SISSA
Scuola Internazionale Superiore di Studi Avanzati

TIME PERCEPTION LAB

Neuroscience Area

PhD course in Cognitive Neuroscience

The processing and learning of temporal patterns: from behaviour to computational models

Candidate:
Dunia Giomo

Advisor:
Prof. Domenica Bueti

Academic year 2021-22



Table of Contents

Chapter 1: Introduction 5

Chapter 2: Global and local deviance effects in the processing of temporal sequences 11

Chapter 3: Structure transition in temporal pattern learning 32

Chapter 4: Learning the statistical properties of temporal patterns 52

Chapter 5: General Discussion 81

Appendix A: Supplementary Material chapter 2 88

Appendix B: Supplementary Material chapter 3 92

Appendix C: Supplementary Material chapter 4 99

References 104

General abstract

Being able to successfully and adaptively interact with our environment often requires us to process and learn various forms of temporal regularities, involving different types of temporal features and structures. In many cases the presence of such regularities makes us experience a series of events as a temporal pattern. We often identify a series of events as a pattern based on its temporal properties, like the duration of the events, their order and temporal relationship.

Temporal patterns feature a multitude of human sensorimotor experiences: musical rhythms, predictions and expectations, complex motor actions, speech perception and productions. Likewise, the interest in temporal patterns and their investigation have involved several research fields and perspectives. Studies from all these fields have shown humans' remarkable abilities in perceiving and learning different types of temporal regularities and patterns, in using them to build expectations, orient attentional resources and organize motor actions. Despite the growing body of knowledge on the basic mechanisms supporting these abilities, there are still many unanswered questions concerning how humans process and learn temporal patterns. A small portion of these open questions is the focus of the current PhD thesis. Given the broad nature of the topic, the current work adopts a "crossroad" perspective that combines theoretical and analytical tools from timing and sequence processing research.

The first part of the work (chapter 2) focuses on the processing of different levels of regularity of a temporal pattern. Specifically, we asked how the reproduction of a temporal pattern is affected by a violation within its elements (i.e. local violation) and in its overall structure (i.e. global violation). To answer this question, we adapted a paradigm originally developed in sequence processing studies to the temporal domain, using a Synchronization-Continuation task. This design was employed in three behavioral experiments, in which we tested the effects of structure complexity and sensory modality on the behavioral signatures associated with local and global violations. Analysis of behavioral data was carried in the frequency domain to obtain a measure of tapping precision, and with Procrustes technique to investigate the internal consistency of the tapping. Results from frequency-domain analysis showed that local violations of a temporal pattern always affected the precision of its reproduction, whereas global violations were detrimental when using visual and complex patterns. Analysis of participants' performances in the time-domain using Procrustes technique showed that in all three experiments the detrimental effects observed in the frequency-domain were partially explained by a systematic tendency of participants to reproduce rescaled versions of the sequences. This last result suggests that the accuracy in reproducing a temporal pattern, does not reflect the ability to acquire and reproduce its global structure.

The second part of the work focuses on the learning of temporal patterns, addressing the topic from two different perspectives. In the first study (chapter 3) we asked how humans learn to switch from a familiar to a novel temporal structure and how this switch is modulated by trial predictability.

This question was tackled employing a novel experimental design in which we parametrically changed the probability of occurrence of two mirroring patterns across trials. This design was applied to a synchronization task presenting the patterns within an auditory stream. Analysis of behavioral performance combined standard tapping measures and a model-based approach. The general purpose of the model-based analysis was to describe and estimate the parameters underlying the learning transition between the two temporal structures, as expressed in subjects' tapping. Results showed that the models implemented were able to retrieve the probabilities in which the two patterns were reproduced by our participants, reflecting the learning transition they experienced during the task. The combination of standard measures and models' results allowed to identify different types of performance and learning strategies among participants. A group of "fast learners" learned to synchronize with the first unknown pattern almost immediately and showed higher sensitivity to the switch in pattern across trials. On the contrary the "slow learners" took more time to both learn to synchronize with a new pattern and switch to a different one.

In the second study (chapter 4) we asked whether and how humans extract and learn low- and high-level statistical properties underlying non-deterministic sequences of temporal intervals. To this aim we designed three generative Markov processes, each defined by a set of statistical properties. Based on these processes we created multiple sequences and presented them to our participants in a delayed reproduction task. Model-based and data-driven analysis were applied to behavioral performances to investigate the statistical properties learned by our participants. Overall results showed that participants were able to learn the generative processes of the sequences they experienced, but also that additional statistics interacted with those designed in our stimuli, namely a central tendency effect and a shrinkage of the duration categories' space.

The work of this thesis gives insights on important aspects concerning the processing and learning of temporal patterns. A first insight is that the reproduction of a temporal pattern seems to require an interplay between individual constitutive durations and overall structure. These two levels are differently weighted according to the complexity of the pattern and the sensory modality in which it is experienced. The extent to which the weights of individual durations and overall structure depend on the "status" of the pattern (rhythmic or aperiodic) and the type of temporal task is not fully clear. Another insight comes from the role of the predictability of the environment in shaping the learning of temporal patterns. The level of predictability not only influences the ability to learn new temporal structures, but also interacts with subject-specific strategies employed to perform the task. When learning the statistical properties of non-deterministic sequences of durations, instead, the level of predictability modulates the interaction between statistics of the stimuli and low- and high-level perceptual priors.

Chapter 1
Introduction

Listening to music, dancing, playing sports, producing and perceiving speech, orienting our attention in the visual and acoustic environment are all examples of sensory motor activities requiring the processing of milliseconds and seconds time. Humans are continuously engaged in such perceptual and sensorimotor experiences that are structured in time and feature multiple levels of temporal regularity. These series of sensorimotor events are often experienced as temporal patterns, i.e. they are identified as a pattern based on their temporal structure.

The ability to detect and learn temporal regularities and patterns, use them to organize our motor actions, as well as adapt to them and build expectations about them is essential for successfully interact with our environment. Given the widespread involvement of these abilities in many cognitive and sensorimotor processes, it is no surprise that they have been investigated from several research perspectives: sensorimotor synchronization studies (Merchant et al., 2015), rhythm perception and production (Grahn, 2012), temporal expectations and temporal orienting of attention (Coull & Nobre, 2008; Nobre & Van Ede, 2018), auditory scene perception and sequence processing (Dehaene, 2015; Helibron & Chait, 2018), spatiotemporal sequence learning (e.g. Heideman et al., 2018).

Just as the quest for the basic mechanisms supporting the processing and learning of temporal patterns involves multiple research fields, also the work presented in this thesis stems from a contamination of different approaches and perspectives on the topic. In the next two paragraphs, however, only the most significant background literature will be briefly outlined, with a major focus on the investigation of temporal patterns in timing research.

1.1 Temporal patterns in timing research

A general assumption in timing research is that processing a single interval and processing a temporal pattern are two very different temporal computations for the human brain, at least in the subsecond range (for a review covering both, see Paton & Buonomano, 2018). However, even among temporal patterns a distinction is usually made between aperiodic and periodic patterns (Hardy & Buonomano, 2016). The main difference lies in the cyclical nature of periodic patterns: they are series of highly regular repeating events with fixed base frequency and phase, i.e. their main feature is the periodicity (Kotz et al., 2018; Rimmele et al., 2018).

Periodic patterns are often interchangeably called rhythms, even though not all rhythmic patterns are strictly periodic, despite their repetitive structure (Rimmele et al, 2018). In rhythm perception literature, rhythms are generally defined as regular series of durations perceptually grouped together, characterized by an explicit or underlying beat (also called “pulse” or “tactus”), i.e. a

perceived regular pulse in the rhythmic sequence. Series of beats are temporally organized into higher order regularities forming a hierarchical structure, or temporal framework, called "meter" (London, 2004; McAuley, 2010; Grahn, 2012; Honing, 2013). The extent to which a rhythm is periodic and its beat perceptually clear gives a measure of its complexity (Povel & Essens, 1985; Grahn & Brett, 2007). In fact, the simplest forms of rhythmic patterns are also the most periodic: isochronous rhythms, i.e. patterns made of only one repeating duration. Isochronous rhythms have particular biological relevance, as they feature many animal physiological and sensorimotor processes (e.g., breathing, walking, chewing). Highly regular patterns are characteristic of music, several motor behaviors and, with more flexibility, speech perception and production. Aperiodic temporal patterns can be found in complex motor or spatiotemporal sequences, some forms of music, environmental acoustic stimuli and other non-rhythmic series of sensorimotor events (Hardy & Buonomano, 2016; Nobre & Van Ede, 2018; Kotz et al., 2018; Rimmele et al., 2018).

Beside these terminological specifications, the feature of periodicity is at the base of an operational and theoretical distinction commonly acknowledged in timing models. According to this distinction, two separate mechanisms, or strategies, are employed to process durations within periodic patterns and process single intervals or durations within aperiodic patterns. In the first type of processing (relative or *beat-based* timing) the duration is encoded relative to an underlying temporal reference, i.e. the beat, whereas in the second type (absolute, interval- or *duration-based* timing) the duration is encoded discretely, as an absolute quantity, whether it is part of a pattern or not (McAuley & Jones, 2003; Teki et al., 2011, 2012; Leow & Grahn, 2014). These two strategies are thought to support different types of temporal judgments, temporal predictions and attentional processes (Large & Jones, 1999; Nobre & van Ede, 2018), as well as motor timing tasks (Merchant & Honing, 2014). In particular, beat-based mechanisms are generally considered to be recruited for rhythm perception and production, rhythm-based predictions and motor synchronization with rhythmic stimuli (Grahn, 2012; Merchant et al., 2015).

