditory and visual modality, respectively. Solid and dotted lines are data from session 1 (session 'Small') and session 2 (session 'Large'), respectively. The error bar represents one standard error.

Figure 3 shows the results of Experiment 2. By visual inspection, the mean estimation follows the test duration in linear fashion, with auditory duration being overestimated in general as compared to the visual duration, which is consistent with previous findings on the auditory/visual differences of duration judgments (T. B. Penney et al., 2000a; Wearden et al., 2006). Second, the visual durations presented with small patches were perceived shorter than the same duration with the large patch. To quantify the estimation errors, we calculated the Relative Estimation Error (RRE), which was defined by the proportion of the estimation error (the estimated duration subtracted the test duration) to the test duration. As shown in the right panel of Figure 3, this metric provided a better illustration as to how response deviated from each sample point. A three-way ANOVA with the factor of modality, session, and duration revealed a significant of the main effect of the modality, $F(1,13) = 4.77$, $p=.048$, $\eta_g = .09$, indicating the auditory durations were generally reproduced longer than the visual durations with the same physical length. There was also a significant effect of session, $F(1,13) = 5.33$, $p=.038$, $\eta_g = .11$, suggesting that the mixing with the large visual stimuli shifts the reproduced durations higher than the session with the small visual stimuli. But there was no significant difference among different durations, $F(1,13) = 1.74$, $p=.21$, $\eta_g = .03$, indicating that the estimation errors follow the scalar property.

Thus, in Experiment 2, we replicated the classical auditory-visual effect found in previous studies: when the auditory and visual durations were presented in the same testing block, auditory durations were perceived longer than the visual durations at the equivalent sample duration. Interestingly, we also found that the non-temporal feature - size of the visual stimuli also influenced the duration judgment, not only for the visual duration, but also for the auditory duration. In the large visual stimulus session, the estimation of the auditory durations was also

shifted upward. This suggests that the non-temporal feature influenced the general prior of the perceived duration range (Shi et al., 2013), having an amodal effect on time judgment.

Experiment 3

In Experiment 3, the non-temporal features, visual sizes, were systematically manipulated in two separate sessions. Given that the auditory and visual modalities have different temporal resolutions, here we further test the mixing effect with manipulation in the auditory modality. Experiment 1 we found the classic modality mixing effect. In order to further examine if the effect is truly due to the context mixing, we conducted Experiment 3, in which we tested the visual and auditory duration judgments in separate sessions.

Methods

Participants

16 participants took part in Experiment ‘Separate’ (8 females, mean age of 22.1). All had normal hearing and normal or corrected-to-normal vision and gave written informed consent, which was conducted in accordance with the declaration of Helsinki 2008 and received 9 Euro/hour for their participation. The study was approved by the Ethics Board of the LMU Faculty of Pedagogics and Psychology.

Apparatus and stimuli

The experiment was conducted in the same sound-isolated and dimly lit (0.76 cd/m^2) experimental cabin as in Experiment 1. The same equipment was used for generating auditory and visual signals. The size of visual stimuli was the same as in the ‘Small’ session of Experiment 1 (subtended $3^\circ \times 3^\circ$). Auditory stimulus was a sinusoidal 500 Hz tone (60 dBA), which was embedded in a continuous pink noise background (60 dB).

Design and procedure

The identical paradigm and experimental task were adopted in Experiment ‘Separate’, except that in each session participants only experienced stimuli from one modality, instead of both

auditory and visual stimuli intermixed for presentations. There were two separate modality-specific sessions: one session for vision and one session for audition. The order of two sessions were counterbalanced among participants. Each session consisted of 5 blocks with each of 30 trials of each. The experiment lasted approximately 45 minutes to complete and participants were asked to take a 10-minute break between sessions.

Results and discussion

Figure 3 depicts the relative duration estimation errors in Experiment 2. By visual inspection, there was a numerical trend that the estimation error decreased as the function of the duration, and the errors were numerically larger in vision than in audition. However, a repeated-measures ANOVA with the modality and the duration as two main factors revealed neither the main effects (Duration: $F(1,15)=1.96$, $p=.18$, Modality: $F(1,15)=1.89$, $p=.19$), nor their interaction ($F(1,15)=1.72$, $p=.2$) were significant. That is, the relative estimation errors were not significantly different between auditory and visual modalities, and between the short and long durations.

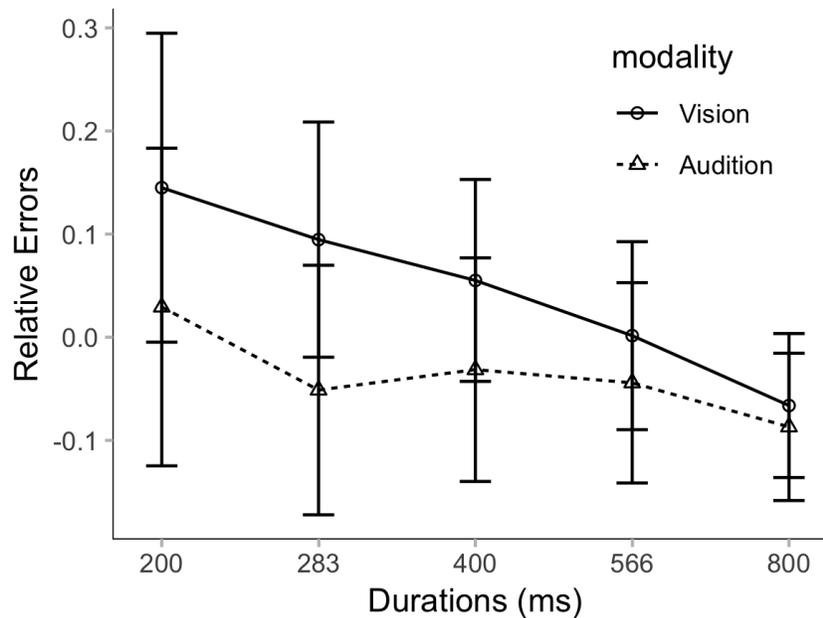


Figure 4. Mean relative estimation error as a function of the test duration in Experiment 3, separated for the visual and auditory durations. The error bar represents one standard error.

In Experiment 1, when the auditory and visual durations were randomly mixed within blocks, we found the classical memory mixing effect (T. B. Penney et al., 2000a; Wearden et al., 2006). However, in Experiment 3, when we separate two modalities in separate sessions, the mixing effect diminished. Combining both experiments, we are relatively confident that the modality mixing effect comes from the cross-modality context. That is, when both auditory and visual durations are available for comparison, this modality context leads to the mixing effect.

General Discussion

The present study investigated the interaction between the modality mixing and non-temporal intensities (such as size and SNR) in duration estimation. We mixed auditory and visual durations within each session in Experiments 1 and 2, and compared to the baseline condition where the auditory and visual durations were separated tested in different sessions. We replicated the modality mixing effect both in Experiments 1 and 2, with auditory duration judged longer than visual ones with the same physical length within the same test session. However, this modality mixing effect was not observed when the test was separated in separate sessions (Experiment 3). More interestingly, we found that the manipulation of the non-temporal feature, such as visual size, not only influenced the visual duration judgments, but also the duration estimation of the auditory modality in the same session. The large visual size shifted the overall estimation bias upward, as compared to the small visual size session. By contrast, the manipulation of the signal-noise ratio (SNR) in the auditory modality failed to change the overall bias. These results demonstrated an asymmetric role of non-temporal characteristics between visual and auditory modality.

