Forschungsberichte zum Themenschwerpunkt

Time as the Ink That Music Is Written With: A Review of Internal Clock Models and Their Explanatory Power in Audiovisual Perception

Zeit als Grundlage der Musik: Ein Überblick zu Modellen innerer Uhren und deren Erklärungswert für die audiovisuelle Wahrnehmung

Xinyue Wang*[a], Clemens Wöllner[a]

[a] Institut für Systematische Musikwissenschaft, Universität Hamburg, Hamburg, Germany.

Abstract

The current review addresses two internal clock models that have dominated discussions in timing research for the last decades. More specifically, it discusses whether the central or the intrinsic clock model better describes the fluctuations in subjective time. Identifying the timing mechanism is critical to explain and predict timing behaviours in various audiovisual contexts. Music stands out for its prominence in real life scenarios along with its great potential to alter subjective time. An emphasis on how music as a complex dynamic auditory signal affects timing accuracy led us to examine the behavioural and neuropsychological evidence that supports either clock model. In addition to the timing mechanisms, an overview of internal and external variables, such as attention and emotions as well as the classic experimental paradigms is provided, in order to examine how the mechanisms function in response to changes occurring particularly during music experiences. Neither model can explain the effects of music on subjective timing entirely: The intrinsic model applies primarily to subsecond timing, whereas the central model applies to the suprasecond range. In order to explain time experiences in music, one has to consider the target intervals as well as the contextual factors mentioned above. Further research is needed to reconcile the gap between theories, and suggestions for future empirical studies are outlined.

Keywords: internal clock models, Dynamic Attending Theory, Scalar Expectancy Theory, music perception, audiovisual timing

Zusammenfassung

Dieser Überblick befasst sich mit zwei Modellen der inneren Uhr, die in den letzten Jahrzehnten die Diskussion in der Forschung zur Zeitwahrnehmung und -gestaltung bestimmt haben. Insbesondere wird diskutiert, ob das zentrale oder das intrinsische Uhrenmodell Schwankungen der subjektiven Zeit besser erklärt. Dabei ist das Erkennen des zugrundeliegenden Mechanismus' entscheidend, um das Zeiterleben im Nachhinein zu erklären oder in verschiedenen audiovisuellen Kontexten vorherzusagen. Musik zeichnet sich durch ihre Bedeutung in realen Szenarien sowie durch ihr großes Potenzial zur Veränderung des subjektiven Zeiterlebens aus. Musik kann als komplexes dynamisches Audiosignal die zeitliche Genauigkeit beeinflussen. Dies ist der Hintergrund, verhaltensbezogene und neuropsychologische Belege zu diskutieren, die eines der Uhrenmodelle oder beide unterstützen. Neben den Zeitmechanismen wird ein Überblick auf interne und externe Variablen wie Aufmerksamkeit und Emotion, sowie auf klassische experimentelle Paradigmen gegeben. Dadurch wird dargelegt, welche Rolle den Mechanismen zukommt hinsichtlich der Reaktion auf Änderungen im Stimulusmaterial, insbesondere beim Erleben von Musik. Im Ergebnis kann kein Modell die Auswirkungen von Musik auf das subjektive Zeiterleben vollständig erklären. Während das intrinsische Modell in erster Linie das Zeiterleben für sehr kurze Dauern unterhalb einer Sekunde zu erklären vermag, bietet das zentrale Modell einen höheren Erklärungswert für den Suprasekundenbereich, das heißt für das Timing von Sekunden bis Minuten. Um Zeiterfahrungen in der Musik zu erklären, müssen die Zielintervalle sowie die oben genannten Kontextfaktoren berücksichtigt werden. Weitere Forschungen sind erforderlich, um die Kluft zwischen den Theorien zu schließen, wobei Vorschläge für künftige empirische Studien skizziert werden.

Schlüsselwörter: Innere Uhrenmodelle, Dynamic Attending Theory, Scalar Expectancy Theory, Musikwahrnehmung, audiovisuelles Zeiterleben

*Korrespondenzanschrift: Institut für Systematische Musikwissenschaft, Universität Hamburg, Neue Rabenstr. 13, 20354 Hamburg, Germany. E-Mail:
xinyue.wang@uni-hamburg.de

Properties of time have attracted the interest of researchers since long. From a Newtonian perspective, time is seen as an arrow flying eternally forward, whereas for classical thermodynamics, the passage of time shares similarity with the irreversible increase of entropy, or the degree of disorder in the universe (Lieb & Yngvason, 1999). The psychological study of time perception, in comparison, held different opinions. One of the earliest efforts in capturing an internal timing system stemmed from doctors' skills in making accurate estimates of time based on heartbeats and breathing (Goudriaan, 1921). Composers, as experts of time and timing, have discovered the link between the tempo in which their works are played and the impression of durations among the audience. Ravel once complained to Furtwängler that his Boléro, when played too fast, would feel unjustifiably long (Nichols, 2011). Ravel's somewhat paradoxical observation comes close to that of cognitive scientists, such that music played at various tempi could induce corresponding time distortions (e.g., Droit-Volet et al., 2010).

Much as Ravel perceived, music is a form of art closely intertwined with time. A rich vein of literature has pointed out that rhythmic patterns, or beats, are fundamentally embedded in all genres of music, leading to perceptual periodicities (London, 2004; Nozaradan, 2014). The perceived periodicities provide a sequence of external events, which could subsequently be internalized as representations of time (Droit-Volet et al., 2013; Gibbon et al., 1984; Jones & Boltz, 1989). Music- or beat-induced movements are noticed across a wide range of ages as the substantiation of an individuals' anticipation of the rhythmic patterns – among infants (Zentner & Eerola, 2010), pre-schoolers (Eerola et al., 2006), and adults (Burger et al., 2018). By either proactively or passively synchronising to the musical beats, an individual's perceptual time is subject to modifications. The general tendency indicates that fast music leads to duration overestimation, whereas slow music to underestimations (Droit-Volet et al., 2013; Wang & Shi, 2019).

In this review, timing refers to the active process of monitoring temporal order by explicit (e.g., tapping) or implicit (e.g., silent counting) actions, with an emphasis on the efforts involved in the task. Meanwhile, duration estimation, the action of gauging past time, composes the second part of time perception. It encompasses the concepts of both prospective (knowing before that the duration of an event should be judged) and retrospective timing (judging duration afterwards) after the target duration has elapsed (Zakay & Block, 1995). In this sense, one could passively experience the passage of time with or without attending to the passage of time itself, which affects subsequent judgments.

In relation to duration estimation, it is equally important to underline the role of beat perception, or inner timing as the ability to perceive and predict the temporal location of events. As a supporter of the cerebral clock model, Pöppel (1989) hypothesized that maintaining a constant tempo in music production, especially in classical music, has to do with a time keeping mechanism that functions mainly by tracking the temporal order such as synchrony and succession in addition to measuring durations. In the same vein, a "3-second window of temporal integration" (Pöppel, 1989, p. 86) was assumed to constitute the psychological present. This has consequences for perceiving musical tempo and the integration of beats, and hence subjective experiences of time. Similar attempts for developing clock models were based on beat perception (Langner, 2002; Schulze, 1978). Schulze, in particular, pivoted the *Dynamic Attending Theory* (Jones & Boltz, 1989) that emerged later by emphasizing the variability of internal clock speed under the influence of environmental cues (accelerating and decelerating beat patterns).

While some studies investigated musical tempo in order to formulate hypotheses about clock models, other research found that the variables embedded in tempo, such as isochrony, salience, or complexity, directly evoked changes in the functioning of the internal clock. Povel and Essens (1985) observed in their experiments that different grouping of rhythmic beats led to various temporal reproductions, giving rise to a best fit for the internal clock. Explanations lie in the coupling of beat accents and the clock 'tick': the stronger the beat pattern is (in this case, higher metrical level), and the less complex the metrical structure is, the more likely the beat would activate the internal time recording system and be represented in temporal processing. Recognizing beat perception not only helps understanding human timing better, but also particularly with music listening and performing, which can also be understood in terms of proactive and active timing. In fact, frequent exposure to musical beat production appears to enhance one's temporal sensitivity, and this effect may transcend to other sensory modalities (from audition to vision; Cicchini et al., 2012). Apart from external training, the stability of intrinsic rhythm was also positively correlated with tempo reproduction performances (McPherson et al., 2018). It is therefore essential to look into the complexity in musical tempo itself.