Whereas beat-based and duration-based mechanisms are conceptually distinct, several neuroimaging and lesion studies have demonstrated an extensive overlap in the brain regions supporting them, involving a wide network of cortical and subcortical regions (premotor, supplementary motor, temporal, inferior parietal, insular and prefrontal cortices, basal ganglia and cerebellum). The same studies have also shown a functional specialization or a major engagement of the olivo-cerebellar system for interval-based timing and of the cortico-striatal circuit for beat-based timing (Grahn & Brett, 2009; Grube et al., 2010a,b; Teki et al., 2011; Grahn, 2012; Merchant et al., 2015; Breska & Ivry, 2018). Over the years, however, the entrainment of neuronal oscillatory processes has been proposed and widely accepted as the underlying neurophysiological signature of beat-based timing (Grahn, 2012; Merchant et al., 2015). Despite its popularity in the field, the dichotomy beat vs interval timing has been challenged by recent behavioral and electrophysiology works on pattern-based temporal predictions (Morillon et al., 2016; Breska & Dowell, 2016; Bouwer et al., 2020). These studies have questioned the putative separation between the neural and behavioral signatures associated with beat-based and interval-based predictions (also referred as memory-based), revealing that they are more similar than what was previously thought.

The distinction between periodic and aperiodic patterns is of great relevance in the motor domain, as several human (and animal) motor behaviors are temporally organized and often rhythmically structured. In this regard, beat-based mechanisms have been traditionally considered at the basis of humans' ability to coordinate motor actions with periodic stimuli, a complex motor behavior known as sensorimotor synchronization (SMS, for a comprehensive review see Repp, 2005, and Repp & Su, 2013). More elementary forms of synchronization and motor entrainment have also been found in other animal species (Kotz et al., 2018; Henry et al., 2021), in particular in non-human primates (Merchant et al., 2011; Bartolo et al., 2014; Cadena-Valencia et al., 2018; Gámez et al., 2018, 2019).

In general, SMS studies have shown that the ability to synchronize body movements with an external rhythmic source is a spontaneous and highly developed skill in humans. Human SMS is predictive (i.e. the movement tends to anticipate the stimulus), is subject to error correction mechanisms (i.e. adjustment of the movement to minimize temporal mismatches with the stimuli) and reaches its maximal accuracy when coordinating with auditory isochronous patterns. SMS is typically studied with tasks that require to tap the finger in synchrony with an external pattern and later continue the tapping in the absence of it. These two types of coordination have been shown to differentially recruit a large network of areas that comprises both interval timing and beat perception substrates (Repp & Su, 2013; Iversen & Balasubramanian, 2016). It has also been demonstrated that the activity of these areas is modulated by the temporal complexity of the external stimulus and the sensory modality involved (Lewis et al., 2004; Vuust et al., 2006; Chen et al., 2008; Pollok et al, 2009). Concerning pattern's complexity, some SMS and rhythm production studies have focused on either the synchronization with or the production of aperiodic and complex patterns (i.e. rhythms with complex or absent metrical structure), in contrast with simple or isochronous rhythms (Large et al., 2002; Lewis et al., 2004; Vuust et al., 2006; Grahn & Brett, 2007; Chen et al., 2008). Overall these studies showed that coordinating with or reproducing complex patterns is more difficult compared to simple rhythms, as would be expected, and can be influenced by the level of musical expertise. It is interesting, though, that across studies metrically complex and irregular (non-metric) patterns were performed with comparable levels of accuracy, suggesting the absence of qualitative differences among the two types, at least in their reproduction.

A less mainstream perspective on the topic comes from works on aperiodic motor pattern learning (Laje et al., 2011; Hardy & Buonomano, 2018), combining psychophysics and computational models. These studies showed that aperiodic patterns can be learned and reproduced with extreme accuracy by human participants and are characterized by specific behavioral signatures that differentiate them from single interval reproduction. Namely, reproduction variability is best explained by a "continuous" fit model (i.e. increase of the variance in proportion with the total duration of the pattern, not with each interval); and learning effects result in the ability to produce rescaled versions of the pattern. These findings are well accounted for by a population clock model of temporal processing (Karmarkar & Buonomano, 2007), according to which durations are

encoded implicitly in the pattern of activity of neural circuits (i.e. neither in an absolute nor in a beat-based way).

Concerning other computational approaches, a group of studies have applied Bayesian inference to describe rhythm perception and production. Some of these studies used predictive coding and Bayesian inference to describe the dynamic tracking of rhythmic patterns (e.g. Vuust & Witek, 2014), or the perceptual and motor entrainment to rhythmic stimuli (Canon, 2021). Whereas other studies have focused more on humans' preference toward specific temporal features of rhythms and how they link to the perception and production of rhythmic patterns (e.g., Sadakata et al., 2005).

1.2 Temporal regularities in sequence processing

A great interest in temporal regularities in a broader sense has historically come from sequence processing studies. In this context the temporal regularities being investigated can span multiple levels of a sequence's structure, can involve qualitatively different temporal features (i.e. timing, serial order), be more or less deterministic, and combine with other forms of regularities (Dehaene et al., 2015). These studies have generally demonstrated that, whenever we are exposed to a series of stimuli characterized by some form of temporal regularity, our brain quickly exploits and builds expectations about such regularity, and reacts to its violations or omissions. In particular, studies that fall within the general framework of predictive coding have focused on the neural responses associated with the detection and the violation of different hierarchical levels of sequences' regularities (for a review see Helibron & Chait, 2018). For example, a series of studies on auditory scene analysis (Andreou et al., 2015; Barascud et al., 2016) have shown specific neural signatures of unexpected omissions in ongoing temporally structured sound sequences (latency of offset responses) and signatures of the detection of a temporal regularity arising from randomness (increase in the sustained amplitude of evoked responses). Another noteworthy example comes from an electrophysiology study by Gavornik and Bear (2014), who trained mice to learn visual temporal sequences which varied only in the durations and order of their elements. The authors demonstrated that neural populations in the primary visual cortex showed an evoked response potentiation that was specific for stimulus order and timing, as well as a predictive response when elements of the learned sequence were omitted. These studies highlight how fast the brain detects and learn temporal regularities, and how early in the sensory processing hierarchy these regularities are encoded.

A closely related perspective on the processing of temporal regularities comes from probabilistic approaches to sequence learning. Indeed, many temporal structures are not fully deterministic and often come with some degree of uncertainty, but nevertheless are characterized by statistical properties that can be learned. Probabilistic approaches have been successful in accounting for the learning of simple statistical rules in sequences (Frost et al., 2015) either made of simple auditory or visual stimuli (Creel et al., 2004; Kirkham et al., 2002), musical notes (e.g. Pearce et al., 2010),

or phonetic/linguistic stimuli (e.g. Saffran et al., 1996). Studies that have combined this framework with Bayesian inference-based modelling (e.g., Maheu et al., 2019) have shown that brain activity and behavioral responses reflect the dynamical extraction of multiple levels of statistical regularities within sequences of stimuli. And that these processes can be well described in terms of Bayesian inference. Overall, sequence processing studies can offer powerful experimental paradigms and analytical tools to investigate how humans process and learn temporal patterns per se, whether they are fully predictable or not, whether they have simple or complex structure.

1.3 General aim and research questions of the current work

The general focus of the current thesis is on how humans process and learn temporal patterns. This broad topic is addressed using a "crossroad" perspective: investigating questions arising from timing research on temporal patterns by using a toolkit that combines theoretical frameworks, experimental paradigms, tasks and analytical approaches from sequence processing and motor timing studies. In this sense one of the aims of the work presented here is to offer novel combinations of perspectives and tools to investigate temporal patterns.

Chapter 2. The aim of the series of studies presented in the first chapter concerns the description of temporal pattern processing, beside the distinction between beat- and interval-based timing strategies. Specifically, we asked whether the processing of a temporal pattern encompasses all its features (i.e. each single interval, the order of the intervals and their overall structure) or only part of them. Adopting an experimental paradigm developed in the realm of predictive coding studies on sequence processing (Bekinschtein et a., 2009), we investigated how the reproduction of a temporal pattern is affected by different levels of violation of its regularity, i.e. a violation within its elements, or in its overall structure. The paradigm was employed in combination with a Synchronization-Continuation task in three behavioral experiments, in which we tested whether the behavioral signatures of each level's violation would change if the patterns are experienced in the visual or in the auditory domain, and whether their structure is simple or complex. Behavioral data were analyzed in the frequency domain to obtain a measure of tapping performances' precision, and with a shape mapping technique to explore the internal consistency of the tapping series. This analytical approach allowed us to simultaneously take into account the reproduction signatures of the single elements and the overall structures of the patterns performed.

Chapter 3. The work presented in this chapter focuses on the learning of temporal patterns. The leading research question here is how a new temporal structure is discovered and learned. In particular, we asked how humans learn to switch from a known to a novel temporal pattern, which is made by the same elements but has a different structure, and how this switch is modulated by trial predictability. To address these questions, we developed a novel experimental design in which we parameterized a gradual transition between two mirroring temporal patterns, by manipulating their probability of occurrence trial by trial. This design was applied to a synchronization task in which the patterns were embedded in auditory streams presented to our participants. Behavioral

performances were analyzed using standard tapping measures and a model-based approach. The general purpose of our modelling approach was to describe and estimate the parameters underlying the learning transition between the two temporal structures, as expressed in subjects' performance. To this extent we devised two models that give a trial- and tap-wise estimation of the proportions of the two patterns in participants' tapping series, as well as a formal description and estimation of the learning parameters governing the transition from one pattern to the other. These results represent the first stage of an ongoing work on model optimization, which will include a dynamic and more precise estimation of the learning parameters and the performance variability.

Chapter 4. The last experimental chapter also focuses on the learning of temporal patterns, but the topic is addressed from a different perspective. In this last chapter we adopted a probabilistic approach and asked whether and how humans extract and learn low- and high-level statistical properties underlying non-deterministic sequences of temporal intervals. To this aim we defined three generative Markov processes based on which multiple sequences were created and presented to our participants in a delayed reproduction task. The detection and learning of the underlying statistical regularities, as expressed in behavioral performance, was modelled as an optimal Bayesian inference process, and, in parallel, analyzed with a data-driven approach. The combination of the two analytical approaches allowed to fully describe the generative processes learned by our participants.

Chapter 5 presents a summary of all experimental findings, as well as a general discussion concerning the main insights these studies bring with respect to the current literature on temporal patterns.

Chapter 2

Global and local deviance effects in the processing of temporal sequences

ABSTRACT

Perceptual and sensorimotor events are often endowed with a temporal structure and experienced as temporal patterns, i.e. series of co-occurring sensory or motor events identified as sequences based on their temporal features.