Modality mixing effect

As outlined in the Introduction, the modality mixing effect has been observed when the auditory and visual durations were randomly mixed together, and auditory duration is often perceived longer than the visual duration (T. Penney, 2003; T. B. Penney et al., 2000a). Penney and colleagues suggested that the occurrence of such modality difference occurred in the memory

stage where the auditory and visual duration representations were stored in the same working memory, which makes the comparison across modality possible. As it has been shown in the literature that the auditory duration is often perceived longer than the visual duration (Wearden et al., 1998, 2006), thus the memory contrast causes the whole auditory duration to be overestimated while the visual durations are relatively underestimated. However, whether the modality difference is required for memory mixing has been questioned by Wearden and colleagues (Wearden et al., 2006). Wearden and colleagues argued that the modality difference is inherited from modality-specific ‘pacemaker speed’, ‘mixing’ auditory and visual stimuli for duration presentation is not necessarily required. In the present study, we replicated the modality mixing effect when the auditory and visual durations were randomly interleaved together within one session. However, when the auditory and visual durations were tested separated (Experiment 2), the modality difference became insignificant. Thus, our findings favor the memory mixing account (T. B. Penney et al., 2000b).

It should be noted, though, our findings do not reject the difference of the time processing in the visual and auditory modalities. In fact, we found the temporal precision in duration judgments were different for the auditory and visual modalities. The temporal precision of the auditory modality was higher than the visual modality, which is consistent with previous findings stating that duration judgment is more precise in the auditory domain (Thomas H. Rammsayer et al., 2015a). This might be an indication of an inherent difference between auditory and (Thomas H. Rammsayer et al., 2015b)visual in subsecond timing. For example, higher auditory temporal sensitivity may be due to faster and more accurate processing of auditory information, considering the pathway to process auditory information is shorter than in vision (Pinel et al., 2004). Electrophysiological evidence also supported this idea with investigating the N1 component, an early negative component of the event-related potential (ERP), which reflects the analysis of physical stimulus properties, identified differences in latencies of N1 components elicited by visual and auditory stimuli (Allison et al., 1984). Thus, in spite of the adopted experimental paradigm and selection of stimulus configuration, temporal judgments from audition always exhibit higher precision.

Influences of non-temporal magnitudes

Non-temporal magnitudes (such as intensities, contrast) have been shown to be influential factors in duration judgments. It has been suggested that high intensity may be more aroused than the low intensity stimuli, thus speeding up the internal pacemaker speed, according to the information-processing pacemaker-accumulator models (Gibbon & Church, 1990; Penton-Voak et al., 1996). This has been supported by studies that found the magnitude effect in sub-second duration estimations (Busch et al., 2004; Casini & Macar, 1997). By adopting different levels of nontemporal magnitudes in duration representations, such as the number of dots, size and luminance of visual stimuli, Xuan and colleagues (Xuan et al., 2007) demonstrated that visual items with larger magnitudes were judged to last longer. By using a size contrast illusion (the Ebbinghaus illusion), Ona and Kawahara also showed that perceived duration also depended on the subjective visual size (Ono & Kawahara, 2007).

However, it is important to mention the influence of non-temporal features/intensities is only concerned with the duration of that related stimulus intensity, which is often assumed to be independent of the other stimuli. That is, the theory does not consider the ‘magnitude’ context. Here, we demonstrated that the intensity changes in visual duration (i.e., the size of the visual stimulus) not only altered the visual duration judgment, but also the auditory duration that was being tested in the same session. Our findings suggest that the influence of the magnitude in duration judgment is not short-live for the test duration itself, it has a broader influence in terms of the context. It is a second type of memory mixing. That is, the non-temporal magnitude contributes to the bias of the current estimation, which the estimation is incorporated into the modality non-specific amodal prior. As a result, the global general prior influences the duration estimation in the other modality.

Interestingly, though, the manipulation of the intensity in auditory modality in terms of the SNR (0 dB vs. -10 dB) did not cause a general shift between two high vs. low SNR sessions (Experiment 3), which is in contrast to the visual size manipulation. From the temporal estimation, however, we do find that the temporal precision in the auditory modality was higher than that in the visual modality, and the temporal precision of the auditory modality in the high

and low SNR sessions did not differ from each other. This suggests that the failure modulation of the SNR may be owing to the fact that the temporal precision of the auditory modality is relatively high, which is resistant to any different SNR manipulations. Alternatively, the SNR manipulation here (the difference was only 10 dB) was not large enough to introduce a general bias, given that the auditory has a high temporal precision.

Conclusion

In the current study, we found two types of memory mixing in duration judgment - modality mixing and non-temporal magnitude mixing. We replicated the modality mixing effect when the auditory and visual durations were tested in a randomly mixed session, while the modality effect diminished when each modality was tested separately. In addition, we found the non-temporal magnitude not only influenced the test duration itself, but could also introduce a general bias in the test session. However, the influence of non-temporal magnitude depends on the modality. The influence is more marked when the manipulated modality has less temporal precision.

Author Contributions

This work was carried out under the supervision of Zhuanghua Shi; Z.S. conceived the idea, Y.R. and Z.S. designed the research, Y.R. collected and analyzed the data. Y.R. and Z.S. discussed the results and wrote the paper. H.J.M commented on the manuscript.

Acknowledgement

This work was supported by the German Research Foundation (DFG) grants SH166/3-2, awarded to ZS. Correspondence concerning this article should be addressed to Zhuanghua Shi, General and Experimental Psychology, Department of Psychology, LMU Munich, 80802, Munich, Germany. Email: shi@lmu.de

References for Chapter IV

- Allison, T., Hume, A. L., Wood, C. C., & Goff, W. R. (1984). Developmental and aging changes in somatosensory, auditory and visual evoked potentials. In *Electroencephalography and Clinical Neurophysiology* (Vol. 58, Issue 1, pp. 14–24).
[https://doi.org/10.1016/0013-4694\(84\)90196-2](https://doi.org/10.1016/0013-4694(84)90196-2)
- Busch, N. A., Debener, S., Kranczioch, C., Engel, A. K., & Herrmann, C. S. (2004). Size matters: effects of stimulus size, duration and eccentricity on the visual gamma-band response. In *Clinical Neurophysiology* (Vol. 115, Issue 8, pp. 1810–1820).
<https://doi.org/10.1016/j.clinph.2004.03.015>
- Casini, L., & Macar, F. (1997). Effects of attention manipulation on judgments of duration and of intensity in the visual modality. *Memory & Cognition*, 25(6), 812–818.
- Droit-Volet, S., & Wearden, J. (2002). Speeding up an internal clock in children? Effects of visual flicker on subjective duration. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, 55(3), 193–211.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. In *Psychological Review* (Vol. 84, Issue 3, pp. 279–325). <https://doi.org/10.1037/0033-295x.84.3.279>
- Gibbon, J., & Church, R. M. (1990). Representation of time. In *Cognition* (Vol. 37, Issues 1-2, pp. 23–54). [https://doi.org/10.1016/0010-0277\(90\)90017-e](https://doi.org/10.1016/0010-0277(90)90017-e)
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar Timing in Memory. In *Annals of the New York Academy of Sciences* (Vol. 423, Issue 1 Timing and Ti, pp. 52–77).
<https://doi.org/10.1111/j.1749-6632.1984.tb23417.x>
- Goldstone, S., & Lhamon, W. T. (1974). Studies of Auditory-Visual Differences in Human Time