Musical tempo is subject to ambiguity. The complexity of tempo structures in music has long been recognized (e.g., Pressnitzer et al., 2011). It is not only marked by the number of note events in a melody (Behne, 1976), nor only by the patterns of percussion instruments, but rather by changes in pitch, timbre, or loudness (Brochard et al., 2003), as well as phrasing and articulation (Auhagen & Busch, 1998). Multiple sound sources of the instruments in a symphony orchestra vary tremendously across different sections and therefore constitute auditory streams that are hard to disentangle (Shamma & Micheyl, 2010), especially for non-musicians. Note that the difficulty of correctly identifying temporal structures in music is not equal to that of correctly identifying the tempo of music, considering the latter has more to do with detecting the absolute 'speed' and tempo changes. Attempts have been made to examine the thresholds of detecting musical tempo acceleration and deceleration, for instance, among musically trained and untrained groups (e.g., Ellis, 1991). There are several assumptions of how we cope with "noisy" auditory signals in terms of time and tempo perception. Some argued that the process of tempo extraction depends mainly on periodic regularities (McDermott et al., 2011), while others emphasized the importance of learning, regardless of tempo structure complexities (Agus et al., 2010).

A small number of studies aimed at the disentanglement of auditory rhythmic features and revealed how tempo salience affects perceptual time. One study investigated different metrical levels and found effects on listeners' sense of time (Hammerschmidt & Wöllner, 2020). More specifically, the lower the metrical level individuals at-

tended to by tapping (e.g., eight notes versus half notes), the longer a music excerpt was perceived, providing some evidence for the impact of event density (cf. Behne, 1976). In this case, the count of time was affected by the number of beats registered in memory.

Apart from music, inputs from other sensory modalities may also affect temporal processing. Indeed, psychological research has often used visual stimuli such as flashes or flickering lights to investigate time. For instance, studies of the entrainment effect for independent modalities showed that the presence of either visual flickers or pure tones led to higher entrainment (e.g., Ortega & López, 2008; Treisman & Brogan, 1992; Treisman et al., 1990). The effect, nevertheless, is not limited to one modality. Past research suggests that auditory signals of various complexities could enhance the entrainment effect for visual sequences and, in some cases, were transferable to the attention acuity of the other modality (Bolger et al., 2013; Escoffier et al., 2010). In Bolger et al.'s study, participants were able to perform equally well in a target detection task regardless of the target modality (auditory or visual) when entrained with tone sequences. Another case in point is the cross-modal transfer in tempo discrimination between auditory and tactile domains, where training with rhythmic sounds led to enhanced performance in that of the latter (Nagarajan et al., 1998). These studies provide evidence that the cognitive processes involved in timing and time perception should function at a domain-general level.

In this review, an overview is provided of internal clock models that were established or further developed in recent years. In particular, the aim was to show how each model accounts for the experience of musical time in auditory and audiovisual contexts. We will tackle questions such as: How does music facilitate temporal processing? What are the timing mechanisms and models, and how do they explain the inference between music and perceptual time, respectively? What are the implications of studying music and time perception?

The Internal Clock

Comparable to an actual clock, the internal clock has been an analogy for the timing mechanism in human and animals (Eagleman et al., 2005; Ivry & Schlerf, 2008). The temporal order of events is recorded by multiple sensory modalities and processed in, according to different theories, a variety of pathways before becoming representations of time, that is, the occurrence of "clock ticks". Early in the discussion, hypotheses stated that time perception was a form of information processing that highly depended on the recording capacity (Ornstein, 1969). Researchers such as Barry (1990) and Schulze (1978) both emphasized the importance of music as an environmental construct of attention that shaped both the perceived time (in terms of its duration) and the passage of time (the perceived speed).

In the past years, two major theories were on the forefront in discussions as to how the temporal units are recorded, both postulating the presence of a specific cognitive module dedicated to timing. The 'no clock' hypothesis, or state-dependent network, and the 'central clock' hypothesis have both received increasing attention in research (Grondin, 2010b). The latter, in particular, encompasses two theories: The Dynamic Attending Theory, based on a non-linear cumulation of temporal units, as well as the Scalar Expectancy Theory, which assumed that the emission of temporal pulses follows a linear approach (Ivry & Schlerf, 2008). Such as Stern (1897) had already pointed out for "Präsenzzeit" (the experienced present moment) and the time range for other cognitive processes, it appears that different theories function best at specific time ranges (Figure 1).
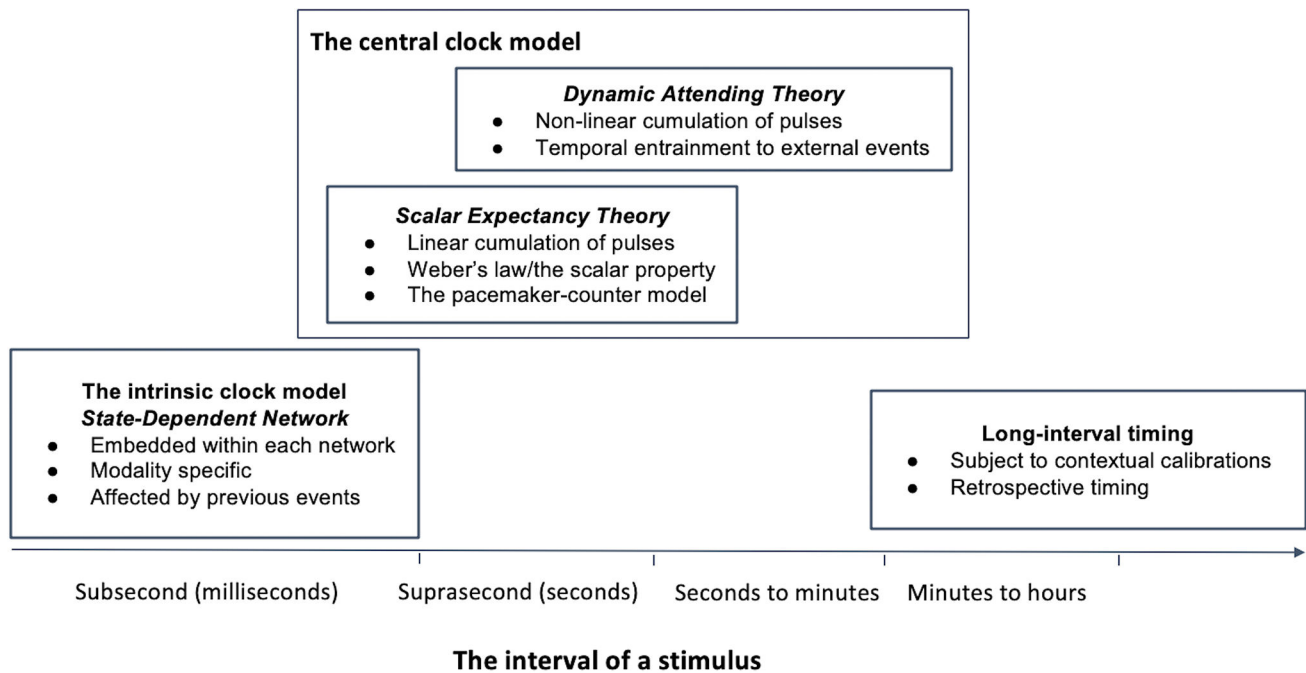
The central clock model

> ***Dynamic Attending Theory***
> - Non-linear cumulation of pulses
> - Temporal entrainment to external events

> ***Scalar Expectancy Theory***
> - Linear cumulation of pulses
> - Weber's law/the scalar property
> - The pacemaker-counter model

The intrinsic clock model
State-Dependent Network
- Embedded within each network
- Modality specific
- Affected by previous events

Long-interval timing
- Subject to contextual calibrations
- Retrospective timing

Subsecond (milliseconds)   Suprasecond (seconds)   Seconds to minutes   Minutes to hours

The interval of a stimulus

Figure 1. An overview of the internal clock models specified by interval ranges (subsecond, suprasecond, seconds to minutes, and minutes to hours) as well as by the division of central vs. intrinsic model.