Previous research on sensorimotor learning, rhythm perception and sequence processing has shown that humans are able to process, learn and form expectations on temporal regularities embedded in external stimuli. In addition to this, theoretical accounts from time perception literature have been proposed and tested to dissociate the underlying mechanisms supporting the processing of single intervals and the processing of temporal patterns. What is still unclear is whether and how humans encode and represent all the features that characterize a temporal pattern (e.g., the single durations it is made of, their order, the overarching structure connecting them). In other words, what remains unclear is whether the processing of a temporal pattern involves an additive processing of its single elements, or a holistic representation of its structure, or both of them. To narrow these general questions down, we focused on investigating how local changes within the elements of a temporal sequence or changes in its overall structure differently affect its reproduction. And how these behavioral effects vary when reproducing either simple or complex patterns, marked by either auditory or visual stimuli.

To this aim we employed the local-global paradigm (Bekinschtein et al., 2009), originally developed in auditory neuroscience research to differentiate the neural signatures associated to the detection of different levels of sequence violations, namely a violation within its elements ("local deviance") or in its overall structure ("global deviance"). We adapted this paradigm to the temporal domain in three separate experiments, using a Synchronization-Continuation task. Analysis of tapping performances in the frequency domain showed that the precision of reproduction of simple patterns is mostly affected by local changes, whereas global changes have comparable detrimental effects when reproducing complex patterns. Additional analysis in which we applied a shape mapping technique to tapping signals, revealed that the detrimental effects observed were not simply due to random noise in the reproduction, but partially explained by the presence of systematic biases and sequence rescaling effects. Taken together, these results suggest that the ability to precisely reproduce a temporal pattern depends on the internal regularity within its elements, but also on the regularity of its overall structure when using visual and complex patterns. The accuracy of the reproduction, however, does not reflect the ability to acquire and reproduce the global structure of the pattern, as highlighted by internal consistency measures.

2.1 INTRODUCTION

Experiencing an awkwardly long moment of silence during a conversation or distinguishing a "SMS" from a "SOS" message delivered in Morse code represent two very different types of temporal computations that the human brain is equally able to perform efficiently and effortlessly. In the first case, evaluating the awkwardness requires to estimate the duration of a single event (the moment of silence), while in the second case the brain needs to compare multiple features of two temporal patterns, i.e. the identity of their durations and the way they are related. The ability to process and represent such different forms of temporal information is critical for a wide range of sensory and motor experiences.

Although the underlying mechanisms and neural underpinnings are still matter of investigation and debate in time perception literature, it is generally acknowledged that, at least in the hundreds of millisecond range, processing a single duration is for the human brain quite a different operation than processing a temporal pattern (for a recent review see Paton & Buonomano, 2018). In line with this distinction, a long-standing model in the field of timing proposes two distinct mechanisms for the processing of the absolute duration of events (interval-based) and the processing of their relative duration (beat-based) (McAuley & Jones, 2003; Grube et al., 2010a,b; Bouwer et al., 2020). These two mechanisms recruit both shared (pre-motor and pre-frontal cortices) and more specialized networks of brain areas, namely the olivo-cerebellar system for interval-based timing and the cortico-striatal network for beat-based timing (Teki et al., 2011, 2012; Grahn & Brett, 2009; Breska & Ivry, 2018). Despite this model can account for a wide range of temporal judgements and types of temporal expectations, its core assumption is essentially a dichotomy of timing "strategies". It does not offer an exhaustive explanation on whether and how humans encode and represent all the features that characterize a temporal pattern (e.g., the single durations it is made of, their order, the overarching structure connecting them). In other words, it is currently unclear whether processing a temporal pattern encompasses an additive processing of its single elements, or a holistic representation of its structure, or both of them.

A potentially useful approach in this sense comes from several studies on sequence processing, in particular those investigating how the brain processes different hierarchical levels of sequences' regularity and how it builds level-specific expectations on them (e.g., Dehaene et al., 2015). These studies usually fall within the general framework of predictive coding models and their focus is on the neural responses associated to the recognition or to the violation of these regularities (for a review see Helibron & Chait, 2018). The paradigms used in these studies represent useful tools to investigate also how humans process temporal sequences, by focusing on the signatures associated to different violations of their features.

Adopting this approach, we conducted a series of experiments to investigate how the reproduction of a temporal sequence is affected by different levels of violation of its regularity, namely a violation within its elements and of its overall structure. Moreover, we asked whether the behavioral signatures of these violations would change if the sequences are experienced in the visual or auditory domain, and whether their structure is simple or complex.

To tackle these questions, we employed one of the aforementioned paradigms, the local-global paradigm, originally developed in the field of consciousness studies to differentiate neurophysiological signatures associated to the detection of violations of embedded levels of regularity of a sequence (Bekinschtein et al., 2009). Specifically, the paradigm allows to orthogonally manipulate violations within the single elements composing the sequence (i.e., local violations) and violations in its overall structure (i.e. global violations). The original version and variants of the paradigm have subsequently been employed in numerous studies, using different techniques and/or theoretical framework to account for the neural responses it elicits (e.g., Wacongne et al., 2011; El Karoui et al., 2015; Strauss et al., 2015), overall demonstrating its ability to dissociate automatic, low-level prediction-related neural signatures of local violations of a sequence (i.e., Mismatch negativity, increased activation localized in superior temporal gyrus) from higher-level prediction-related and awareness-dependent signatures of global violations (i.e., P3b component, activation of a distributed network of brain areas).

We adapted the local-global paradigm to the temporal domain and conducted three separate behavioral experiments in which we manipulated the temporal features of the sequences (i.e. instead of their pitch as in the original version) and asked our participants to reproduce them in a Synchronization-Continuation task (SCT, Zarco et a., 2009; Merchant et al., 2011). In Experiment 1 (Exp1V) we employed simple sequences marked by visual stimuli, in Experiment 2 (Exp2A) we used the auditory version of the same sequences of Exp1V, and in Experiment 3 (Exp3VC) we used complex sequences, again using visual stimuli as markers. Given that the main feature of the local-global paradigm is the combination of hierarchical violations of a sequence and our focus was on how each sequence was reproduced according to its local-global status, we opted for a data-analysis approach that would allow us to consider simultaneously the reproduction signatures of single elements and overall structures of the sequences, by combining signal processing (frequency-domain) and shape mapping (time-domain) techniques.

2.2 METHODS

2.2.1 Participants

Exp1V. Thirty participants (9 males; 2 left-handed; mean age 25 ± 5.1; age range 19-45) were recruited for Exp1-Visual. Eleven of them reported some form of musical expertise, either as amateurs musicians (N=5, 6.2 years on average) or as a result of formal musical/dance training (N=6, 7.8 years on average).

Exp2A. Twenty-nine participants (11 males; 3 left-handed; mean age 23 ± 3; age range 19-32) were recruited. Eighteen participants reported some form of musical expertise, either as amateurs musicians (N=10, 5.2 years on average) or as a result of formal musical/dance training (N=8, 8.3 years on average).

Exp3VC. Twenty-nine participants (11 males; 3 left-handed; mean age 23.4 ± 3.5; age range 18-35) were recruited. Thirteen participants reported some form of musical expertise, either as

amateurs musicians (N=6, 4.2 years on average) or as a result of formal musical/dance training (N=7, 7.4 years on average).

In all three studies each participant gave written informed consent before starting the experiment, the protocol of which was approved by the SISSA Ethics Committee (protocol number 11175-III/13) and complied with the Declaration of Helsinki. All recruited participants were completely naïve to the purposes of the study. Four of them were recruited for two out of three experiments (two participated to Exp1V and Exp2A; one to Exp1V and Exp3VC; one to Exp2A and Exp3VC): for these participants a complete debriefing on the purposes of the studies was given only after the end of the data collection. All participants received a fixed monetary compensation after completing the task.

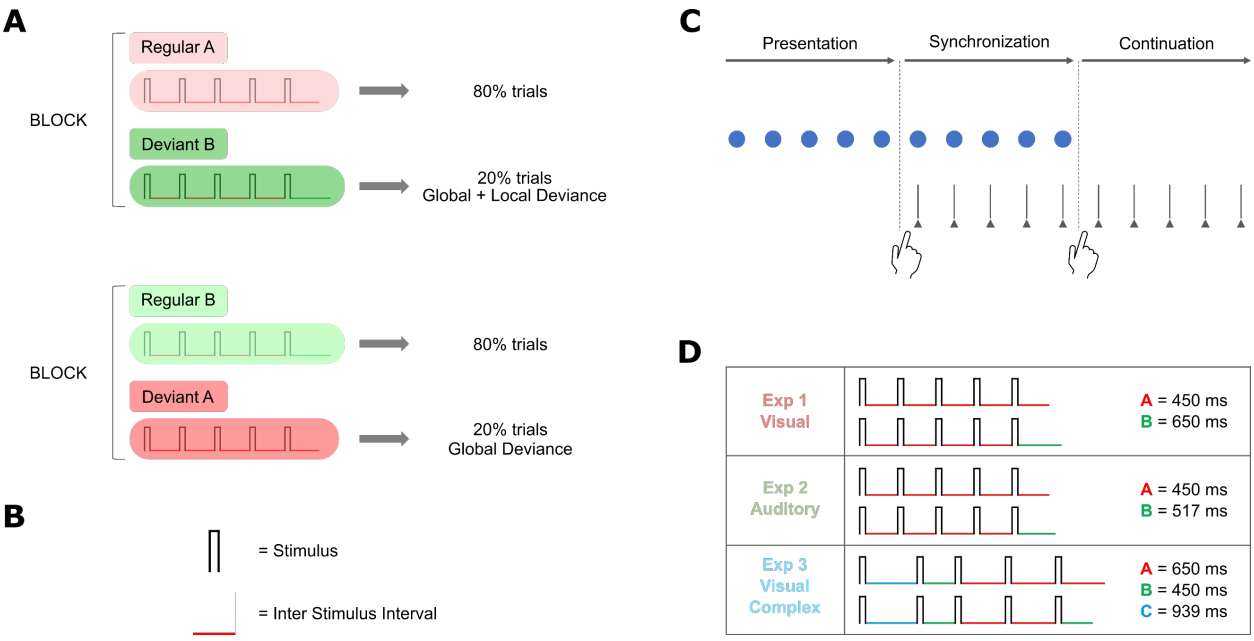


Figure 2.1. **Local-global paradigm and task structure**. (A) Representation of the local-global paradigm. In one type of block regular trials contain the locally regular sequence AAAAA (shaded in light red) and the deviant trials contain the locally deviant sequence AAAAB (shaded in green): in this block the deviance is both at the local and global level. In the other type of block regular trials contain the locally deviant sequence AAAAB (shaded in light green) and the deviant trials the locally regular sequence AAAAA (shaded in red): in this block the deviance is only at the global level. (B) The temporal sequences used in all experiments were made of empty intervals (*Inter Stimulus Interval*) marked by brief visual or auditory stimuli (*Stimulus*). (C) Synchronization-Continuation task. Each trial of the task has three phases: first the participant simply attend to a rhythm given by an external stimulus (Presentation), then he has to synchronize his finger movement with this rhythm (Synchronization) and finally he has to carry on the rhythm without any external reference (Continuation). (D) The specific set of durations used to create the sequences of each experiment.