- Judgment: 1. Sounds are Judged Longer than Lights. In *Perceptual and Motor Skills* (Vol. 39, Issue 1, pp. 63–82). <https://doi.org/10.2466/pms.1974.39.1.63>
- Goldstone, S., Lhamon, W. T., & Sechzer, J. (1978). Light intensity and judged duration. In *Bulletin of the Psychonomic Society* (Vol. 12, Issue 1, pp. 83–84). <https://doi.org/10.3758/bf03329633>
- Gu, B. M., & Meck, W. H. (2011). New perspectives on Vierordt's law: memory-mixing in ordinal temporal comparison tasks. In A. Vatakis, A. Esposito, M. Giagkou, F. Cummins, & G. Papadelis (Eds.), *Multidisciplinary Aspects of Time and Time Perception* (Vol. 6789, pp. 67–78).
- Lejeune, H. (1998). Switching or gating? The attentional challenge in cognitive models of psychological time. In *Behavioural Processes* (Vol. 44, Issue 2, pp. 127–145). [https://doi.org/10.1016/s0376-6357\(98\)00045-x](https://doi.org/10.1016/s0376-6357(98)00045-x)
- Matthews, W. J., Stewart, N., & Wearden, J. H. (2011). Stimulus intensity and the perception of duration. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 37, Issue 1, pp. 303–313). <https://doi.org/10.1037/a0019961>
- Ono, F., & Kawahara, J.-I. (2007). The subjective size of visual stimuli affects the perceived duration of their presentation. In *Perception & Psychophysics* (Vol. 69, Issue 6, pp. 952–957). <https://doi.org/10.3758/bf03193932>
- Penney, T. (2003). Modality Differences in Interval Timing. In *Functional and Neural Mechanisms of Interval Timing*. <https://doi.org/10.1201/9780203009574.ch8>
- Penney, T. B., Gibbon, J., & Meck, W. H. (2000a). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human*

Perception and Performance, 26(6), 1770–1787.

Penney, T. B., Gibbon, J., & Meck, W. H. (2000b). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology. Human Perception and Performance*, 26(6), 1770–1787.

Penton-Voak, I. S., Edwards, H., Percival, A., & Wearden, J. H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *Journal of Experimental Psychology. Animal Behavior Processes*, 22(3), 307–320.

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, 41(6), 983–993.

Proulx, M. J. (2010). Size Matters: Large Objects Capture Attention in Visual Search. In *PLoS ONE* (Vol. 5, Issue 12, p. e15293). <https://doi.org/10.1371/journal.pone.0015293>

Rammsayer, T. H., Borner, N., & Troche, S. J. (2015a). Visual-auditory differences in duration discrimination of intervals in the subsecond and second range. *Frontiers in Psychology*, 6, 1626.

Rammsayer, T. H., Borner, N., & Troche, S. J. (2015b). Visual-auditory differences in duration discrimination of intervals in the subsecond and second range. *Frontiers in Psychology*, 6, 1626.

Rammsayer, T. H., & Verner, M. (2015). Larger visual stimuli are perceived to last longer from time to time: The internal clock is not affected by nontemporal visual stimulus size. In *Journal of Vision* (Vol. 15, Issue 3, pp. 5–5). <https://doi.org/10.1167/15.3.5>

Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends*

in Cognitive Sciences, 17(11), 556–564.

Treisman, M. (1963). Temporal discrimination and the indifference interval. Implications for a model of the “internal clock.” *Psychological Monographs*, 77(13), 1–31.

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why “sounds are judged longer than lights”: application of a model of the internal clock in humans. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, 51(2), 97–120.

Wearden, J. H., Philpott, K., & Win, T. (1999). Speeding up and (...relatively...) slowing down an internal clock in humans. *Behavioural Processes*, 46(1), 63–73.

Wearden, J. H., Todd, N. P. M., & Jones, L. A. (2006). When do auditory/visual differences in duration judgements occur? In *Quarterly Journal of Experimental Psychology* (Vol. 59, Issue 10, pp. 1709–1724). <https://doi.org/10.1080/17470210500314729>

Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision*, 7(10), 2.1–5.

Chapter V. General Discussion

The main research objective of this accumulative work was to examine the roles of three contextual factors in subjective interval timing in the range of sub-seconds to seconds. First, we focused on the influence of temporal context (e.g., the range of sample durations) in duration reproductions and examined whether the use of different temporal contexts could affect general timing performance, such as the accuracy and (relative) precision of interval reproductions. The second study addressed the question of whether internal subjective intervals are logarithmic encoded. We used the rapid and intuitive ensemble perception in extracting statistical information (such as the mean) from the ensemble temporal context as a probing tool to investigate this. In a third study, we concentrated on the interplay of non-temporal and temporal context in time judgments. Specifically, we manipulated the ‘mixed’ condition where both auditory and visual stimuli were presented in the same testing sessions with different configurations of non-temporal contexts.

Summary of findings

The first two studies (Chapters 2 and 3) focused on two perspectives of the classic laws of psychophysics in human interval timing: 1) the psychophysical law, which describes the relationship between the magnitude of sample duration and perceived time, and 2) scalar timing, which states a constant relationship between JND and mean estimations of time (S. Grondin, 2003). We demonstrated that temporal context (e.g., sample distributions) played a vital role in both aspects of the laws of psychophysics in interval timing. Through manipulating the temporal context among testing sessions, we compared the duration reproductions of the same target samples from different temporal contexts (see Chapter 2) and observed both conformity and violation of scalar timing (see Figures 4 and 5 in Chapter 2), as mean coefficient of variations (CVs) showed both constant and varied trends against targets intervals according to the experimental contexts (e.g., the temporal mixed vs. blocked contexts, or visual vs. auditory contexts). We further explained how the variations of CVs could happen on each stage of the reproduction-reproduction task using the Bayesian model. Although it is still under debate whether human interval timing conforms to the scalar timing, our findings suggested that the observed CV is not strictly tied to the internal scalar property

and using the CV alone to reject the scalar property is not warranted by the behavioral findings.

Focusing on the other fundamental law of psychophysics in time, the second study (Chapter 3) aimed to investigate the internal encoding of the subjective timeline by adopting an intuitive process - ensemble perception - in the time domain. The idea was, since ensemble perception in time work at a very fast, low level of coding and might possibly bypass many high-level cognitive decoding strategies, the task of averaging of time intervals, which requires rapid response to extract the mean from a train of consecutively presented durations, may thus reflect an average of primitive encoded time. By manipulating the mean values of three different sample distributions, those different averaging strategies could be disassociated: while the arithmetic mean could reflect a linear encoding, the geometric mean could reflect a logarithmic encoding. The results provided evidence to support that our perceived (subjective) time, at least in the subsecond range, is likely to be logarithmically encoded.