Note. The most important features described in the overview were discussed in detail in the following sections. For a review of the research methods adopted in timing studies, see Grondin (2010b).

The Intrinsic Clock Model

Unlike the traditional view of a clock, some researchers believe that there might be no clock at all. Such a 'no-clock' model is known as a state-dependent timing system or intrinsic model (Ivry & Schlerf, 2008). The 'state' here describes the specific circumstances generated by a neural network in response to external changes. Timing is seen as an implicit function of each neural network that is activated for a given sensory modality and is sensitive to pre- and post-interval changes. It is postulated that activity-elicited changes in neural networks directly reflect the inherent temporal structures and therefore serve as references for timing in sub-second intervals (e.g., Karmarkar & Buonomano, 2007). The process is also referred to as "a temporal-to-spatial transformation" (Karmarkar & Buonomano, 2007, p. 3), or the "intrinsic model", as researchers hypothesize that timing is an integral function of neural activity (Ivry & Schlerf, 2008).

The model essentially suggests that timing as a function is distributed to a variety of neural structures in which oscillatory patterns stay consistent, known as the recurrent neural network (Buonomano & Laje, 2011). Ramping, or climbing activities in neural oscillations ranging from primarily low frequency gamma band to higher frequency such as beta and alpha band (e.g., Wittmann, 2013) have been revealed as the physiological basis for the model, in addition to neural spikes across a wide range of brain regions such as the striatum (Gu et al., 2015). The time stamps, or accumulated states, are hypothesized to be expressed on both micro (individual neurons) as well as macro (populatory neuron excitation/inhibition) levels (Buonomano & Laje, 2011).

The intrinsic model proposes the possibility that timing is an inherent function of multiple dynamic neural networks. The flexibility allows the network to take into account calibrations towards previous durations and to judge the duration of the current event on this basis. Furthermore, by hypothesizing the implicity of timing, there is no need for external triggers even when an event is absent. However, the state-dependent network also has its shortcomings. Studies suggested that the model is only applicable to the subsecond range. That is to say, the cumulative effects of previous events, either enhancing or reducing one's temporal sensitivity, diminished within up to 300 milliseconds (Buonomano et al., 2009). On the other hand, it does not offer a clear explanation of cross-modal temporal information integration. This is where the central clock model provides a useful alternative perspective.

The Central Clock Model

The central timing mechanism, also known as the dedicated clock model (e.g., Allman et al., 2014), stemmed from Treisman's (1963) work. Decades of research into the human timing mechanism were based on this model and have assumed that timing is a specific cognitive module, hypothetically located across the global neural network (e.g., Allman & Meck, 2012).

Where is the 'clock' in our brain? Neurological studies supporting the timing mechanism as an independent cognitive module that is dedicated solely to this function, however, do not necessarily assume one single structure in the brain for it. Research rather supports the roles of a wide range of brain regions working collaboratively in order to process time (e.g., Buhusi & Meck, 2005). The cerebellum, for example, is involved in short duration judgments, arguably from a few hundred milliseconds to 30 seconds (Allman & Meck, 2012). Disruptions to other cortical and subcortical structures including the basal ganglia can lead to timing deficits also in larger time frames. In Schwartze and colleagues' (2011) study, participants with basal ganglia lesion failed to detect tempo acceleration and deceleration of tones and could not entrain tapping movements with the signals. The evidence implicates a global network where multiple brain regions are involved; impairments at any part of the chain could possibly lead to malfunctions of the timing mechanism. There are two prominent theories with consequences for time processing in music that will be explained below.

The Dynamic Attending Theory (DAT)

DAT, also known as the oscillator model, hypothesizes that the ability to estimate the duration of past events depends on the coupling between attentional pulses and the occurrences of external events (Jones & Boltz, 1989). The theory supports the presence of a central clock, in the sense that the allocation of attention as a limited resource is based on the expectation of the next event on the timeline (Large & Jones, 1999). An exogenous stimulus, when aligned with the peak of attention, is best retained in working memory and transformed into representations of time (Barnes & Jones, 2000). Essential to this theory is that, like the unstable periodicities of external events, the emission of attentional pulses or oscillations is a non-linear process (Large, 2008).

Holding the central clock premise, DAT suggests that attention plays a critical role in regulating the frequency of the pacemaker pulses according to the Attentional Gate Model (Block & Gruber, 2014; Zakay & Block, 1995). More specifically, the "gate" through which temporal units pass before registering with the counter device opens wider when more attention is assigned to the specific time point. When one's attention is shifted elsewhere irrelevant to temporal cues, fewer pulses are recorded, leading to duration underestimation. Some argue that

DAT applies exclusively to prospective timing, while in retrospective timing, it is subject to contextual influences and memory retrieval (Block & Gruber, 2014; Gu et al., 2015).

Note that DAT is hypothesized to function mostly within the suprasecond range, because prospective timing recedes with time due to limited capacities of working memory (WM) (for a review, see Gu et al., 2015). Concurrent tasks that require extra attentional resources could reduce timing accuracy (Brown & Boltz, 2002). Polti et al. (2018) attempted to explore the interval boundary of attention in prospective timing and found that the magnitude of WM interference on time estimation tasks increased proportionally with interval lengths (30 to 90s). In a more naturalistic setting, gamers were asked to estimate the elapsed time (12, 35, or 58 minutes) either knowingly (prospective timing) or not (retrospective; Tobin et al., 2010). The 12-minute session was estimated significantly longer in the prospective than the retrospective paradigm, while estimation differences were less pronounced in 35- and 58-minute conditions, suggesting that DAT's predictive power may be reduced in longer intervals. However, no evidence so far has made a clear cut of the interval ranges where each model fits best. This question clearly deserves further exploration and will be discussed in the conclusion.

The Scalar Expectancy Theory (SET)

Another well-known model for the internal timing mechanism argues that the perceived amount of time is composed of regularly emitted pulses from a pacemaker, as an analogy of an internal clock, and accumulated by a counter device, therefore also known as the pacemaker-counter model (Gibbon, 1977; Gibbon et al., 1984; Treisman, 1963). This model specifies that the temporal process is accomplished through roughly three different steps: the clock, the memorisation, and the judgment of time. Accordingly, in order to explain the temporal flow, SET proposes that subjective time is composed of (a) the representation of objective durations, and (b) the estimation variance or error rate by Weber's fraction (Allman & Meck, 2012; Grondin, 2010b). The variance is hypothesized to stem from transferring the clock readings to working memory (Meck, 1984; Treisman, 1963). Inaccuracy in duration estimation, according to SET, is also subject to the influence of attention, clock speed error, task switch, decision error and other factors (Allman & Meck, 2012). The longer the duration, the larger the variance. It is, however, observed that SET has a controversial applicability in the sub-second to supra-second range. Grondin (2010a) focused on the violation of scalar property when examining participants' timing performance in a subsecond range, and found a tendency for Weber's fraction increasing as the interval approached 1s. This suggests that SET does not provide a powerful explanation of timing behaviour in the millisecond range. The applicability of SET in the millisecond to second range was further supported by the audiovisual evidence for this theory. Evidence is still needed for a clear boundary of 'time ranges of the best fit' for SET.

From a neurobiological perspective, the striatal beat frequency (SBF) theory offers an explanation for the original pacemaker-counter model (Miall, 1989; van Rijn et al., 2014). Unlike the latter, SBF theory instantiated the biological structure of the clock pulses as the oscillations of striatal medium spiny neurons (MSN), locating at the suprachiasmatic nucleus at the anterior hypothalamus. Researchers proposed that the neurons at different oscillatory frequencies reset when the timing begins, receiving inputs from cortical neurons firing as the consequences of dopaminergic releases (Merchant et al., 2013). Detection of the synchronous neural oscillation is known as the coincidental detection (Buhusi & Meck, 2005). The MSNs are capable of detecting coincidental oscillations from the cortical neurons that fire at similar frequencies, also known as the input, then translate to temporal units as the output. To justify the scalar property, that is variance in accumulative timing, it's been proposed that neural oscillations phase out and disperse into the inherent frequency of each neuron

after the initial alignment. As a result, the discrepancy increases proportionally until the neurons that were firing together completely desynchronize in the end. The MSNs, however, retain the robust ability to detect temporal patterns up to minutes despite the complexity of the inputs thanks to ironically the large number of cortical neurons (e.g., Matell et al., 2003), making the theory viable for a wider range of durations.