2.2.2 The Local-Global paradigm

The local-global paradigm employs two types of sequences that alternate across trials: the first is composed of five identical elements (i.e. AAAAA) and the second is made of four identical elements plus a fifth different one (i.e. AAAAB). Sequence AAAAA is defined "locally regular", given the absence of variation within its elements, whereas sequence AAAAB is defined "locally deviant" because it comes with an intrinsic violation within its elements. In each block the relative frequency of these two sequences across trials establishes which of them is identified as the "global regular" sequence and which is the "global deviant" one. Specifically, in one experimental condition (block type) sequence AAAAA is presented in 80% of the trials, thus becomes the global regular sequence, whereas sequence AAAAB is presented in the remaining 20% of the trials, hence is the global deviant sequence (**Figure 2.1A**). In the other condition the relative frequency is reversed: sequence AAAAB is presented in 80% of the trials, hence it is the global regular sequence, whereas sequence AAAAA is presented in 20% of the trials, thus represents the global deviant sequence. This orthogonal manipulation results in two violation conditions: in the first block type the violation is both at the local and global level (both within the elements and in the overall structure of the sequence); in the second block type the violation is only at the global level (the change is only in the overall structure of the sequence, not within the elements).

In the current study we adapted this paradigm to the temporal domain, using sequences made of empty intervals interleaved by brief visual or auditory markers (**Figure 2.1B**), hence the changing feature was the length of the durations. As shown in **Figure 2.1D**, for Exp1V and Exp2A we employed the same sequences' structures used in the original paradigm (AAAAA and AAAAB), whereas for Exp3VC we used two complex sequences (i.e. non isochronous), whilst maintaining the local regularity/deviance contraposition between the fourth and the fifth element (CBAAA and CBAAB).

2.2.3 Stimuli and apparatus

Stimuli. The sequences used were made of empty durations marked either by brief blue disks (33 ms, radius 32 pixels) for Exp1V and Exp3VC, or beeps (33 ms pure tones of 1000 Hz, 2-ms rise and fall time filter) for Exp2A. Both types of markers were created using custom codes in Matlab (MathWorks). For Exp1V, the empty intervals selected were A=450 ms and B=650, with B 45% longer than A, to ensure a large enough discriminability among them. For Exp2A the interval B was reduced to 517 ms (15% longer than A), to take into account the finer temporal sensitivity of the auditory system and maintain a comparable level of difficulty in the auditory task compared to visual task (Grondin, 2010). For Exp3VC the empty intervals selected were A=650 ms, B=450ms and C=939 ms, each 45% longer than the closest one. In this case the shortest duration represented also the local deviant element of the sequence: this choice allowed us to control for overlooked detrimental effects in Exp1V and Exp2A due to duration B being longer than A, and not to its "deviant" status (**Figure 2.1D**).

Apparatus. To collect participants' response and keep track of each stimuli-related event, in the three experiments we employed a custom-made set up (www.cynexo.com) that allowed us to have a single common recording and timescale for all events in the trial. The set up was made of a microcontroller Arduino Due (www.arduino.cc) connected to a response pad and a photodiode attached to the monitor. In Exp2A the same apparatus was used also for delivering auditory stimuli, by adding to the system a second microcontroller (Arduino Uno) paired with a shield for sound delivery (Adafruit Wave Shield for Arduino, www.adafruit.com), into which headphones were directly plugged. Each trial event series was generated using Matlab and Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007), whereby each stimulus onset was coded as a visual trigger appearing on one corner of the screen (hidden from the participant). This trigger was read by the photodiode, whose instantaneous change of status and timestamp were recorded by the microcontroller. For Exp1V and Exp3VC each visual trigger was also paired with the simultaneous appearance of the visual marker at the center of the screen. In Exp2A, instead, the change of status of the photodiode triggered in turn the two microcontrollers serially to play the beep sound, marking the onset of each auditory event. Preliminary technical tests showed a constant delay of 23 ms in sound delivery from the onset of the visual trigger to the headphones, with a negligible standard deviation (in the microsecond scale). The response pad was equipped with a plastic button, where participants were instructed to tap, and its signal was a binary value (1 if pressed, 0 if not pressed) sampled at 500 Hz. To avoid any perceptual feedback produced by the button click, in Exp1V and Exp3VC white noise was delivered through headphones throughout the whole experimental session, whereas in Exp2A we employed noise-cancelling headphones. In Exp1V and Exp3VC participants' gaze was tracked using an EyeLink ® 1000 (SR Research) device, to ensure that visual fixation was maintained at the center of the screen.

2.2.4 Procedure and Task

Participants sat comfortably in front of a computer screen (monitor BenQ-XL2720Z, 35 x 28 cm, resolution 1920x1080 pixels), in a dimly lit room. After receiving general instructions on the experimental procedure, they were asked to wear headphones and had their head placed on a chin rest (used also in Exp2A, to limit differences in the set up). A short training to familiarize with the task preceded the main experiment. To avoid any transfer of temporal learning, the sequences used in the training were made of different durations than those used in the main experiment (A=500 ms, B=700ms in Exp1V; A=500 ms and B=575 ms in Exp2A; A=700 ms, B=500 ms and C=980 ms). Both in the training and in the main experiment, the structure of each block was the same: to establish the condition-specific global regularity, the first 5 trials (i.e. familiarization trials, 2 in the training) contained only the global regular sequence; familiarization trials were followed by 10 experimental trials (5 in the training), 8 of which contained the global regular sequence and 2 the global deviant sequence (for the training this proportion was 4/1). Within a block, the order of the experimental trials was randomized, so that any of them could contain the global deviant sequence. Participants performed two training blocks, one per condition type, and received a

feedback on their tapping after each trial. In the main experiment each condition type was repeated for 3 consecutive blocks, so that one condition was performed in the first half of the session and the other condition in the second half. The presentation order of the two block-triplets was counterbalanced across subjects. In total, for each condition type, participants performed 24 trials containing the global regular sequence and 6 trials containing the global deviant one. No feedback was given during the main experiment. Participants were allowed to take short breaks between blocks, if necessary.

Synchronization-Continuation Task. Each trial consisted of three phases (**Figure 2.1C**). 1) *Presentation*: after a brief random interval (jittered between 0.9-1.7 seconds) the sequence was presented 5 times to the participant, who was asked to simply pay attention to it. 2) *Synchronization*: the subject was cued by a message on the screen to start tapping on the response pad in synchrony with the external stimulus (i.e., blue disk or tone), for a total of 5 presentation of the sequence. 3) *Continuation*: once the external stimulus disappeared, the subject had to continue tapping at the same pace, only relying on his/her internal representation of the sequence. A message on the screen signaled the end of the trial, approximately after 3 repetitions of the sequence plus an interval randomly selected between 0.9-1.7 seconds. The beginning of the next trial was self-paced. During Presentation and Synchronization phases each repetition of the sequence immediately followed the previous one, so that the sixth marker of one sequence corresponded to the first of the next sequence. This was done in order to have a continuous stream of stimuli and avoid further temporal manipulation (i.e., inter sequence interval) that could bias the way participants group the elements of the sequence in their tapping.

2.3 DATA ANALYSIS

2.3.1 Pre-processing

All pre-processing steps to extract events' onsets were performed trial by trial for each participant, using custom codes written in Matlab (MathWorks). During each trial stimuli and presses' signals were simultaneously recorded by our apparatus as parallel time series of 0 and 1, where 1 represents an event's onset. Specifically, each stimulus' onset corresponded to a change of status of the photodiode, elicited by the appearance of the visual trigger on the screen, whereas each press' onset corresponded to a button press. For both time series, each positive derivative was identified as an onset and stored with its timestamp. For Exp2A the actual timing of the onset of each beep was calculated by adding a constant delay of 23 ms.

As additional pre-processing steps for the tapping series, presses' onset following a previous one within a window of 250 ms were considered either as artifacts or "false" onsets (e.g. double tap), hence discarded. Possible isolated onsets detected during Presentation phase or at the end of Continuation phase of the trial were also excluded (more than 2 sec before/after the next/previous onset). All familiarization trials were excluded from the analysis.

A few participants were excluded from the three original datasets due to reported difficulty in performing the task (i.e., inability to synchronize with the external stimulus) and/or poor performance (frequent long gaps in the tapping series in more than 25% of trials). The sample's size used for the analysis was N=26 for Exp1V, N=27 for Exp2A and N=26 for Exp3VC.

2.3.2 Frequency-domain analysis

To explore the effects of our experimental manipulations on the precision of tapping performance we analyzed the tapping data in the frequency domain, by computing and comparing the power spectra of stimuli's sequences and of participants' performance.