The third study (Chapter 4) investigated how non-temporal characteristics of stimuli could affect sub-second timing when timing information was delivered by mixed auditory and visual signals. We investigated the role of two non-temporal factors, one for vision and one for audition, which could potentially influence the classic visual/auditory difference in sub-second interval timing. We mixed auditory and visual durations within each session and compared to the baseline condition where the auditory and visual durations were separately tested in different sessions. We replicated the modality mixing effect with auditory duration judged longer than visual ones with the same physical length within the same test session. However, this modality mixing effect was not observed when the test was separated into separate sessions. More interestingly, we found that the manipulation of the non-temporal features, such as visual size, not only influenced the visual duration judgments, but also the duration estimation of the auditory modality in the same session. The large visual size shifted the overall estimation bias upward, as compared to the small visual size session. By contrast, the manipulation of the signal-noise ratio (SNR) in the auditory modality failed to change the overall bias. These results demonstrated an asymmetric role of non-temporal characteristics between visual and auditory modality.

Temporal Context and Bayesian Inference Model

As outlined in the Introduction, up to date the constancy of CVs in time domain still widely functions as the ‘golden’ standard of examining the scalar timing (Buhusi et al., 2018; Wearden & Lejeune, 2008). While the constant CVs conforms to Weber's law, a ubiquitous law for magnitude perception, reporting the inconstancy of the scalar property often led to the conclusion of a systematic violation of such law (Simon Grondin, 2012, 2014). By observing both constant and inconstant CVs on the same sample durations but from different experimental contexts (e.g., different temporal ranges and modalities), we argued that the role of context factors should be more granted. Additionally, the two-stage Bayesian estimator model proposed in this thesis (see Chapter 2) successfully predicted both the mean and sensitivities - two key parameters that evaluate the general goodness of timing performances - in temporal reproductions.

Previous studies using the Bayesian inference estimator have enjoyed a great amount of success in predicting interval timing. For example, Jazayeri and Shadlen's BLS model successfully predicted the central tendency effect of duration reproductions (Jazayeri & Shadlen, 2010). Their model adopted a uniform distribution as the prior, which assumed that each sample duration contributed equally to the prior knowledge, whereas Cicchini's model (Cicchini et al., 2012) predicted a better fit on the central tendency by assuming a normal distribution depends on both the mean and the standard deviation of sample distributions. In our model, we considered a third parameter - a shift term (Δ), which incorporates additional information that cannot be directly reflected by simple statistical properties from sample durations. Our results demonstrated that this mean shift of the prior proposed in our model could capture the change in the center of the sample durations, in spite of the modality difference between stimulus presentations (see Table 1 in Chapter 2). To assimilate the whole process of the production-reproduction task, we introduced another additional parameter in the ‘reproduction’ phase of the task. Compared to the classic bisection task of intervals, the production-reproduction task may require high-level cognitive skills, as the reference memory is updated in each trial for explicit time judgment (Block et al., 1999; S. Droit-Volet et al., 2015). We considered a variance component of decision uncertainty (σ_r) using a

zero-mean Gaussian distribution and demonstrated this stimulus-independent variance had a large impact on the variations of CVs (also see Table 1 in Chapter 2).

Modality difference between Vision and Audition

Interestingly, we observed a strong modality difference in interval timing between vision and audition in all three studies of this thesis. In study one, duration reproductions from auditory modality showed less estimation bias and less central tendency towards the sample mean (see Figure 2 in Chapter 2). In addition, CV under auditory modality showed overall smaller values indicating higher precision of reproductions (see Figure 5 in Chapter 2). In the second study, estimations in the auditory modality showed both better estimations of PSE and JND of the psychometric function in the mean discrimination task (see Figures 3 and 4 in Chapter 3). In the third study, when auditory and visual signals were mixed in the same testing session, auditory stimuli were always judged longer than visual stimuli (see Figure 2 in Chapter 4), which is consistent with the classic audio-visual differences where ‘sounds are judged longer than lights (Goldstone & Lhamon, 1974; Goldstone et al., 1978)’. In a bisection task, JNDs of the psychometric function were smaller in the auditory than visual domain indicating a higher sensitivity of time judgments (see Figure 4 in Chapter 4).

These findings provide converging evidence for employment of modality-specific timing mechanisms, in which the modality difference was mainly due to different pacemaker rate in both ‘clock’ and ‘memory’ stage of the interval clock (Sylvie Droit-Volet et al., 2007; Ogden et al., 2010), or achieved by climbing neural activation within modality-specific brain areas (Wittmann, 2013). Also, our findings support the idea of an inherent difference where temporal information tended to be processed faster and more efficiently in the auditory than in the visual domain (Allman et al., 2014; Buetti et al., 2008; Zélandi & Droit-Volet, 2012).

Implications of logarithmic encoding of Time

It was suggested that space, time and number share similar 'nature' in the sense that providing us ‘a priori intuitions’ to experience our environment (Stanislas Dehaene & Brannon, 2010). Studies from numerosity have demonstrated a logarithmic ‘intuition’ of perception, only that culture-specific or attentional-related experiences ‘linearized’ the form of mapping to the

observable level (Anobile et al., 2012; Stanislas Dehaene, 2003; Feigenson et al., 2004). For example, when probed number-space mappings in the Mundurucu, an Amazonian indigene group with a reduced numerical lexicon and little or no formal education, mapped symbolic and non-symbolic numbers onto a logarithmic scale, whereas Western adults used linear mapping with small or symbolic numbers and logarithmic mapping when numbers were presented non-symbolically under conditions that discouraged counting (S. Dehaene et al., 2008). When performing a dual-task, educated adults showed logarithmic scaling for numbers, whereas a linear scaling was shown when performing a single-task (Anobile et al., 2012). The discovery of number neurons from Nieder and Miller might be a 'game-changer' in solving the psychophysical debate: the neural representation of number lines seems to favor the logarithmic scale rather than linear (Nieder & Miller, 2003). Recently, Roberts and colleagues have demonstrated that both subjective timeline and number line were logarithmic in animal timing (e.g., in pigeons) (Roberts, 2005, 2006). Through using the ensemble coding in the time domain (see Chapter III), we provided further evidence that the subjective timeline in the sub-second range is also likely to be logarithmic and joined this line of research demonstrating a potentially universal representation of perceptual magnitude in both number and time (Varshney & Sun, 2014; Walsh, 2003).

Conclusions

The thesis aimed to uncover the role of three context factors in human intervals timing in the range of sub-seconds to seconds. We argue for more careful examinations upon both temporal and non-temporal factors that emerged from the experimental environment, which may influence observed timing behaviors in psychophysical research. On the one hand, because psychophysical studies of timing can provide a solid behavioral foundation in forming timing models, variations of observed timing behaviors may result in different theoretical developments, and furthermore, lead to potentially false investigations of the neurobiological mechanisms. Thus, fully addressing the role of context factors could decrease the variations of those interpretations of timing theories. On the other hand, although the validity of behavioral studies has been increasingly acknowledged in recent years, lacking direct information about the neural basis makes it difficult to substantiate the internal representation of time. To achieve a full understanding of human timing requires further concerted research effort from both the psychophysical and neural perspectives.