Factors Overarching Both Clock Models

Sensory Modalities

Multisensory inputs often interact with one another in our daily life. A vase dropping to the ground is usually followed by a shattering sound. A knock on the door leads to a knocking sound. To a broader extent, signals from vision, hearing, touch, smell and taste constitute the intangible framework of timing references together. Hence it is critical to understand the specificity of each sensory modality and their joint effects in temporal perception.

The dominant role of audition in temporal processing has been evidenced by a series of studies (e.g., Boltz, 2017; Chen et al., 2018; Repp & Penel, 2002). A number of studies supported higher precision in temporal discrimination in audition compared to vision (e.g., Large, 2008; Phillips & Hall, 2002). Furthermore, auditory temporal processing is capable of interfering with visual timing. In this case, participants' performances in identifying the correct rhythmic visual patterns were most heavily compromised when the task was accompanied by a new string of isochronous sounds rather than visual display (Guttman et al., 2005). One may assume that temporal information derived from auditory events weighs more than that of visual inputs. The auditory dominance view is, however, not without dispute. van Wassenhove and colleagues (2008) found that incongruent visual displays could distort temporal perception of auditory information in both directions. A recent finding, in addition, suggests that temporal perception was biased towards the visually perceived tempo of natural human movements rather than that of the drumbeats when the two sensory modalities were incongruent (Wang et al., 2019).

Research showed that auditory stimuli can effectively distort visual perception (e.g., Burr et al., 2013). The auditory driving effect emphasizes the perceived coupling of fluttering sounds to visual flicker rates, if the temporal gap between the flutter and the flicker does not exceed a certain range (Shipley, 1964). In other words, perceptual integration is accomplished by averaging auditory and visual input frequencies while endowing more weights on the former. A robust auditory driving effect could be observed when the sounds were presented as a brief distractor (Burr et al., 2013). Furthermore, Chen and colleagues' (2018) study suggested that, in addition to traditional regular flutters, irregular auditory inputs accompanying the visual flickers could also lead to distortions in perceiving the latter. Similar observations of the audiovisual bias were reported as fission and fusion illusions (Shams et al., 2002). In this case, the former specifies two visual events perceived as one when presented simultaneously with a beep, while in the latter, one flash is perceived as two when accompanied by two beeps.

Therefore, it is inevitable to take into account the arguably dominant position of the auditory modality when exploring the role of music in temporal processing. It should be noted that music encompasses not only complex acoustic signals, but a rich source of emotions that alter subjective time. Films are an example of how music shapes experiences of time. In a study investigating slow motion film scenes as compared to the

same scenes played back in real time, participants were significantly influenced in their temporal judgments of the scenes' duration when music was present (Wöllner et al., 2018). While slow motion scenes led to an underestimation of time, the same scenes in real time seemed to last relatively longer, and music yielded more accurate time estimations. Furthermore, music led to higher physiological arousal and larger pupil diameters in observers, suggesting that music modulates emotional responses and experiences of time in audiovisual scenes.

Working Memory and Attention

Central to the SET is the memory stage, in which working memory is retained, and the judgment stage, in which the current count of temporal units is compared to references retrieved from long-term memory (Gibbon et al., 1984; van Rijn et al., 2014). Individual differences in short-term memory capacity and discrepancies in timing performances bring attention to the role of working memory in temporal processing (e.g., Broadway & Engle, 2011). More specifically, higher working memory capacities imply higher potential to hold more time units at the second and third stage of timing, thus leading to more precision (Teki & Griffiths, 2014).

Working memory is positively related to other executive functions such as selective and divided attention (Colflesh & Conway, 2007), for both auditory and visual modalities (Wöllner & Halpern, 2016). Both shift in weights in various timing scenarios. This is particularly relevant for understanding the different mechanisms behind prospective and retrospective timing, as mentioned before (Block & Zakay, 1997). In the oscillator model (DAT), attentional pulses are emitted in order to track external beats. These pulses are recorded and transferred to working memory before entering the stage of comparison with a reference duration in long-term memory (Block & Zakay, 1997; Gibbon et al., 1984). Attention diverted from the timing task results in fewer temporal units taken into the count and consequently underestimations of time, while attention directed to timing led to overestimations regardless of test durations (Polti et al., 2018). Despite a lack of evidence, we hypothesize a similar result with music listening. When instructed to time a piece of music before it commences, a listener processes the passage of time differently than when asked to estimate the time elapsed at the end of the excerpt.

The interpretation of the roles of WM and attention also depends on the theories. DAT, compared to SET, highlights the role of attention rather than working memory (e.g., Jones, 2010; Jones & Boltz, 1989). It postulates that attention, when quantified as regular emitted pulses, could synchronize with external periodicity and therefore serve as a reference for time. The periodicities of external events, that is regular or irregular patterns, do affect the strength of their synchronisation with attentional pulses. The more predictable an exogenous pattern is, the better the effect, known widely as the temporal entrainment effect (Barnes & Jones, 2000; Schroeder & Lakatos, 2009), This has been evidenced by a number of visual (Cravo et al., 2013), auditory (Barnes & Jones, 2000; Jones, 2010), and movement (Burger et al., 2018) studies. Jones (1981, 1990) proposed that the characteristics of the information, in this case musical expressions, could distort the perception of time. Empirical studies supporting her claims found that, for instance, music was perceived to be slower when there were more pitch variations and inconsistent metrical accents (Boltz, 1998). We may predict that music genres with more predictable rhythms such as pop and rock, compared to those with less predictability such as Jazz, are associated with higher duration sensitivity and better timing accuracy.

Emotions

"Time flies when you are having fun". Understanding the nature of emotions in time perception is important to comprehend how music distorts subjective time, as it essentially conveys a wide spectrum of emotions. The relatively small number of studies that have directly looked into the effects of musical emotions on subjective time show that information of strongly emotional contents were more engaging and were subsequently better processed and stored in WM, leading to time overestimation (for a review, see Schäfer et al., 2013). Music as a powerful tool to induce emotions was found to induce a sense of timelessness (duration overestimation) as well as faster passage of time when an individual is completely submerged in the experience (Herbert, 2012). Apart from the aesthetic pleasure, other types and intensity of emotions may also have an impact on how music could distort the perception of time. The reasons may lie in the psycho-physiological arousal levels. Higher arousal level is believed to cause time overestimation (e.g., Droit-Volet et al., 2013). A group of participants, for instance, were presented with emotional film excerpts to induce corresponding emotions in them (Droit-Volet et al., 2011). Results indicate that, compared to baseline temporal judgments, participants tended to overestimate the durations after watching scary films. There are nevertheless findings implying the opposite, that is, higher emotional arousal leads to duration underestimation especially from a retrospective point of view (Herbert, 2012).

Another line of studies investigates the impact of emotional valences on temporal processing. Positive emotions, substantiated by happy music, led to duration underestimation, while negative emotions in sad music to duration overestimation with retrospective paradigms (Bisson et al., 2008). It was speculated that the positive emotions gave rise to less contextual changes than did the negative, therefore registering fewer events in the memory. Some evidence, on the contrary, implies that valence does not matter. Further investigations showed that highly arousing emotional pictures accelerated the internal clock speed and caused a leftward shift in the reaction time compared with pictures of low emotional arousal, regardless of its valence (Droit-Volet & Berthon, 2017).

The seemingly puzzling observations may be explained by the mechanism by which emotions take effect on time perception. One approach is rooted in the emission rates of attentional pulses, which can be moderated by the affective states, especially the arousal level. According to the pacemaker-counter model (Treisman, 1963), more attentional pulses are emitted when the arousal level is high, and subsequently be recorded as the sum of clock ticks, that is, the perceived duration. Attention could either facilitate or hinder the interaction between emotions and temporal processing. More specifically, when attention is allocated to sustaining the temporal units, the effect would lead to duration overestimation. In contrast, when attention is shifted from temporal information to the emotionally charged event, fewer ticks are accumulated, resulting in duration underestimation.