For each trial, . all stimuli and presses detected earlier than 500ms before the 1st stimulus' onset of the Synchronization phase were discarded. The timing of the 1st stimulus' onset was then subtracted to all other stimuli and presses' onset. The resulting time series from the same trial type and sequence type (i.e. global regular A, global deviant B, global regular B and global deviant A) were grouped together and convolved with a kernel density estimator (temporal resolution=2 ms, bandwidth=0.1), leading to four convolved tapping signals and four convolved stimulus signals for each participant. Each convolved signal was split into two parts in order to have separated series for Synchronization and Continuation phases (an example from Synchronization phase is shown in **Figure 2.2A**). The cutting point was chosen as the end cycle of the last stimulus' onset of the Synchronization phase. We then computed the FFT of the four convolved stimuli' series to identify the specific frequency power spectrum of the sequences used. Sequence AAAAA, being isochronous, has a power spectrum with an isolated peak around 2 Hz, whereas the spectrum of AAAAB in Exp1V lies within 0.3 and 2.3 Hz with the greatest peak around 2 Hz, and that of AAAAB in Exp2A between 1.6 and 2.3 Hz, with an analogous peak around 2 Hz. In case of Exp3VC, instead, both frequency spectra defining sequences CBAAA and CBAAB lie within 0.5 and 2.8 Hz.

For each participant, task phase, trial type and sequence type, we compared the power spectrum of the convolved tapping signal and the spectrum of its appropriate stimulus signal by means of Magnitude Squared Coherence (MSC), using Welch's overlapped averaged periodogram method, with sampling rate at 50 Hz and overlap of 50% between consecutive segments of the frequency spectrum. MSC is a precision measure, as it evaluates the similarity (i.e. coherence) between the power spectra of two signals, computed as the ratio between their cross-power spectral density and auto-spectral density functions (Kay, 1988). MSC takes values between 0 and 1, thus it gives a standardized measure for making comparisons across subjects. An example of MSC calculated for one participant in one experimental condition is shown in **Figure 2.2B**. As MSC returns a value of coherence for each frequency bin in the spectrum, we focused our analysis only on those bins within the sequence-specific frequency ranges listed above and calculated the mean of their MSC values.

Being a within-subject design, we tested the effects of our experimental manipulation by running a 3-way repeated measures ANOVA on mean MSC values, with trial type (global regular vs global deviant), sequence type (local regular vs local deviant) and task phase (Synchronization vs Continuation) as independent variables in interaction, and the error term split across factors. Frequency-domain analysis were all performed in Matlab (MathWorks) using built-in functions and custom codes. Statistical analyses on MSC values were performed both in Matlab and R (R Core Team, 2020), using built-in function for ANOVA. Post-hoc comparisons were conducted in R using *multcomp* package (Hothorn et al., 2008), with Tukey's method for p-value adjustment. Alpha level was always set at 0.05.

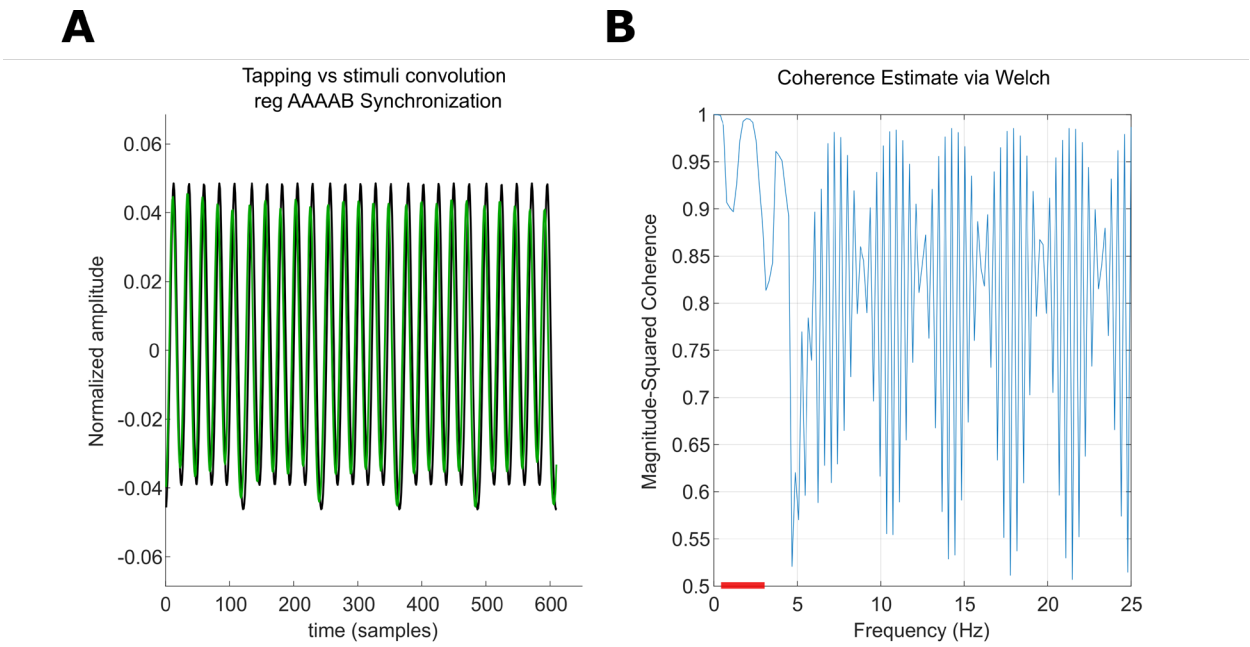


Figure 2.2. **Signal convolution and Magnitude-Squared Coherence (MSC)**. (A) Convolution of tapping (green trace) and stimulus' (black trace) signals from one participant during Synchronization phase of the trials containing the global regular sequence AAAAB. (B) MSC values obtained from the same group of trials plotted by frequency bin. The red bar on the x-axis highlights the frequency band on which statistical analysis on MSC values were carried out.

2.3.3 Procrustes analysis

To explore further the effects of our experimental manipulations on participants' performance, independently of precision measures, we carried out a second analysis more focused on the internal consistency of tapping signals. Specifically, we employed Procrustes analysis (Gower & Dijksterhuis, 2004), a mapping technique that applies a Euclidean transformation to find the best match between two shapes (or two linear signals) by simultaneously considering different non-random sources of distortions in the to-be-matched signal. These distortions are quantified as a

bias (translation) component, a scaling component and, in case of shapes, a rotation (reflection) component. Tapping signals are linear time series, hence the Procrustes transformation we applied included only the bias and the scaling components. In tapping behavior (see **Figure 2.3A**) these two components translate respectively in the presence of a constant anticipation or delay of presses' onsets with respect to stimuli's onsets (i.e. bias, with negative values expressing an anticipation and positive values a delay), and in a slowed down/speeded up reproduction of the stimuli series (i.e. scaling, with values above 1 indicating acceleration and below 1 deceleration of the tapping). Once these components are taken into account, the remaining distortions in the transformed signal are essentially noise-related and less quantifiable (**Figure 2.3B**).

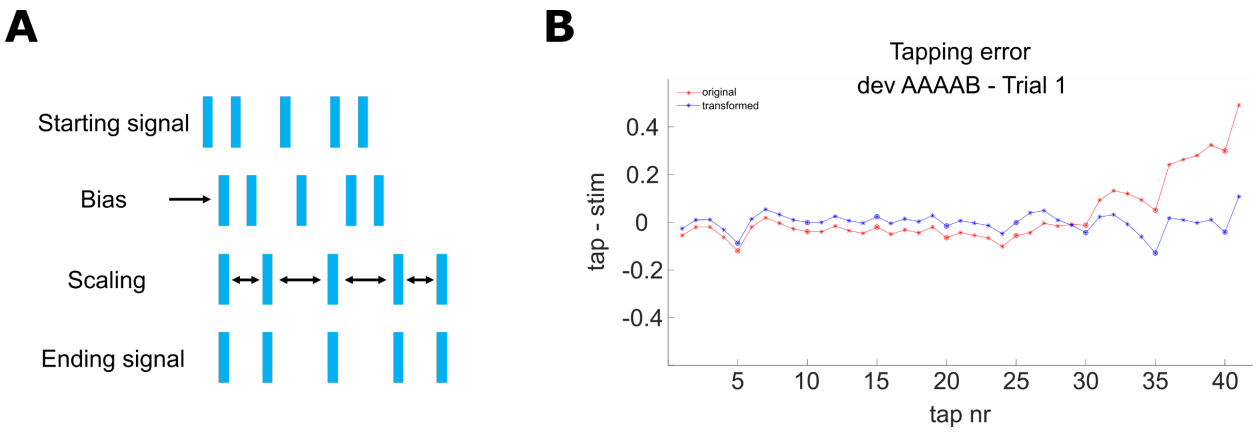


Figure 2.3. **Procrustes analysis**. (A) Procrustes applies a linear transformation to find the best match between the starting and ending signals by quantifying a bias and a scaling component. The bias corresponds to a translation in a positive (delay) or negative (anticipation) direction of the intervals' onsets, whereas the scaling represents either a stretching or squeezing of the intervals of the sequence. (B) Example of tapping signal from one trial before and after applying Procrustes. The tapping is represented in terms of error in respect to the stimulus' signal. The original signal (in red) shows a systematic bias across the whole trial and a scaling component that appears in the last 15 taps. Once these components are quantified, a general noise component remains in the resultant signal (in blue).

To apply Procrustes algorithm, in order to pair the length of tapping and stimuli signals we performed some additional pre-processing steps. First, all stimuli and presses' onsets detected earlier than 250ms before the 1st stimulus' onset of the Synchronization phase were removed. Of these tap series, those with less than 31 taps were discarded (<0.05% of the total number of trials) to ensure that all series used for Procrustes analysis contained enough data from both Synchronization and Continuation phases. We then aligned the tapping series to the stimuli series based on a minimum absolute difference criterion in their timing onsets during the first part of the Synchronization phase. In case of gaps between consecutive tapping onsets longer than twice the time of the longest difference in stimuli onset, we considered them as indicative of a "missing tap" and deleted the respective closest stimulus onset. The two resulting aligned series were then matched in length by cutting their starting point, based on the timing of the first tap and its closest

stimulus onset, and ending point, based on the length of the shortest series among the two. Each tap-stimulus time series was further split in its Synchronization and Continuation part, based on the onset timing of the first stimulus of the Continuation phase, and separately transformed by Procrustes algorithm. This led us to obtain two bias and two scaling values for every trial (one per task phase). For each of these four values, the effects of trial type (global regular vs global deviant) and sequence type (local regular vs local deviant) were assessed using LME models, with participants as random intercept. The described steps were done for all three experiments. Procrustes analysis was conducted in Matlab (MathWorks) using built-in functions and custom codes. Statistical analyses on scaling and bias values were performed in R (R Core Team, 2020). LME models were fitted using the packages *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017), and *MuMIn* package (Barton, 2020) was used for calculating R^2 statistics. Overall effects were tested using Type III Analysis of Variance with Satterthwaite's method for degrees of freedom. Post-hoc comparisons were conducted using *emmeans* package (Lenth, 2020), using Kenward-Roger's method for degrees of freedom and Tukey's method for p-value adjustment. Alpha level was always set at 0.05.