References

- Aagten-Murphy, D., Cappagli, G., & Burr, D. (2014). Musical training generalises across modalities and reveals efficient and adaptive mechanisms for reproducing temporal intervals. *Acta Psychologica*, *147*, 25–33.
- Allan, L. G. (1998). The influence of the scalar timing model on human timing research. In *Behavioural Processes* (Vol. 44, Issue 2, pp. 101–117).
[https://doi.org/10.1016/s0376-6357\(98\)00043-6](https://doi.org/10.1016/s0376-6357(98)00043-6)
- Allan, L. G. (1998). The influence of the scalar timing model on human timing research. *Behavioural Processes*, *44*(2), 101–117.
- Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. In *Learning and Motivation* (Vol. 22, Issues 1-2, pp. 39–58).
[https://doi.org/10.1016/0023-9690\(91\)90016-2](https://doi.org/10.1016/0023-9690(91)90016-2)
- Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learning and Motivation*, *22*(1-2), 39–58.
- Allman, M. J., Penney, T. B., & Meck, W. H. (2016). A Brief History of “The Psychology of Time Perception.” In *Timing & Time Perception* (Vol. 4, Issue 3, pp. 299–314).
<https://doi.org/10.1163/22134468-00002071>
- Allman, M. J., Teki, S., Griffiths, T. D., & Meck, W. H. (2014). Properties of the internal clock: first- and second-order principles of subjective time. *Annual Review of Psychology*, *65*, 743–771.
- Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*(3), 454–459.

- Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. In *Psychological Science* (Vol. 12, Issue 2, pp. 157–162). <https://doi.org/10.1111/1467-9280.00327>
- Baird J. C., & Noma E. J. (1978). *Fundamentals of scaling and psychophysics*. John Wiley & Sons.
- Bausenhart, K. M., Dyjas, O., & Ulrich, R. (2014). Temporal reproductions are influenced by an internal reference: Explaining the Vierordt effect. In *Acta Psychologica* (Vol. 147, pp. 60–67). <https://doi.org/10.1016/j.actpsy.2013.06.011>
- Bizo, L. A., Chu, J. Y. M., Sanabria, F., & Killeen, P. R. (2006). The failure of Weber’s law in time perception and production. In *Behavioural Processes* (Vol. 71, Issues 2-3, pp. 201–210). <https://doi.org/10.1016/j.beproc.2005.11.006>
- Block, R. A., Zakay, D., & Hancock, P. A. (1999). Developmental Changes in Human Duration Judgments: A Meta-Analytic Review. In *Developmental Review* (Vol. 19, Issue 1, pp. 183–211). <https://doi.org/10.1006/drev.1998.0475>
- Brown, S. W. (1995). Time, change, and motion: The effects of stimulus movement on temporal perception. In *Perception & Psychophysics* (Vol. 57, Issue 1, pp. 105–116). <https://doi.org/10.3758/bf03211853>
- Bueti, D., Bahrami, B., & Walsh, V. (2008). Sensory and association cortex in time perception. *Journal of Cognitive Neuroscience*, 20(6), 1054–1062.
- Buhusi, C. V., Oprisan, S. A., & Buhusi, M. (2018). Biological and Cognitive Frameworks for a Mental Timeline. *Frontiers in Neuroscience*, 12, 377.
- Callender, C. (2019). Time, flow, and space [Review of *Time, flow, and space*]. *The Behavioral and Brain Sciences*, 42, e246.

- Chong, S. C., & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900.
- Church, R. (2003). A Concise Introduction to Scalar Timing Theory. In *Functional and Neural Mechanisms of Interval Timing*. <https://doi.org/10.1201/9780203009574.sec1>
- Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology. Animal Behavior Processes*, *3*(3), 216–228.
- Church, R. M., Meck, W. H., & Gibbon, J. (1994). Application of scalar timing theory to individual trials. *Journal of Experimental Psychology. Animal Behavior Processes*, *20*(2), 135–155.
- Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal Encoding of Interval Timing in Expert Percussionists. In *Journal of Neuroscience* (Vol. 32, Issue 3, pp. 1056–1060). <https://doi.org/10.1523/jneurosci.3411-11.2012>
- Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(3), 1056–1060.
- Curtis, D. W., & Mullin, L. C. (1975). Judgments of average magnitude: Analyses in terms of the functional measurement and two-stage models. In *Perception & Psychophysics* (Vol. 18, Issue 4, pp. 299–308). <https://doi.org/10.3758/bf03199378>
- de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, *62*(9), 1716–1722.
- Dehaene, S. (2001). Subtracting pigeons: logarithmic or linear? [Review of *Subtracting pigeons: logarithmic or linear?*]. *Psychological Science*, *12*(3), 244–246; discussion

247.

- Dehaene, S. (2003). The neural basis of the Weber–Fechner law: a logarithmic mental number line. In *Trends in Cognitive Sciences* (Vol. 7, Issue 4, pp. 145–147).
[https://doi.org/10.1016/s1364-6613\(03\)00055-x](https://doi.org/10.1016/s1364-6613(03)00055-x)
- Dehaene, S., & Brannon, E. M. (2010). Space, time, and number: a Kantian research program. *Trends in Cognitive Sciences*, *14*(12), 517–519.
- Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures. In *Science* (Vol. 320, Issue 5880, pp. 1217–1220). <https://doi.org/10.1126/science.1156540>
- Droit-Volet, S., Meck, W. H., & Penney, T. B. (2007). Sensory modality and time perception in children and adults. *Behavioural Processes*, *74*(2), 244–250.
- Droit-Volet, S., Wearden, J. H., & Zélanti, P. S. (2015). Cognitive abilities required in time judgment depending on the temporal tasks used: A comparison of children and adults. *Quarterly Journal of Experimental Psychology*, *68*(11), 2216–2242.
- Eisler, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. In *Psychological Bulletin* (Vol. 83, Issue 6, pp. 1154–1171).
<https://doi.org/10.1037/0033-2909.83.6.1154>
- Fechner, G. T. (1860). *Elemente der Psychophysik*.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. In *Trends in Cognitive Sciences* (Vol. 8, Issue 7, pp. 307–314).
<https://doi.org/10.1016/j.tics.2004.05.002>
- Fraisse, P. (1984). Perception and Estimation of Time. In *Annual Review of Psychology* (Vol.

35, Issue 1, pp. 1–37). <https://doi.org/10.1146/annurev.ps.35.020184.000245>

Gallistel, C. R. (1990). Representations in animal cognition: an introduction [Review of *Representations in animal cognition: an introduction*]. *Cognition*, 37(1-2), 1–22.

Gibbon, J. (1971). Scalar timing and semi-markov chains in free-operant avoidance. In *Journal of Mathematical Psychology* (Vol. 8, Issue 1, pp. 109–138).

[https://doi.org/10.1016/0022-2496\(71\)90025-3](https://doi.org/10.1016/0022-2496(71)90025-3)

Gibbon, J. (1972). Timing and discrimination of shock density in avoidance. In *Psychological Review* (Vol. 79, Issue 1, pp. 68–92). <https://doi.org/10.1037/h0031887>

Gibbon, J. (1999). multiple time scales is well named. *Journal of the Experimental Analysis of Behavior*, 272–275.