Modality-Specific Evidence for the Internal Clock Models

Audiovisual Evidence for the Intrinsic Clock Model

Time-dependent neural oscillations are specific to sensory modalities. Studies have revealed that neuron excitation and inhibition could be elicited according to a specific type of sensory input, such as sound (Schnupp et al., 2006) and visual flicker (Burr et al., 2007). Researchers found that the time-dependent decodability of

visual objects with MEG in a window of 1000ms varied significantly, suggesting that time might be an inherent feature in the local visual network (Carlson et al., 2013). Furthermore, transcranial magnetic stimulation studies revealed that auditory timing could be dissociated with that in other sensory modalities (Bueti et al., 2008), as participants performed worse in duration discrimination task (pure tones, 10 to 40ms) when receiving disruptions in the auditory cortex. We might as well propose that, when listening to complex auditory signals such as music, particular groups of neurons in the human auditory cortex generate time-dependent responses, which simultaneously serve as time codes. However, relatively few studies with humans have directly confirmed the time-dependent variability of the local auditory network (Toiviainen et al., 2019).

The disassociation in timing abilities among different sensory modalities also showed that time is processed as a local flow of information. Early findings entail significantly higher timing precision with hearing than with vision (e.g., Penney et al., 2000), indicating a superiority of audition over vision in providing temporal cues. Timing is a highly selective, localized process even within one modality. Burr and colleagues (2007) successfully modulated the perceived durations of the target visual stimuli by manipulating the apparent rate of flickers in a confined retinal region. Their finding is among one of the first to empirically support (a) the spatial-temporal connection in neural representations, and (b) the modality specificity in temporal processing, particularly the superiority of audition (e.g., Repp & Penel, 2002). In Lustig and Meck's (2011) study, the modality effect was stronger for participants at both ends of the age spectrum. One potential cause was that older adults were more susceptible to varying allocation of attention under different experimental conditions, whereas children might be influenced by developing sensory functions. That is not to say that SDN is a 'one modality, one clock' system, but rather a large network that also covers the interactions between multiple networks.

Taken together, from an intrinsic model's perspective, time is a consequence of cumulative states in a recurrent neural network that represent the amount of changes induced by external stimuli. In this sense, when listening to a piece of music repetitively, the perceived duration of both music and video (as a further stimulus) will be altered if presented again later on.

Audiovisual Evidence for the Dynamic Attending Theory

DAT is endowed with a particular emphasis on attention, given that the count of temporal units depends on how well attentional pulses synchronize with the external event, also known as the temporal entrainment effect. The term specifies the coupling of the tempo of extrinsic temporal cues and that of pacemaker pulses (Jones, 2010). The emphasis on external entrainment like music began in the early days of the formulation of the clock model (Barry, 1990; Pöppel, 1989). Neurobiological evidence suggests that the just noticeable differences for auditory gaps can be modulated when neural activities were entrained with specific frequency bands and amplitudes (Henry et al., 2014). Regarding music, the synchronization between neural oscillations and musical beats was substantiated as the steady-state event potential (SS-EP) evoked by periodicity in musical beats (Nozaradan, 2014).

Behavioural evidence provided similar findings. Fast tempo was found to lead to overestimation, or "time dilation", and slow tempo to underestimation, or "time contraction" in both auditory (Wang & Shi, 2019) and visual perception (Ortega & López, 2008). In addition, behavioural entrainment to external beats were found across age ranges and stimulus types, including auditory sequences and music excerpts (for a review, see Repp & Su, 2013). The experimental paradigms usually provide participants with a rhythmic beat that ceases

(or not) after a short period of entrainment and require them to continue tapping or moving along with the beats. Boasson and Granot (2012) adopted a paradigm of tapping to pitch rises and drops in multiple melodic sequences, in order to examine the entrainment effect. In their study, however, musicians and non-musicians uniformly exhibited faster-paced tapping behavior with rising pitch. This is consistent with other findings which revealed no difference in predictive timing between musically trained and untrained groups (e.g., Repp, 2010), whereas other studies indicated that musicians (percussionists) exhibited better entrainment performance when exposed to intense beat production activities (Cicchini et al., 2012). These studies suggest that individuals actively entrain with external rhythms and perceive past durations accordingly, and may provide evidence of the wide applicability of DAT.

Building upon simple click paradigms as previously discussed (Treisman et al., 1990), research in recent years used naturalistic stimuli, since DAT is most applicable in music and speech. Periodic tone entrainment studies yielded new results: Wearden et al. (2017) found the residual effect of the classic click train paradigm, that is, the higher the preceding click frequency, the longer the following duration would be perceived. They have also observed similar effects with irregular tones as well as white noise. This study revealed multiple approaches to activate and to speed up the internal clock. Periodic and aperiodic clicks, as well as rhythmic visual flickers and even white noise influenced results. In addition, the entrainment effect was also verified to transcend as long as 8s after hearing high-frequency clicks, indicating that the emission of attentional pulses has a latency between activation and cessation.

More complex stimuli such as music are processed similarly. Fast music compared to slow one was perceived to be longer due to the accumulation of more temporal units. A study adopted Mozart's Sonata for two pianos (K.448), where participants tended to overestimate the duration when the excerpt was at the "fast" (120BPM) end of the spectrum (Wang & Shi, 2019). The effect, nevertheless, is subject to the allocation of attention. Keller and Burnham (2005) emphasized the flexibility of attention when listening to musical meter, which could be composed of multiple metrical layers. Therefore, tracking high and low metrical structures is expected to have its corresponding effects on psychological time (cf. Hammerschmidt & Wöllner, 2020), as the former should hypothetically lead to fewer mental counts and thus time compression. Neurological evidence also indicated that focusing on different temporal structures led to alignments in steady state event potential (SS-EP) frequencies, deciphered from EEG recordings (Nozaradan et al., 2012). In this case, neural entrainment reflects that attending to local features in complex auditory signals could form mental representations of time by modulating the original neural oscillations.

When more attention is allocated to the temporal features of music, Cocenas-Silva et al. (2011) observed a time dilation effect. When participants were asked to group excerpts of various arousal levels based on their estimated lengths, those which were highly arousing tended to be overestimated. The finding is consistent with Droit-Volet et al.'s (2013) observation that faster music, which was thought to be more arousing, was judged to be longer than the slow, less arousing ones. We might reason that, when individuals attend to temporal features of the auditory signals, the temporal entrainment effect is stronger compared to situations when they attend to other features such as key chords and pleasantness.

Audiovisual Evidence for the Scalar Expectancy Theory

The following examines the evidence for multiple sensory modalities that either support or disagree with SET. To establish a solid ground for SET, researchers tried to find evidence for Weber's fraction, or a constant variance to subjective timing, across different sensory modalities, durations, populations, and other conditions. Wearden and Jones (2007) probed the scalar property of subjective timing using two variations of the duration comparison task with auditory tones ranging from 600ms to 10s. They found a linear increase in subjective timing that conforms to Weber's law. This effect is consistent also in the visual domain. In a duration discrimination study, Grondin (2001) found that participants exhibited similar sensitivity towards intervals marked by visual flickers between 600 to 900 ms, in accordance with Weber's law. However, the ratio changed when the inter-stimulus interval went beyond 900ms. The violation of Weber's law might be due to potentially explicit counting.

Similarly, mixed findings have been reported in multi-modalities studies. Hypothetically, if the scalar property holds across modalities, one should expect a consistent linear increase in different modalities. This was indeed the case when participants performed predictive saccades, or eye-movement timing, when intervals from 500 to 1000ms were presented either as visual flashes or auditory tone flutters (Joiner et al., 2007). However, comparing Weber's ratios between the two modalities revealed that auditory timing had greater variability than visual timing, as shown in participants' reactive eye movements when tracking the periodic cues. Hence, one might deduct that the scalar property holds but is also subject to stimulus modality. Block and Gruber (2014) argued that the obstacles of finding a cross-modal transfer effect was restricted to below the 3 to 5s window, beyond which the automatic processing should diminish due to the limited capacity of working memory.