2.4 RESULTS

2.4.1 Magnitude Squared Coherence

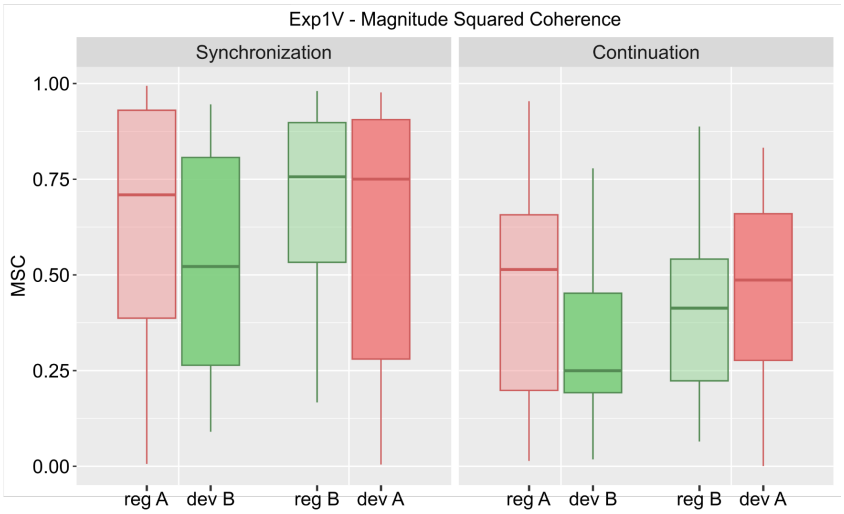
Tapping signals were analyzed in the frequency-domain using a measure of precision of performance (i.e. MSC), which expresses the degree of similarity/coherence of tapping's power spectrum with respect to the stimulus' spectrum. For all three studies a repeated measures ANOVA was performed to compare the effects of trial type (global regular vs global deviant), sequence type (local regular vs local deviant), task phase (Synchronization vs Continuation) and their interaction on mean MSC values (shown for all three experiments in **Figure 2.4**).

Exp1V. Repeated measures ANOVA on MSC values computed for the first experiment (**Figure 2.4A**) showed significant main effects of trial type ($F(1, 25) = 10.712$, $p < 0.01$) and task phase ($F(1, 25) = 28.207$, $p < 0.0001$), as well as a significant interaction between trial and sequence type ($F(1, 25) = 5.892$, $p = 0.022$). The main effect of trial type was due to lower MSC values for global deviant sequences (mean MSC = 0.481) compared to global regular sequences (mean MSC = 0.543). The main effect of task phase was due to lower MSC values during the Continuation phase (mean MSC = 0.408) compared to the Synchronization phase (mean MSC = 0.616). Concerning the interaction between trial and sequence type, post-hoc pairwise comparisons showed a significant difference in MSC values only between global regular AAAAA and global deviant AAAAB (dev B – reg A = -0.112084, z-value = -2.675, $p = 0.037$; all other contrasts are reported in **Appendix A, table A1T**). No main effect of sequence type ($F(1, 25) = 2.988$, $p = 0.096$), nor interactions between trial type and task phase ($F(1, 25) = 1.377$, $p = 0.251$), or between sequence type and task phase ($F(1, 25) = 2.506$, $p = 0.126$), or a 3-way interaction between them ($F(1, 25) = 1.803$, $p = 0.191$) were found. Overall the results of Exp1V revealed that only the passage from the global regular sequence AAAAA to the global deviant AAAAB had detrimental effects on

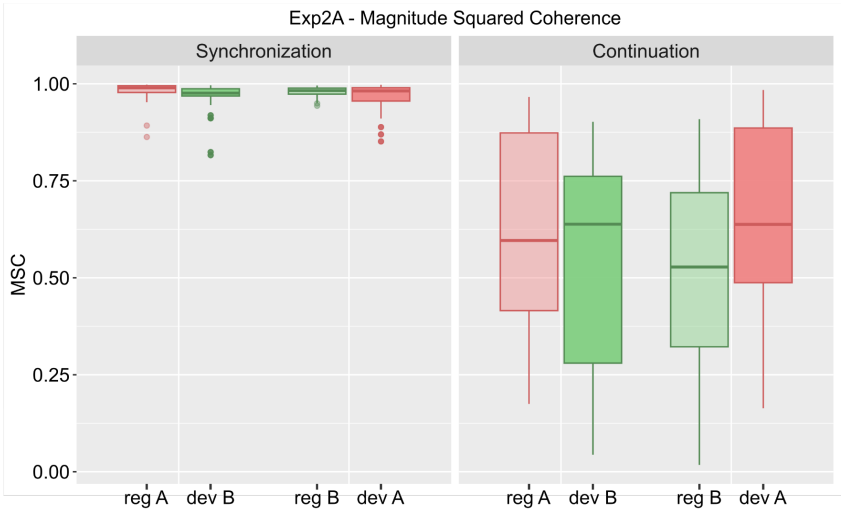
performance, irrespective of the task phase. The same effect was not found in the passage from global regular AAAAB to global deviant AAAAA, thus suggesting that it is the combination of local and global deviances that had a detrimental effect on tapping performance.

Exp2A. Concerning the same ANOVA on MSC values calculated for the auditory experiment (**Figure 2.4B**), results showed again a significant main effect of task phase ($F(1, 26) = 111.73$, $p < 0.0001$) because MSC was on average greater in the Synchronization phase (mean MSC = 0.964) compared to the Continuation phase (mean MSC = 0.570). Differently from Exp1V, we found a main effect of sequence type ($F(1, 26) = 8.971$, $p < 0.01$), due to lower MSC values for the local deviant sequence AAAAB (mean MSC = 0.733) compared to the local regular sequence AAAAA (mean MSC = 0.801). And a significant interaction between sequence type and task phase ($F(1, 26) = 8.026$, $p < 0.01$), since MSC values for the local regular sequence AAAAA were significantly higher than those for the local deviant sequence AAAAB in the Continuation phase only (B cont – A cont = -0.12246, z-value = -3.806, $p < 0.001$; B synch – A synch = -0.01426, z-value = -0.443, $p = 0.97$). No other main effect or interaction was significant (trial type: ($F(1, 26) = 0.014$, $p = 0.9083$; trial type x sequence type: $F(1, 26) = 0.065$, $p = 0.8$; trial type x task phase: $F(1, 26) = 3.269$, $p = 0.082$; trial type x sequence type x task phase: $F(1, 26) = 0.229$, $p = 0.636$). As in Exp1V the worsening of tapping performance during Continuation phase compared to Synchronization, was an expected effect. What is peculiar of this study is the clear presence of a pure local deviance effect when using auditory patterns, meaning that the local deviant sequence AAAAB was always less precisely performed, independently of its global "status". This was observed distinctly during Continuation phase, when the external reference was absent, although potential meaningful differences between sequence AAAAA and AAAAB during Synchronization phase may have been overlooked due to near ceiling levels of MSC values (i.e., close to 1), as can be noticed in **Figure 2.4B**.

A


Exp1V - Magnitude Squared Coherence
Synchronization
Continuation
MSC
1.00
0.75
0.50
0.25
0.00
reg A
dev B
reg B
dev A

B


Exp2A - Magnitude Squared Coherence
Synchronization
Continuation
MSC
1.00
0.75
0.50
0.25
0.00
reg A
dev B
reg B
dev A

C

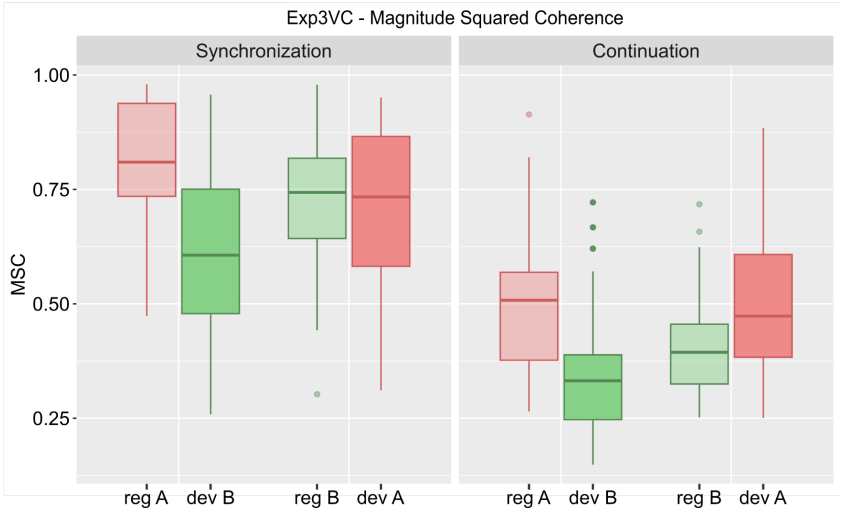

Exp3VC - Magnitude Squared Coherence
Synchronization
Continuation
MSC
1.00
0.75
0.50
0.25
reg A
dev B
reg B
dev A

Figure 2.4. **Magnitude-Squared Coherence values.** (A) Distribution of MSC values for Exp1V grouped by task phase, trial type and sequence type (color-code: light red = global and local regular sequence A; green = global and local deviant sequence B; light green = global regular and local deviant sequence B; red = global deviant and local regular sequence A): a significant drop in MSC values was observed only in the condition in which the deviance is both at the local and global level. (B) Distribution of MSC values for Exp2A grouped and color-coded as in (A): MSC analysis showed a generalized detrimental effect of local deviance, independent of the global status of the sequence. (C) Distribution of MSC values for Exp3VC grouped and color-coded as in (A): for complex sequences both local deviance and global deviance effects were observed, in particular the global deviance effect was significantly present during Synchronization phase irrespective of the local status of the sequence.