Gibbon, J., & Church, R. M. (1981). Time left: Linear versus logarithmic subjective time. In *Journal of Experimental Psychology: Animal Behavior Processes* (Vol. 7, Issue 2, pp. 87–108). <https://doi.org/10.1037/0097-7403.7.2.87>

Gibbon, J., & Church, R. M. (1981). Time left: Linear versus logarithmic subjective time. *Journal of Experimental Psychology*, 7(2), 87–108.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar Timing in Memory. In *Annals of the New York Academy of Sciences* (Vol. 423, Issue 1 Timing and Ti, pp. 52–77). <https://doi.org/10.1111/j.1749-6632.1984.tb23417.x>

Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. R. (1997). Toward a neurobiology of temporal cognition: advances and challenges. In *Current Opinion in Neurobiology* (Vol. 7, Issue 2, pp. 170–184). [https://doi.org/10.1016/s0959-4388\(97\)80005-0](https://doi.org/10.1016/s0959-4388(97)80005-0)

Gibson, J. J. (1975). Events are Perceivable But Time Is Not. In *The Study of Time II* (pp.

295–301). https://doi.org/10.1007/978-3-642-50121-0_22

Gil, S., & Droit-Volet, S. (2012). Emotional time distortions: the fundamental role of arousal. *Cognition & Emotion*, 26(5), 847–862.

Goldstone, S., & Lhamon, W. T. (1974). Studies of Auditory-Visual Differences in Human Time Judgment: 1. Sounds are Judged Longer than Lights. In *Perceptual and Motor Skills* (Vol. 39, Issue 1, pp. 63–82). <https://doi.org/10.2466/pms.1974.39.1.63>

Goldstone, S., Lhamon, W. T., & Sechzer, J. (1978). Light intensity and judged duration. In *Bulletin of the Psychonomic Society* (Vol. 12, Issue 1, pp. 83–84). <https://doi.org/10.3758/bf03329633>

Grondin, S. (2003). Studying Psychological Time with Weber's Law. In *The Nature of Time: Geometry, Physics and Perception* (pp. 33–41). https://doi.org/10.1007/978-94-010-0155-7_5

Grondin, S. (2008). *Psychology of Time*. Emerald Group Publishing.

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. In *Attention, Perception, & Psychophysics* (Vol. 72, Issue 3, pp. 561–582). <https://doi.org/10.3758/app.72.3.561>

Grondin, S. (2012). Violation of the scalar property for time perception between 1 and 2 seconds: Evidence from interval discrimination, reproduction, and categorization. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 38, Issue 4, pp. 880–890). <https://doi.org/10.1037/a0027188>

Grondin, S. (2014). About the (non)scalar property for time perception. *Advances in Experimental Medicine and Biology*, 829, 17–32.

- Gu, B.-M., & Meck, W. H. (2011). New Perspectives on Vierordt's Law: Memory-Mixing in Ordinal Temporal Comparison Tasks. In *Multidisciplinary Aspects of Time and Time Perception* (pp. 67–78). https://doi.org/10.1007/978-3-642-21478-3_6
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology: CB*, *17*(17), R751–R753.
- Haberman, J., & Whitney, D. (2012). Ensemble Perception. In *From Perception to Consciousness* (pp. 339–349). <https://doi.org/10.1093/acprof:osobl/9780199734337.003.0030>
- Hollingworth, H. L. (1910). The Central Tendency of Judgment. In *The Journal of Philosophy, Psychology and Scientific Methods* (Vol. 7, Issue 17, p. 461). <https://doi.org/10.2307/2012819>
- Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026.
- Johnson, K. O., Hsiao, S. S., & Yoshioka, T. (2002). Neural coding and the basic law of psychophysics. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *8*(2), 111–121.
- Jozefowicz, J., Gaudichon, C., Mekkass, F., & Machado, A. (2018). Log versus linear timing in human temporal bisection: A signal detection theory study. *Journal of Experimental Psychology. Animal Learning and Cognition*, *44*(4), 396–408.
- Karaminis, T., Cicchini, G. M., Neil, L., Cappagli, G., Aagten-Murphy, D., Burr, D., & Pellicano, E. (2016). Central tendency effects in time interval reproduction in autism. In *Scientific Reports* (Vol. 6, Issue 1). <https://doi.org/10.1038/srep28570>

- Klapproth, F. (2009). Single-Modality Memory Mixing in Temporal Generalization: An Effect Due to Instructional Ambiguity. In *NeuroQuantology* (Vol. 7, Issue 1).
<https://doi.org/10.14704/nq.2009.7.1.210>
- Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. In *Behavioral and Brain Sciences* (Vol. 12, Issue 2, pp. 251–267).
<https://doi.org/10.1017/s0140525x0004855x>
- Leib, A. Y., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, 7, 13186.
- Lejeune, H., & Wearden, J. H. (2006). Scalar properties in animal timing: conformity and violations. *Quarterly Journal of Experimental Psychology*, 59(11), 1875–1908.
- Lejeune, H., & Wearden, J. H. (2009). Vierordt's The Experimental Study of the Time Sense(1868) and its legacy. In *European Journal of Cognitive Psychology* (Vol. 21, Issue 6, pp. 941–960). <https://doi.org/10.1080/09541440802453006>
- Lewis, P. A., & Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Current Opinion in Neurobiology*, 13(2), 250–255.
- Lhamon, W. T., & Goldstone, S. (1974). Studies of auditory-visual differences in human time judgment. 2. More transmitted information with sounds than lights. *Perceptual and Motor Skills*, 39(1), 295–307.
- Luce, R. D., & Edwards, W. (1958). The derivation of subjective scales from just noticeable differences. *Psychological Review*, 65(4), 222–237.
- Lustig, C., & Meck, W. H. (2011). Modality differences in timing and temporal memory

throughout the lifespan. *Brain and Cognition*, 77(2), 298–303.

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. In *Visual Neuroscience* (Vol. 26, Issue 1, pp. 147–155). <https://doi.org/10.1017/s0952523808080905>

Matthews, W. J., & Meck, W. H. (2014). Time perception: the bad news and the good. *Wiley Interdisciplinary Reviews. Cognitive Science*, 5(4), 429–446.

Matthews, W. J., Stewart, N., & Wearden, J. H. (2011a). Stimulus intensity and the perception of duration. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 37, Issue 1, pp. 303–313). <https://doi.org/10.1037/a0019961>

Matthews, W. J., Stewart, N., & Wearden, J. H. (2011b). Stimulus intensity and the perception of duration. In *Journal of Experimental Psychology: Human Perception and Performance* (Vol. 37, Issue 1, pp. 303–313). <https://doi.org/10.1037/a0019961>

Mauk, M. D., & Buonomano, D. V. (2004). THE NEURAL BASIS OF TEMPORAL PROCESSING. In *Annual Review of Neuroscience* (Vol. 27, Issue 1, pp. 307–340). <https://doi.org/10.1146/annurev.neuro.27.070203.144247>

Meck, W. H., & Benson, A. M. (2002). Dissecting the Brain's Internal Clock: How Frontal–Striatal Circuitry Keeps Time and Shifts Attention. In *Brain and Cognition* (Vol. 48, Issue 1, pp. 195–211). <https://doi.org/10.1006/brcg.2001.1313>

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology. Animal Behavior Processes*, 9(3), 320–334.