On the other hand, evidence against the scalar property has been presented in auditory studies. Grondin (2012) adopted three approaches to measure Weber's ratio: duration discrimination, reproduction and categorization tasks on a spectrum from 1 to 1.9 seconds using pure tones. In all three tasks, Weber's ratio appeared to be higher when the intervals were longer regardless of the number of interval repetitions, in this case either 1, 3, or 5 times. These results indicate the inconsistency in Weber's ratio or temporal sensitivity despite different emphases of each paradigm on the timing process. Grondin (2010a) pointed out that the failure of conforming to the pacemaker-counter model, which SET is built upon, was because this model no longer applied to this duration range (see Figure 1). More specifically, a cut-off point at 1.2 to 1.3s was observed. This aligns with observations from other studies (for a review, see Matthews & Meck, 2014). The question is, how is time processed beyond that point? Some researchers proposed that a learning effect might have altered the variance, as the brain was influenced by multiple exposures to the same interval (Matthews & Grondin, 2012). Findings across timing tasks and sensory modalities, nevertheless, support the presence of a unitary clock system.

Despite the controversial evidence, reports investigating timing precision on multiple sensory modalities align with what the striatal beat frequency theory proposed: a familiarity effect that is reflected by enhanced synaptic communication between neurons. This might lead to higher processing efficiency and smaller variability compared to unfamiliar intervals. Grondin's (2012) experiments revealed that participants performed better in 3- and 5-interval discrimination than when only one interval was presented. Frequent exposure to timing tasks, as a part of music training, may also implicate the benefits of enhanced neural connection. In Rammsayer and Altenmüller's (2006) study, musicians outperformed non-musicians in a perceptual timing task in terms

of showing less variance and thus higher temporal sensitivity for instance in duration discrimination tasks. Musicians, however, did not exhibit significant superiority in a temporal generalization task, where participants compared the duration of an excerpt to the reference at the beginning, hypothetically stored in one's working memory. The authors believed that this was due to the fact that the intervals exceeded working memory capacities. This explanation is equally applicable to Grondin and Killeen's (2009) results, where participants in a reproduction-by-tapping task performed significantly better if they adopted counting or singing, compared to doing nothing. Thus it might be concluded that the SET indeed predicts the timing performance only within short intervals of no more than 2s (for a review, see Ivry & Schlerf, 2008). Nevertheless, it is equally important to understand timing within a few notes as well as in larger musical structures such as phrases.

Conclusions

This review has discussed two internal clock models: the intrinsic and the central clock models. The intrinsic model emphasizes automatic processing of temporal information in the subsecond range, while the central clock model explains the suprasecond (seconds to minutes) range of timing, which demands higher levels of cognitive control. Controversially, the Scalar Expectancy Theory, which can be seen as a specific account of the central clock model, applies to timing in the seconds range only, while the Dynamic Attending Theory works for timing intervals from seconds to minutes. According to SET and DAT, short intervals are represented linearly through the accumulation of pacemaker pulses, while longer intervals are represented nonlinearly, as pulse emission is calibrated to align with external periodicity. As for intervals of hours and longer, the timing process is subject to contextual changes and memory segmentation, and relevant research is scarce.

Audition, among all modalities, shows superiority in temporal processing by entailing higher sensitivity to detect changes and to estimate interval lengths compared to vision and other sensory modalities. In this sense, the modality specificity supports a distributed timing mechanism. Yet more evidence is needed to explain the cross-modal transfer of training effects in, for example, duration discrimination. Despite years of debate on the superiority of one clock model, there is no conclusive evidence to the best of our knowledge. We come to the observation that each model has its best fit at a different time duration scale, and as to whether discrete events (SET) or complex streams (DAT) such as in music are at the core of the investigation.

Regarding the explanatory power of the internal clock models for the perception of musical time, it is therefore necessary to consider an interval-specific approach. Short interval timing within the milliseconds range plays a crucial role in music production such as expressive microtiming, whereas long interval timing is more strongly modified by attention, emotion, and working memory, consequently adding more variables to the equation. In this regard, the timing paradigm adopted in an ecologically plausible environment such as music concerts, movies, or sports should receive more attention. Ways of applying clock models to longer-interval timing and time estimation are yet to be investigated.

Funding

PsychOpen GOLD

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

The authors have no support to report.

References

Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron, 66*(4), 610-618. https://doi.org/10.1016/j.neuron.2010.04.014

Allman, M. J., & Meck, W. H. (2012). Pathophysiological distortions in time perception and timed performance. *Brain, 135*(3), 656-677. https://doi.org/10.1093/brain/awr210

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. https://doi.org/10.1146/annurev-psych-010213-115117

Auhagen, W., & Busch, V. (1998). The influence of articulation on listeners' regulation of performed tempo. In R. Kopiez & W. Auhagen (Eds.), *Controlling creative processes in music* (pp. 69–92). Bern, Switzerland: Peter Lang.

Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive Psychology, 41*(3), 254-311. https://doi.org/10.1006/cogp.2000.0738

Barry, B. R. (1990). *Musical time: the sense of order*. Hillsdale, NY, USA: Pendragon Press.

Behne, K. E. (1976). „Zeitmaße": Zur Psychologie des musikalischen Tempoempfindens. *Die Musikforschung, 29*(2), 155-164.

Bisson, N., Tobin, S., & Grondin, S. (2008). Remembering the duration of joyful and sad musical excerpts: Assessment with three estimation methods. *NeuroQuantology, 7*(1), 46-57. https://doi.org/10.14704/nq.2009.7.1.206

Block, R. A., & Gruber, R. P. (2014). Time perception, attention, and memory: A selective review. *Acta Psychologica, 149*, 129-133. https://doi.org/10.1016/j.actpsy.2013.11.003

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review, 4*, 184-197. https://doi.org/10.3758/BF03209393

Boasson, A. D., & Granot, R. (2012). Melodic direction's effect on tapping. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiadis (Eds.), *Proceedings of 12th international conference on music perception and cognition, and the 8th triennial conference of the European society for the cognitive sciences of music* (pp.110-119). The joint conference ICMPC – ESCOM 2012, Thessaloniki, Greece. Retrieved from http://icmpc-escom2012.web.auth.gr/files/papers/110_Proc.pdf

Bolger, D., Trost, W., & Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychologica, 142*, 238-244. https://doi.org/10.1016/j.actpsy.2012.11.012

Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics, 60*(8), 1357-1373. https://doi.org/10.3758/BF03207998

Boltz, M. G. (2017). Auditory driving in cinematic art. *Music Perception, 35*(1), 77-93. https://doi.org/10.1525/mp.2017.35.1.77

Broadway, J. M., & Engle, R. W. (2011). Individual differences in working memory capacity and temporal discrimination. *PLOS ONE, 6*(10), Article e25422. https://doi.org/10.1371/journal.pone.0025422

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "ticktock" of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science, 14*(4), 362-366. https://doi.org/10.1111/1467-9280.24441

Brown, S. W., & Boltz, M. G. (2002). Attentional processes in time perception: Effects of mental workload and event structure. *Journal of Experimental Psychology: Human Perception and Performance, 28*(3), 600-615. https://doi.org/10.1037/0096-1523.28.3.600

Bueti, D., van Dongen, E. V., & Walsh, V. (2008). The role of superior temporal cortex in auditory timing. *PLOS ONE, 3*(6), Article e2481. https://doi.org/10.1371/journal.pone.0002481

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. https://doi.org/10.1038/nrn1764

Buonomano, D. V., Bramen, J., & Khodadadifar, M. (2009). Influence of the interstimulus interval on temporal processing and learning: Testing the state-dependent network model. *Philosophical Transactions of the Royal Society of London. Series B, 364*(1525), 1865-1873. https://doi.org/10.1098/rstb.2009.0019

Buonomano, D. V., & Laje, R. (2011). Population clocks: Motor timing with neural dynamics. In S. Dehaene, & E. Brannon (Eds.), *Space, time and number in the brain* (pp. 71-85). Cambridge, MA, USA: Academic Press. https://doi.org/10.1016/B978-0-12-385948-8.00006-2

Burger, B., London, J., Thompson, M. R., & Toiviainen, P. (2018). Synchronization to metrical levels in music depends on low-frequency spectral components and tempo. *Psychological Research, 82*(6), 1195-1211. https://doi.org/10.1007/s00426-017-0894-2

Burr, D., Della Rocca, E., & Morrone, M. C. (2013). Contextual effects in interval-duration judgements in vision, audition and touch. *Experimental Brain Research, 230*(1), 87-98. https://doi.org/10.1007/s00221-013-3632-z