Exp3VC. Regarding the study using complex visual sequences, repeated measures ANOVA on MSC values (**Figure 2.4C**) highlighted a quite different combination of effects with respect to the other two studies. All three independent variables showed significant main effects (trial type: $F(1, 25) = 20.828$, $p < 0.001$; sequence type: $F(1, 25) = 36.46$, $p < 0.0001$; task phase: $F(1, 25) = 127.306$, $p < 0.0001$) and a significant interaction between trial type and task phase emerged as well ($F(1, 25) = 9.381$, $p < 0.01$). The main effect of trial type was caused by lower MSC values in global deviant sequences (mean MSC = 0.546) compared to global regular sequences (mean MSC = 0.615). The main effect of sequence type was due to lower MSC values for the local deviant sequence CBAAB (mean MSC = 0.528) compared to the local regular sequence CBAAA (mean MSC = 0.633). The main effect of task phase was again caused by lower MSC values during the Continuation phase (mean MSC = 0.443) compared to Synchronization phase (mean MSC = 0.718).The significant interaction was due to lower MSC values for global deviant sequences compared to global regular sequences during the Synchronization phase (globReg synch – globDev synch = 0.10881, z-value = 4.277, $p < 0.001$; globReg cont – globDev cont = 0.03027, z-value = 1.19, $p = 0.633$). No other interaction was significant. Overall results from Exp3VC revealed that the precision of tapping performance was affected by both levels of violation of the sequence's regularity (local and global): namely, local deviant sequences were always tapped less precisely, whether they were global regular or deviant. At the same time, global deviant sequences, no matter their local "status", were harder to tap in Synchronization phase, as highlighted by the interaction with task phase (lower MSC values for global deviant sequences during Synchronization).

Taken together, MSC analysis results revealed that when using simple temporal patterns, as in Exp1V and Exp2A, a small change within the pattern's elements (i.e. local deviance) has detrimental effects on participants' ability to reproduce it precisely. In the auditory domain this effect is completely independent from the familiarity with the overall structure of the pattern (pure local deviance effect), whereas in the visual modality a worsening of performance is evident when the local change corresponds also to a change in the structure of the sequence (global + local deviance effect). When using more complex temporal patterns, as in Exp3VC, the performance is still affected by local violations of the pattern (local deviance effect), but the ability of participants to precisely synchronize with it is impaired when the overall structure of the pattern is unexpected (global deviance effect during Synchronization).

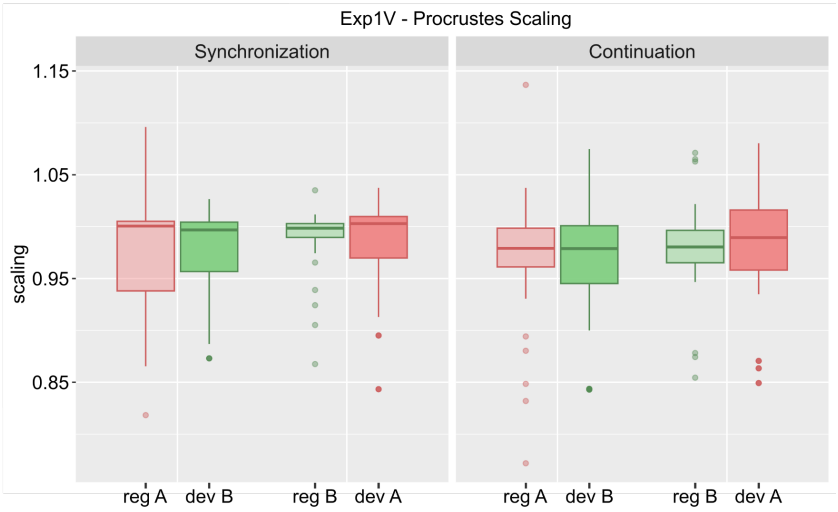
2.4.2 Procrustes analysis

MSC is a static measure of tapping precision with respect to the external stimulus and it is sensitive to phase shifts and amplitude differences (Pereda et al., 2005). In other words, MSC cannot tell whether the detrimental effects observed in the precision arose because the tapping signal was just out of phase with respect to the stimulus, or because its spectrum had a similar density shape but a neighboring frequency, or the tapping was simply noisier than the stimulus signal. Translating this into behavioral signatures, MSC cannot discriminate whether the observed worsening of the performance was due to a general anticipation or delay of the tapping with respect to the stimuli, or to a global slowing down or speeding up of the pattern performed. To quantify these components in tapping signals and separate them from pure noise, we explored the internal consistency of participants' performances by means of Procrustes analysis in all three experiments. For each trial and task phase we derived a bias and a scaling value, resulting in four measures per trial (scaling component values are shown in **Figure 2.5** across the three experiments; see **Appendix A, Figure A1** for bias values across experiments). We fit a separate LME model with each of them as dependent variable, and trial type (global regular vs global deviant) and sequence type (local regular vs local deviant) as independent variables in interaction, with participants as random intercept (model formula: component ~ trial type * sequence type + (1 | subject)).

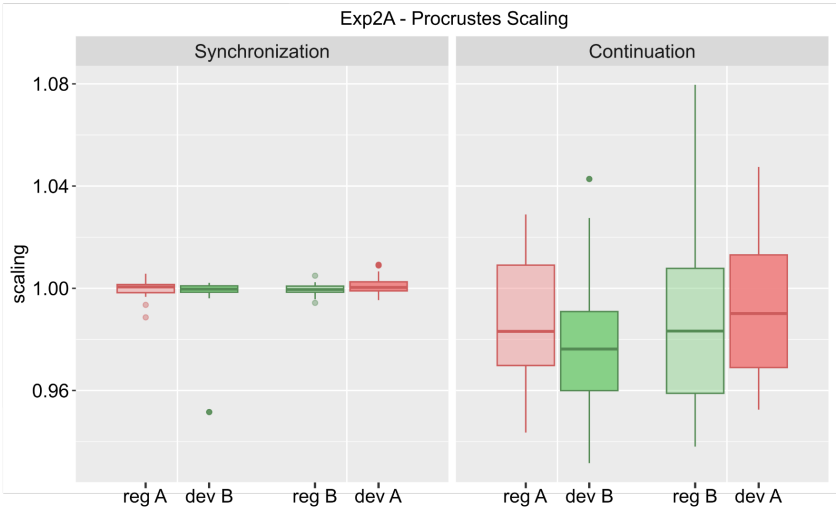
Exp1V. Goodness of fit of the four models was evaluated using R^2 statistic (scaling-synch: $R^2_{marginal}$ = 0.0082, $R^2_{conditional}$ = 0.8336; scaling-cont: $R^2_{marginal}$ = 0.0098, $R^2_{conditional}$ = 0.8372; bias-synch: $R^2_{marginal}$ = 0.006, $R^2_{conditional}$ = 0.8328; bias-cont: $R^2_{marginal}$ = 0.0107, $R^2_{conditional}$ = 0.8128). Type III ANOVA on the two scaling models estimates showed only a significant interaction between trial type and sequence type (scaling-synch: $F(1, 75) = 4.1595$, $p = 0.045$; scaling-cont: $F(1, 75) = 5.4576$, $p = 0.022$), although in both cases all pairwise comparisons failed to reach significance in post-hoc tests (all contrasts reported in **Appendix A, tables A2T and A3T**). No main effects of trial type (scaling-synch: $F(1, 75) = 0.0557$, $p = 0.814$; scaling-cont: $F(1, 75) = 0.6627$, $p = 0.418$) and sequence type (scaling-synch: $F(1, 75) = 0.8806$, $p = 0.351$; scaling-cont: $F(1, 75) = 0.1038$, $p = 0.748$) were found. Type III ANOVA on the two bias models estimates showed quite similar results: only a trend toward significance for the interaction between trial type and sequence type in the Synchronization model (bias-synch: $F(1, 75) = 3.593$, $p = 0.0618$), that reached significance in the Continuation model (bias-cont: $F(1, 75) = 4.74$, $p = 0.032$), although, again, all pairwise comparisons failed to reach significance in post-hoc tests (all contrasts reported in **Appendix A, tables A4T and A5T**). No main effects of trial type (bias-synch: $F(1, 75) = 0.0136$, $p = 0.907$; bias-cont: $F(1, 75) = 0.5604$, $p = 0.456$) and sequence type (bias-synch: $F(1, 75) = 0.1166$, $p = 0.733$; bias-cont: $F(1, 75) = 0.6029$, $p = 0.439$) emerged. Though not fully supported by statistical results, however scaling and bias values distributions (**Figure 2.5A and Appendix A, Figure A1A**) inform on the type and direction of performances' distortions observed in this specific study: specifically, a slight tendency to slow down and delay the tapping when passing from global regular AAAAA to global deviant AAAAB, and to speed up and anticipate it when going from global regular AAAAB to global deviant AAAAA.