Meck, W. H., Komeily-Zadeh, F. N., & Church, R. M. (1984). Two-step acquisition:

- Modification of an internal clock's criterion. In *Journal of Experimental Psychology: Animal Behavior Processes* (Vol. 10, Issue 3, pp. 297–306).
<https://doi.org/10.1037/0097-7403.10.3.297>
- Mioni, G., Stablum, F., McClintock, S. M., & Grondin, S. (2014). Different methods for reproducing time, different results. In *Attention, Perception, & Psychophysics* (Vol. 76, Issue 3, pp. 675–681). <https://doi.org/10.3758/s13414-014-0625-3>
- Nieder, A., & Miller, E. K. (2003). Coding of Cognitive Magnitude. In *Neuron* (Vol. 37, Issue 1, pp. 149–157). [https://doi.org/10.1016/s0896-6273\(02\)01144-3](https://doi.org/10.1016/s0896-6273(02)01144-3)
- Ogden, R. S., Wearden, J. H., & Jones, L. A. (2010). Are memories for duration modality specific? In *Quarterly Journal of Experimental Psychology* (Vol. 63, Issue 1, pp. 65–80).
- Oprisan, S. A., & Buhusi, C. V. (2014). What is all the noise about in interval timing? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1637), 20120459.
- Penney, T. (2003). Modality Differences in Interval Timing. In *Functional and Neural Mechanisms of Interval Timing*. <https://doi.org/10.1201/9780203009574.ch8>
- Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6), 1770–1787.
- Petzschnner, F. H. (2012). *Magnitude Estimation in Humans: A Bayesian Approach to Characteristic Behavior in Path Integration*.
- Petzschnner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: a study on human path integration. *The Journal of*

Neuroscience: The Official Journal of the Society for Neuroscience, 31(47),
17220–17229.

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, 19(5), 285–293.

Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological Science*, 24(8), 1389–1397.

Poirier, L. J., McGeer, E. G., Larochelle, L., McGeer, P. L., Bedard, P., & Boucher, R. (1969). The effect of brain stem lesions on tyrosine and tryptophan hydroxylases in various structures of the telencephalon of the cat. In *Brain Research* (Vol. 14, Issue 1, pp. 147–155). [https://doi.org/10.1016/0006-8993\(69\)90036-5](https://doi.org/10.1016/0006-8993(69)90036-5)

Proulx, M. J. (2010). Size Matters: Large Objects Capture Attention in Visual Search. In *PLoS ONE* (Vol. 5, Issue 12, p. e15293). <https://doi.org/10.1371/journal.pone.0015293>

Rammsayer, T. H., & Verner, M. (2015). Larger visual stimuli are perceived to last longer from time to time: The internal clock is not affected by nontemporal visual stimulus size. In *Journal of Vision* (Vol. 15, Issue 3, pp. 5–5). <https://doi.org/10.1167/15.3.5>

Rhodes, D. (n.d.). *Bayesian Inference in Human Time Perception*.
<https://doi.org/10.31234/osf.io/7fzbn>

Roberts, W. A. (2005). How do pigeons represent numbers? In *Behavioural Processes* (Vol. 69, Issue 1, pp. 33–43). <https://doi.org/10.1016/j.beproc.2005.01.005>

Roberts, W. A. (2006). Evidence that pigeons represent both time and number on a logarithmic scale. *Behavioural Processes*, 72(3), 207–214.

- Shi, Z., & Burr, D. (2016). Predictive coding of multisensory timing. In *Current Opinion in Behavioral Sciences* (Vol. 8, pp. 200–206). <https://doi.org/10.1016/j.cobeha.2016.02.014>
- Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences*, *17*(11), 556–564.
- Staddon, J. E., & Higa, J. J. (1999). Time and memory: towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior*, *71*(2), 215–251.
- Stubbs, D. A. (1976). Scaling of stimulus duration by pigeons. *Journal of the Experimental Analysis of Behavior*, *26*(1), 15–25.
- Sun, J. Z., Wang, G. I., Goyal, V. K., & Varshney, L. R. (2012). A framework for Bayesian optimality of psychophysical laws. *Journal of Mathematical Psychology*, *56*(6), 495–501.
- Taatgen, N., & van Rijn, H. (2011). Traces of times past: Representations of temporal intervals in memory. In *Memory & Cognition* (Vol. 39, Issue 8, pp. 1546–1560). <https://doi.org/10.3758/s13421-011-0113-0>
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the “internal clock.” In *Psychological Monographs: General and Applied* (Vol. 77, Issue 13, pp. 1–31). <https://doi.org/10.1037/h0093864>
- Treisman, M., Faulkner, A., Naish, P. L., & Brogan, D. (1990). The internal clock: evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*, *19*(6), 705–743.
- Trujano, R. E., & Orduña, V. (2015). Reducing bias and analyzing variability in the time-left procedure. *Behavioural Processes*, *113*, 132–142.

- Ungan, P., & Yagcioglu, S. (2014). Significant variations in Weber fraction for changes in inter-onset interval of a click train over the range of intervals between 5 and 300 ms. *Frontiers in Psychology, 5*, 1453.
- van Rijn, H. (2016). Accounting for memory mechanisms in interval timing: a review. In *Current Opinion in Behavioral Sciences* (Vol. 8, pp. 245–249).
<https://doi.org/10.1016/j.cobeha.2016.02.016>
- Varshney, L. R., & Sun, J. Z. (2013). Why do we perceive logarithmically? *Significance. Statistics Making Sense, 10*(1), 28–31.
- Varshney, L. R., & Sun, J. Z. (2014). Why Do We Perceive Logarithmically? In *The Best Writing on Mathematics 2014* (pp. 64–73). <https://doi.org/10.1515/9781400865307-007>
- Vatakis, A., & Spence, C. (2006a). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neuroscience Letters, 393*(1), 40–44.
- Vatakis, A., & Spence, C. (2006b). Temporal order judgments for audiovisual targets embedded in unimodal and bimodal distractor streams. *Neuroscience Letters, 408*(1), 5–9.
- Vierordt, K. (1868). *Der Zeitsinn nach Versuchen*.
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. In *Attention, Perception, & Psychophysics* (Vol. 72, Issue 4, pp. 871–884).
<https://doi.org/10.3758/app.72.4.871>
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*(11), 483–488.
- Wearden, J. (2016). SET and Human Timing. In *The Psychology of Time Perception* (pp.

27–63). https://doi.org/10.1057/978-1-137-40883-9_3

Wearden, J. H., & Lejeune, H. (2008). Scalar properties in human timing: conformity and violations. *Quarterly Journal of Experimental Psychology*, *61*(4), 569–587.

Webster, J., Kay, P., & Webster, M. A. (2014). Perceiving the average hue of color arrays. In *Journal of the Optical Society of America A* (Vol. 31, Issue 4, p. A283).

Wittmann, M. (2013). The inner sense of time: how the brain creates a representation of duration. *Nature Reviews. Neuroscience*, *14*(3), 217–223.