Burr, D., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nature Neuroscience, 10*(4), 423-425. https://doi.org/10.1038/nn1874

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision, 13*(10), 1-19. https://doi.org/10.1167/13.10.1

Chen, L., Zhou, X., Müller, H. J., & Shi, Z. (2018). What you see depends on what you hear: Temporal averaging and crossmodal integration. *Journal of Experimental Psychology: General, 147*(12), 1851-1864. https://doi.org/10.1037/xge0000487

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, 32*(3), 1056-1060. https://doi.org/10.1523/JNEUROSCI.3411-11.2012

Cocenas-Silva, R., Bueno, J. L. O., Molin, P., & Bigand, E. (2011). Multidimensional scaling of musical time estimations. *Perceptual and Motor Skills, 112*(3), 737-748. https://doi.org/10.2466/11.24.PMS.112.3.737-748

Colflesh, G. J., & Conway, A. R. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin & Review, 14*(4), 699-703. https://doi.org/10.3758/BF03196824

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience, 33*(9), 4002-4010. https://doi.org/10.1523/JNEUROSCI.4675-12.2013

Droit-Volet, S., & Berthon, M. (2017). Emotion and implicit timing: The arousal effect. *Frontiers in Psychology, 8*, Article 176. https://doi.org/10.3389/fpsyg.2017.00176

Droit-Volet, S., Bigand, E., Ramos, D., & Bueno, J. L. O. (2010). Time flies with music whatever its emotional valence. *Acta Psychologica, 135*, 226-232. https://doi.org/10.1016/j.actpsy.2010.07.003

Droit-Volet, S., Fayolle, S. L., & Gil, S. (2011). Emotion and time perception: Effects of film-induced mood. *Frontiers in Integrative Neuroscience, 5*, Article 33. https://doi.org/10.3389/fnint.2011.00033

Droit-Volet, S., Fayolle, S., Lamotte, M., & Gil, S. (2013). Time, emotion and the embodiment of timing. *Timing & Time Perception, 1*, 99-126. https://doi.org/10.1163/22134468-00002004

Eagleman, D. M., Peter, U. T., Buonomano, D., Janssen, P., Nobre, A. C., & Holcombe, A. O. (2005). Time and the brain: How subjective time relates to neural time. *The Journal of Neuroscience, 25*(45), 10369-10371. https://doi.org/10.1523/JNEUROSCI.3487-05.2005

Eerola, T., Luck, G., & Toiviainen, P. (2006). An investigation of pre-schoolers' corporeal synchronization with music. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the 9th international conference on music perception and cognition* (pp. 472-476). The Society for Music Perception and Cognition and European Society for the Cognitive Sciences of Music Bologna, Bologna, Italy.

Ellis, M. C. (1991). Research note: Thresholds for detecting tempo change. *Psychology of Music, 19*(2), 164-169. https://doi.org/10.1177/0305735691192007

Escoffier, N., Sheng, D. Y. J., & Schirmer, A. (2010). Unattended musical beats enhance visual processing. *Acta Psychologica, 135*, 12-16. https://doi.org/10.1016/j.actpsy.2010.04.005

Gibbon, J. (1977). Scalar Expectancy Theory and Weber's law in animal timing. *Psychological Review, 84*(3), 279-325. https://doi.org/10.1037/0033-295X.84.3.279

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences, 423*, 52-77. https://doi.org/10.1111/j.1749-6632.1984.tb23417.x

Goudriaan, J. C. (1921). Le rythme psychique dans ses rapports avec les fréquences cardiaque et respiratoire. *Archives Néerllandaises de Physiologie, 6*, 77-110.

PsychOpen GOLD

Grondin, S. (2001). Discriminating time intervals presented in sequences marked by visual signals. *Perception & Psychophysics, 63*(7), 1214-1228. https://doi.org/10.3758/BF03194535

Grondin, S. (2010a). Unequal Weber fractions for the categorization of brief temporal intervals. *Attention, Perception & Psychophysics, 72*(5), 1422-1430. https://doi.org/10.3758/APP.72.5.1422

Grondin, S. (2010b). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics, 72*(3), 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. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 880-890. https://doi.org/10.1037/a0027188

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. https://doi.org/10.3758/APP.71.7.1649

Gu, B. M., van Rijn, H., & Meck, W. H. (2015). Oscillatory multiplexing of neural population codes for interval timing and working memory. *Neuroscience and Biobehavioral Reviews, 48*, 160-185. https://doi.org/10.1016/j.neubiorev.2014.10.008

Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science, 16*(3), 228-235. https://doi.org/10.1111/j.0956-7976.2005.00808.x

Hammerschmidt, D., & Wöllner, C. (2020). Sensorimotor synchronization with higher metrical levels in music shortens perceived time. *Music Perception, 37*(4), 263-277. https://doi.org/10.1525/mp.2020.37.4.263

Henry, M. J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences of the United States of America, 111*(41), 14935-14940. https://doi.org/10.1073/pnas.1408741111

Herbert, R. (2012). *Everyday music listening: Absorption, dissociation and trancing*. Abingdon, United Kingdom: Routledge.

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences, 12*(7), 273-280. https://doi.org/10.1016/j.tics.2008.04.002

Joiner, W. M., Lee, J. E., Lasker, A., & Shelhamer, M. (2007). An internal clock for predictive saccades is established identically by auditory or visual information. *Vision Research, 47*(12), 1645-1654. https://doi.org/10.1016/j.visres.2007.02.013

Jones, M. R. (1981). Music as a stimulus for psychological motion: Part I. Some determinants of expectancies. *Psychomusicology: Music, Mind, and Brain, 1*(2), 34-51. https://doi.org/10.1037/h0094282

Jones, M. R. (1990). Musical events and models of musical time. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 207-240). NJ, USA: Lawrence Erlbaum Associates.

Jones, M. R. (2010). Attending to sound patterns and the role of entrainment. In A. C., Nobre, & J. T., Coull (Eds.), *Attention and time* (pp. 317-330). Oxford, United Kingdom: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199563456.003.0023

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*(3), 459-491. https://doi.org/10.1037/0033-295X.96.3.459

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: Encoding time in neural network states. *Neuron, 53*(3), 427-438. https://doi.org/10.1016/j.neuron.2007.01.006

Keller, P. E., & Burnham, D. K. (2005). Musical meter in attention to multipart rhythm. *Music Perception, 22*(4), 629-661. https://doi.org/10.1525/mp.2005.22.4.629

Langner, J. (2002). *Musikalischer Rhythmus und Oszillation* (Vol. 13). Bern, Switzerland: Peter Lang Publishing.

Large, E. W. (2008). *Resonating to musical rhythm: Theory and experiment*. In S. Grondin (Ed.), *Psychology of time* (pp. 189-232). Bingley, United Kingdom: Emerald Group Publishing. https://doi.org/10.1016/B978-0-08046-977-5.00006-5

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119-159. https://doi.org/10.1037/0033-295X.106.1.119

Lieb, E. H., & Yngvason, J. (1999). The physics and mathematics of the second law of thermodynamics. *Physics Reports, 310*, 1-96. https://doi.org/10.1016/S0370-1573(98)00082-9

London, J. (2004). *Hearing in time: Psychological aspects of musical meter*. Oxford, United Kingdom: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195160819.001.0001

Lustig, C., & Meck, W. H. (2011). Modality differences in timing and temporal memory throughout the lifespan. *Brain and Cognition, 77*(2), 298-303. https://doi.org/10.1016/j.bandc.2011.07.007

Matell, M. S., Meck, W. H., & Nicolelis, M. A. (2003). Interval timing and the encoding of signal duration by ensembles of cortical and striatal neurons. *Behavioral Neuroscience, 117*(4), 760-773. https://doi.org/10.1037/0735-7044.117.4.760

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. https://doi.org/10.3758/s13414-012-0282-3

Matthews, W. J., & Meck, W. H. (2014). Time perception: The bad news and the good. *Cognitive Science, 5*(4), 429-446. https://doi.org/10.1002/wcs.1298

McDermott, J. H., Wrobleski, D., & Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proceedings of the National Academy of Sciences of the United States of America, 108*(3), 1188-1193. https://doi.org/10.1073/pnas.1004765108

McPherson, T., Berger, D., Alagapan, S., & Fröhlich, F. (2018). Intrinsic rhythmicity predicts synchronization-continuation entrainment performance. *Scientific Reports, 8*(1), Article 11782. https://doi.org/10.1038/s41598-018-29267-z

Meck, W. H. (1984). Attentional bias between modalities: Effect on the internal clock, memory, and decision stages used in animal time discrimination. *Annals of the New York Academy of Sciences, 423*(1), 528-541. https://doi.org/10.1111/j.1749-6632.1984.tb23457.x

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annual Review of Neuroscience, 36*, 313-336. https://doi.org/10.1146/annurev-neuro-062012-170349

Miall, C. (1989). The storage of time intervals using oscillating neurons. *Neural Computation, 1*, 359-371. https://doi.org/10.1162/neco.1989.1.3.359

Nagarajan, S. S., Blake, D. T., Wright, B. A., Byl, N., & Merzenich, M. M. (1998). Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 18*(4), 1559-1570. https://doi.org/10.1523/JNEUROSCI.18-04-01559.1998

Nichols, R. (2011). *Ravel*. London, United Kingdom: Yale University Press. https://doi.org/10.1093/ml/gcs082

Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical Transactions of the Royal Society of London. Series B, 369*, Article 20130393. https://doi.org/10.1098/rstb.2013.0393

Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *The Journal of Neuroscience, 32*(49), 17572-17581. https://doi.org/10.1523/JNEUROSCI.3203-12.2012

Ornstein, R. E. (1969). *On the experience of time*. London, United Kingdom: Penguin Publisher.

Ortega, L., & López, F. (2008). Effects of visual flicker on subjective time in a temporal bisection task. *Behavioural Processes, 78*(3), 380-386. https://doi.org/10.1016/j.beproc.2008.02.004

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. https://doi.org/10.1037/0096-1523.26.6.1770

Phillips, D. P., & Hall, S. E. (2002). Auditory temporal gap detection for noise markers with partially overlapping and non-overlapping spectra. *Hearing Research, 174*(1-2), 133-141. https://doi.org/10.1016/S0378-5955(02)00647-0

Polti, I., Martin, B., & van Wassenhove, V. (2018). The effect of attention and working memory on the estimation of elapsed time. *Scientific Reports, 8*(1), Article 6690. https://doi.org/10.1038/s41598-018-25119-y

Pöppel, E. (1989). The measurement of music and the cerebral clock: A new theory. *Leonardo, 22*, 83-89. https://doi.org/10.2307/1575145

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*(4), 411-440. https://doi.org/10.2307/40285311

Pressnitzer, D., Suied, C., & Shamma, S. (2011). Auditory scene analysis: The sweet music of ambiguity. *Frontiers in Human Neuroscience, 5*, Article 158. https://doi.org/10.3389/fnhum.2011.00158

Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception, 24*(1), 37-48. https://doi.org/10.1525/mp.2006.24.1.37

Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: Effects of music training and task experience. *Human Movement Science, 29*(2), 200-213. https://doi.org/10.1016/j.humov.2009.08.002

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance, 28*(5), 1085-1099. https://doi.org/10.1037/0096-1523.28.5.1085

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review, 20*(3), 403-452. https://doi.org/10.3758/s13423-012-0371-2

Schäfer, T., Fachner, J., & Smukalla, M. (2013). Changes in the representation of space and time while listening to music. *Frontiers in Psychology, 4*, Article 508. https://doi.org/10.3389/fpsyg.2013.00508

Schnupp, J. W., Hall, T. M., Kokelaar, R. F., & Ahmed, B. (2006). Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *The Journal of Neuroscience, 26*(18), 4785-4795. https://doi.org/10.1523/JNEUROSCI.4330-05.2006

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences, 32*(1), 9-18. https://doi.org/10.1016/j.tins.2008.09.012

Schulze, H. H. (1978). The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research, 40*(2), 173-181. https://doi.org/10.1007/BF00308412

Schwartze, M., Keller, P. E., Patel, A. D., & Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behavioural Brain Research, 216*(2), 685-691. https://doi.org/10.1016/j.bbr.2010.09.015

Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology, 20*(3), 361-366. https://doi.org/10.1016/j.conb.2010.03.009

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Research: Cognitive Brain Research, 14*(1), 147-152. https://doi.org/10.1016/S0926-6410(02)00069-1

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science, 145*(3638), 1328-1330. https://doi.org/10.1126/science.145.3638.1328

Stern, L. W. (1897). Psychische Präsenzzeit. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 13*, 325-349.

Teki, S., & Griffiths, T. D. (2014). Working memory for time intervals in auditory rhythmic sequences. *Frontiers in Psychology, 5*, Article 1329. https://doi.org/10.3389/fpsyg.2014.01329

Tobin, S., Bisson, N., & Grondin, S. (2010). An ecological approach to prospective and retrospective timing of long durations: A study involving gamers. *PLOS ONE, 5*(2), Article e9271. https://doi.org/10.1371/journal.pone.0009271

Toiviainen, P., Burunat, I., Brattico, E., Vuust, P., & Alluri, V. (2019). The chronnectome of musical beat. *NeuroImage,* Article 116191. Advance online publication. https://doi.org/10.1016/j.neuroimage.2019.116191

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the" internal clock. *Psychological Monographs, 77*(13), 1-31. https://doi.org/10.1037/h0093864

Treisman, M., & Brogan, D. (1992). Time perception and the internal clock: Effects of visual flicker on the temporal oscillator. *The European Journal of Cognitive Psychology, 4*(1), 41-70. https://doi.org/10.1080/09541449208406242

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), 705-742. https://doi.org/10.1068/p190705

PsychOpen GOLD

van Rijn, H., Gu, B. M., & Meck, W. H. (2014). Dedicated clock/timing-circuit theories of time perception and timed performance. In H. Merchant & V. D. Lafuente (Eds.), *Neurobiology of interval timing* (pp. 75-99). Berlin, Germany: Springer. https://doi.org/10.1007/978-1-4939-1782-2_5

van Wassenhove, V., Buonomano, D. V., Shimojo, S., & Shams, L. (2008). Distortions of subjective time perception within and across senses. *PLOS ONE, 3*(1), Article e1437. https://doi.org/10.1371/journal.pone.0001437

Wang, X., & Shi, Z. (2019, September 10-12). *Temporal entrainment effect: Can music enhance our attention resolution in time?* [Poster presentation]. The 12th International Conference of Students of Systematic Musicology. SysMus, Berlin, Germany.

Wang, X., Wöllner, C., & Shi, Z. (2019, September 6-8). *Perceiving tempo in incongruent audiovisual contexts: An exploratory study with a temporal bisection paradigm* [Poster presentation]. Jahrestagung der Deutsche Gesellschaft für Musikpsychologie, Eichstätt, Germany.

Wearden, J. H., & Jones, L. A. (2007). Is the growth of subjective time in humans a linear or nonlinear function of real time? *Quarterly Journal of Experimental Psychology, 60*(9), 1289-1302. https://doi.org/10.1080/17470210600971576

Wearden, J. H., Williams, E. A., & Jones, L. A. (2017). What speeds up the internal clock? Effects of clicks and flicker on duration judgements and reaction time. *Quarterly Journal of Experimental Psychology, 70*(3), 488-503. https://doi.org/10.1080/17470218.2015.1135971

Wittmann, M. (2013). The inner sense of time: How the brain creates a representation of duration. *Nature Reviews. Neuroscience, 14*(3), 217-223. https://doi.org/10.1038/nrn3452

Wöllner, C., & Halpern, A. R. (2016). Attentional flexibility and memory capacity in conductors and pianists. *Attention, Perception & Psychophysics, 78*(1), 198-208. https://doi.org/10.3758/s13414-015-0989-z

Wöllner, C., Hammerschmidt, D., & Albrecht, H. (2018). Slow motion in films and video clips: Music influences perceived duration and emotion, autonomic physiological activation and pupillary responses. *PLOS ONE, 13*(6), Article e0199161. https://doi.org/10.1371/journal.pone.0199161

Zakay, D., & Block, R. A. (1995). An attentional-gate model of prospective time estimation. *Time and the Dynamic Control of Behavior,* 167-178.

Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences of the United States of America, 107*(13), 5768-5773. https://doi.org/10.1073/pnas.1000121107