Woodrow, H. (1930). The reproduction of temporal intervals. In *Journal of Experimental Psychology* (Vol. 13, Issue 6, pp. 473–499). <https://doi.org/10.1037/h0070462>

Woodrow, H. (1933). Individual Differences in the Reproduction of Temporal Intervals. In *The American Journal of Psychology* (Vol. 45, Issue 2, p. 271).
<https://doi.org/10.2307/1414278>

Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision*, *7*(10), 2.1–5.

Yanagisawa, H. (2015). A Computational Model of Human Perception With Prior Expectation: Bayesian Integration and Efficient Coding. In *Volume 1B: 35th Computers and Information in Engineering Conference*. <https://doi.org/10.1115/detc2015-46669>

Yi, L. (2009). Do rats represent time logarithmically or linearly? *Behavioural Processes*, *81*(2), 274–279.

Zélanti, P. S., & Droit-Volet, S. (2012). Auditory and visual differences in time perception? An investigation from a developmental perspective with neuropsychological tests. *Journal of Experimental Child Psychology*, *112*(3), 296–311.

Acknowledgment

‘Don’t get sentimental, it always ends up drivel.’ - I guess I will ignore this advice and release all my sentiments on this page and in the meanwhile hope nobody would notice.

I feel I cannot express my gratitude enough to all the people that I encountered during my stay in Germany. I would like to first thank Professor Hermann Müller, for accepting me as his student in the first place. I regretted a bit for not having more communications with him since most of our deeper ‘conversations’ lied in specific questions while working on the same manuscripts. But even from limited chances, I’ve gained some lifelong lessons just learning by his side and being amazed by his beautiful writing and rigorous attitude.

My deepest gratitude goes to Dr. Zhuanghua Shi (Strongway), who guided me through my whole Ph.D. study. I had on many occasions (behind his back) described him as an omnipotent super-man-visor to my friends. It doesn’t matter if a hardware issue, programming error or mathematical problem occurred, he is always capable of providing solutions. On top of his superior skills, he is the most patient and positive person one could imagine. Without his guidance and help, I couldn’t have fought against my ridiculous procrastination and wrote to this page of my thesis.

Special thanks to my TAC meeting advisor Dr. Flagin (Virginia) for being such an intelligent, patient and talented role model. Many thanks to all the lovely members of the Msense lab at the Experimental Psychology department of LMU. Special thanks to our modeling master Dr. Allenmark (Fredrik) for his patience and all the nice discussions in the Mensa Cafe, where I learned so much about data modeling.

I would like to thank the GSN management team, in particular Lena, Stephanie, Nadine, for always making me at ease, for always being there for any student who needs help. Special thanks to Professor Benedikt Grothe, for providing the best environment of a graduate school, both academically and socially, for always being very inspiring and caring. I’m sincerely grateful for all the conversations we had and the fact that he always has time for students.

I always say that I couldn’t have survived without my friends and this being the most truthful truth I ever tell. I give my hug to Adriana, Anja, Aurore, Calandra, Ebru, Efsun, Felix, François, Jairo, Judita, Luyao, Mario, Martin, Melanie, Natan, Ola, Yue (Zhang), Yang (Wei), thank you all for being so warm-hearted, inspiring and pretty, for always be there for this troublemaker, inviting me to your houses and feeding me, and providing me endless support and consolation. I feel being the luckiest person to be able to have all the genuine conversations and beautiful moments with you guys.

最后，爸爸妈妈，想让你们知道作为你们的女儿是一件多幸运的事情。

Curriculum Vitae

EDUCATION

Ludwig-Maximilians-University	Munich, Germany
Ph.D. (candidate) Neuroscience	07/2020 (defense date)
Tianjin University	Tianjin, China
Master of Biomedical Engineering	06/2013
Tianjin University	Tianjin, China
Bachelor of Biomedical Engineering	06/2010

RESEARCH EXPERIENCE

Department of Psychology, LMU	Munich
Research Fellow	10/2014 - 04/2020
Graduate School of Systemic Neurosciences, LMU	Munich
Ph.D. student	since 10/2014
Advisory Committee: Hermann J. Müller; Zhuanghua Shi; Virginia L. Flanagin	
Department of Tianjin University	Tianjin, China
Neural Engineering and Rehabilitation Lab	
Student Research Assistant	10/2011 - 06/2013

List of Publications

Yue Ren, Fredrik Allenmark, Hermann J Müller, Zhuanghua Shi (2020). Logarithmic encoding of ensemble time intervals. *Scientific Reports* 10: 18174.

Yue Ren, Fredrik Allenmark, Hermann J Müller, Zhuanghua Shi (2021). Variation in the ‘coefficient of variation’: rethinking the violation of the scalar property in time-duration judgments. *Acta Psychologica* 214: 103263.

Yue Ren, Hermann J Müller, Zhuanghua Shi. Two types of memory mixing: Influences of stimulus modality and non-temporal magnitudes in duration judgment (to be submitted).

Selected Conference Presentations:

‘Is Absolute Pitch a switchable mode? An EEG study’ (2017). The Neurosciences and Music VI: Music, Sound, and Health. Harvard Medical School, Boston, USA.

‘Temporal Averaging: Geometric or Arithmetic mean?’ (2016). Society for Neuroscience 2016. San Diego, USA.

‘Temporal Context influences Coefficient of Variances of Interval Reproduction’ (2015). Donders Discussion. Nijmegen, the Netherlands.

Eidesstattliche Erklärung / Affidavit

Ich versichere hiermit an Eides statt, dass die vorliegende Dissertation „Context Effects in Interval Timing“ selbstständig angefertigt habe, mich außer der angegebenen keener weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation ‘Context Effects in Interval Timing’ is the result of my own work and that I have only used sources or material listed and specified in the dissertation.

München, den

Munich, date

16.07.2021

Yue Ren

Declaration of author contributions:

Manuscript 1:

Yue Ren, Fredrik Allenmark, Hermann J Müller, Zhuanghua Shi. Logarithmic encoding of ensemble time intervals. Scientific Reports 10, 18174 (2020).

This work was carried out under the supervision of Zhuanghua Shi, and in cooperation with Fredrik Allenmark; Z.S. conceived the idea, Y.R. and Z.S. designed the research, Y.R. collected and analyzed the data, Y.R., Z.S. and H.J.M discussed the results and wrote the paper.

Manuscript 2:

Yue Ren, Fredrik Allenmark, Hermann J Müller, Zhuanghua Shi. Variation in the 'coefficient of variation': rethinking the violation of the scalar property in time-duration judgments. Acta Psychologica 214, 2021, 103263.

This work was carried out under the supervision of Zhuanghua Shi, and in cooperation with Fredrik Allenmark; Y.R. and Z.S. conceived the idea, designed the research, Y.R. collected and analyzed the data. Y.R and F.A modeled the data. Y.R., F.A., H.J.M and Z.S. discussed the results and wrote the paper.

Manuscript 3:

Yue Ren, Hermann J Müller, Zhuanghua Shi (in preparation). Two types of memory mixing effects: Influences of the modality and non-temporal magnitudes in duration perception.

This work was carried out under the supervision of Zhuanghua Shi; Z.S. conceived the idea, Y.R. and Z.S. designed the research, Y.R. collected and analyzed the data. Y.R. and Z.S. discussed the results and wrote the paper. H.J.M commented on the manuscript.

München,

Hermann J. Müller

Zhuanghua Shi

Fredrik Allenmark