Exp2A. Goodness of fit of the four models was evaluated using R^2 statistic (scaling-synch: $R^2_{marginal}$ = 0.0466, $R^2_{conditional}$ = 0.3094; scaling-cont: $R^2_{marginal}$ = 0.0691, $R^2_{conditional}$ = 0.7109; bias-synch: $R^2_{marginal}$ = 0.0365, $R^2_{conditional}$ = 0.4249; bias-cont: $R^2_{marginal}$ = 0.0679, $R^2_{conditional}$ = 0.7048). Type III ANOVA on the two scaling model estimates showed a main effect of sequence type (scaling-synch: $F(1, 78) = 4.7011$, $p = 0.033$; scaling-cont: $F(1, 78) = 12.2281$, $p < 0.001$) and a significant interaction between trial type and sequence type only during Continuation (scaling-synch: $F(1, 78) = 2.4727$, $p = 0.119$; scaling-cont: $F(1, 78) = 12.1004$, $p < 0.001$). In both cases these results were indicative of a general slowing down of the tapping when performing the local deviant sequence AAAAB compared to the local regular AAAAA (**Figure 2.5B**), although during Continuation phase this distortion was statistically stronger only when AAAAB was also globally deviant (Reg A – Dev B = 0.0156, t.ratio(78) = 3.261, $p < 0.01$; Dev A – Reg B = 0.00806, t.ratio(78) = 1.684, $p = 0.338$), as revealed by post-hoc comparisons (all scaling-cont contrasts are reported in **Appendix A, table A6T**). No main effect of trial type (scaling-synch: $F(1, 78) = 0.0557$, $p = 0.814$; scaling-cont: $F(1, 78) = 0.6627$, $p = 0.418$) was found. The same pattern of results emerged for the two models on the bias component. Type III ANOVA on models estimates showed a main effect of sequence type (bias-synch: $F(1, 78) = 4.8225$, $p = 0.031$; bias-cont: $F(1, 78) = 12.2632$, $p < 0.001$) and a significant interaction between trial type and sequence type only during Continuation (bias-synch: $F(1, 78) = 1.9559$, $p = 0.165$; bias-cont: $F(1, 78) = 11.1503$, $p < 0.01$). In both cases these results indicated a significant delay when tapping the local deviant sequence AAAAB compared to the local regular AAAAA (**Appendix A, Figure A1B**), although during Continuation this delay was statistically bigger only when AAAAB was also globally deviant (Reg A – Dev B = -0.3909, t.ratio(78) = -3.253, $p < 0.01$; Dev A – Reg B = -0.2041, t.ratio(78) = -1.699, $p = 0.331$), as revealed by post-hoc comparisons (all bias-cont contrasts are reported in **Appendix A, table A6T**). No main effect of trial type (bias-synch: $F(1, 78) = 0.0243$, $p = 0.876$; bias-cont: $F(1, 78) = 1.2079$, $p = 0.275$) was found.

A


Exp1V - Procrustes Scaling
Synchronization
Continuation
scaling
1.15
1.05
0.95
0.85
reg A
dev B
reg B
dev A

B


Exp2A - Procrustes Scaling
Synchronization
Continuation
scaling
1.08
1.04
1.00
0.96
reg A
dev B
reg B
dev A

C

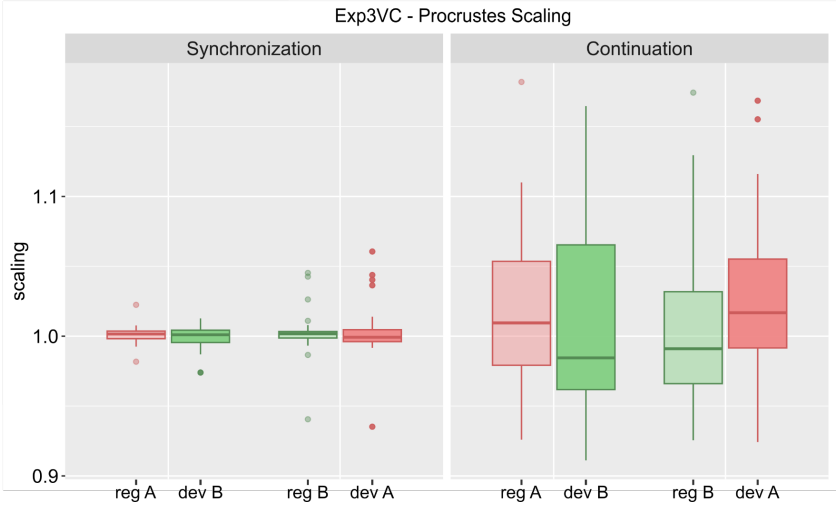

Exp3VC - Procrustes Scaling
Synchronization
Continuation
scaling
1.1
1.0
0.9
reg A
dev B
reg B
dev A

Figure 2.5. **Scaling component results.** (A) Distribution of scaling values for Exp1V grouped by task phase, trial type and sequence type (color-code: light red = global and local regular sequence A; green = global and local deviant sequence B; light green = global regular and local deviant sequence B; red = global deviant and local regular sequence A): in this dataset we observed only a mild tendency to slow down the tapping when the deviance is both at the local and global level and to speed up the tapping when the deviance is only global. (B) Distribution of scaling values for Exp2A grouped and color-coded as in (A): statistical analysis revealed a general local deviance effect expressed as a deceleration of the sequence AAAAB. (C) Distribution of scaling values for Exp3VC grouped and color-coded as in (A): also for complex sequences a statistically significant slowing down of the tapping was observed when performing the local deviant sequence CBAAB.

Exp3VC. Goodness of fit of the four models was evaluated using R^2 statistic (scaling-synch: $R^2_{marginal} = 0.0106$, $R^2_{conditional} = 0.5691$; scaling-cont: $R^2_{marginal} = 0.0119$, $R^2_{conditional} = 0.8722$; bias-synch: $R^2_{marginal} = 0.0073$, $R^2_{conditional} = 0.5404$; bias-cont: $R^2_{marginal} = 0.0086$, $R^2_{conditional} = 0.8748$). Type III ANOVA on the two scaling models estimates showed only a main effect of sequence type during Continuation (scaling-synch: $F(1, 75) = 0.5068$, $p = 0.4787$; scaling-cont: $F(1, 75) = 8.8441$, $p < 0.01$), indicating a similar trend already observed in Exp2A, i.e. a general slowing down of the tapping when performing the local deviant sequence CBAAB compared to the local regular CBAAA during Continuation phase (**Figure 2.5C**). No main effect of trial type (scaling-synch: $F(1, 75) = 0.0401$, $p = 0.8418$; scaling-cont: $F(1, 75) = 0.2955$, $p = 0.588$) nor interaction with sequence type (scaling-synch: $F(1, 75) = 1.9964$, $p = 0.1618$; scaling-cont: $F(1, 75) = 0.4857$, $p = 0.488$) emerged. An overlapping pattern of results was observed for the two models on the bias component. Type III ANOVA on models estimates showed again only a main effect of sequence type for Continuation phase (bias-synch: $F(1, 75) = 0.0898$, $p = 0.7653$; bias-cont: $F(1, 75) = 6.8433$, $p = 0.01$), that was suggestive of a significant tendency to delay the tapping when performing the local deviant sequence CBAAB compared to the local regular CBAAA when the external reference was absent (**Appendix A, Figure A1C**). Also for bias values no main effect of trial type (bias-synch: $F(1, 75) = 0.0498$, $p = 0.824$; bias-cont: $F(1, 75) = 0.2531$, $p = 0.616$) nor interaction with sequence type (bias-synch: $F(1, 75) = 1.5054$, $p = 0.2237$; bias-cont: $F(1, 75) = 0.051$, $p = 0.822$) were found.

Overall, LME models results on Procrustes components showed how the detrimental effects in tapping precision found with MSC analysis were at least partially caused by specific and quantifiable distortions of the stimuli sequences, and not completely due to random noise. Specifically, when performing simple visual sequences that are globally deviant (Exp1V) the tapping is characterized by either a mild slowing down and delay (for global and local deviants AAAAB) or a slight speeding up and anticipation (for global deviants but locally regular AAAAA). When the same type of sequences is experienced in the auditory modality (Exp2A), the general distortions observed are a deceleration and a delay of the local deviant sequence (AAAAB), and they become bigger during Continuation phase for the sequence that is also globally deviant. When the sequences of stimuli have more complex structures (Exp3VC), the tapping of local deviant sequences (CBAAB) appear delayed and slowed down only during Continuation phase, when participants are relying on an internal representation of the sequences experienced. It must be noticed, though, that all the LME models we fit on scaling and bias

components had poor goodness of fit, as $R^2_{marginal}$ values showed, hence the variability explained by our experimental manipulations is quite limited. This suggests that other sources of distortions are at play, not necessarily noise-related.

2.5 DISCUSSION

In the current three studies we investigated the behavioral signatures associated to different levels of violation of the regularity of a sequence, namely a violation within its elements and of its overall structure. Moreover, we asked whether these signatures would change if the sequences used are made either of visual or auditory stimuli, and have simple or complex structures. In all three studies we employed the local-global paradigm (Bekinschtein et al., 2009), adapting it to the temporal domain, specifically to a Synchronization-Continuation Task. Participants' tapping series were analyzed in the frequency domain, using a precision measure (i.e. precision with respect to the external stimuli), and in the time-domain, using a shape analysis technique.

Overall, frequency-domain analysis showed a general effect of local deviance in all three experiments, meaning that precision of performance is highly affected by the regularity of the local temporal context (i.e. the internal regularity of the sequence). However the results also showed different global deviance effects for visual sequences, according to the degree of complexity of their structures. In Exp1V, which made use of simple visual sequences, the worsening of tapping performance was due to a combination of local and global deviance effects: MSC values dropped only in the experimental condition in which the violation was both at the local and global level (when going from AAAAA to AAAAB). In Exp2A, in which we used simple auditory sequences, the local deviance effect was spread across conditions, irrespective of their global status, and it was more pronounced during Continuation than Synchronization phase. In Exp3VC, where instead we employed complex visual sequences, both a local deviance and a global deviance effect were observed: the local deviance effect was generalized, independent of the global status of the sequence and the phase of the task, whereas the global deviance effect was independent of the local status of the sequence but specifically observed during Synchronization phase, meaning that participant's ability to precisely synchronize with a complex pattern was compromised when its overall structure was globally deviant.

This last result seems to suggest that the processing of complex temporal sequences requires perhaps the integration of local elements into a global structure, hence a violation of the overall structure has more detrimental effects. On the contrary, when reproducing a simple rhythm the local regularity is more relevant, and local changes have the greatest detrimental effects on performance both in the visual and auditory modality. This would explain the lack of a pure global deviance effect, in particular in Exp1V: the reproduction of isochronous (or nearly isochronous) sequences does not entail the need of integrating any structure, since all elements are identical, and may instead rely more on entrainment-like mechanisms (Morillon et al., 2016; Cadena-Valencia et al., 2018). It is interesting to notice here that the benefit of local regularity on performance precision were found also with complex sequences in Exp3VC (i.e. local deviance effect). This

result, combined with the simultaneous presence of the global deviance effect, seems to confirm that both local elements and overall structure have equal weight for precisely reproduce complex patterns. And this was manifested in the behavioral effects when their respective regularity was violated.

Overall, we observed a greater precision in performance (i.e. higher MSC values) when using auditory sequences, compared to visual ones, even if they had the same structure. This is consistent with all the literature on sensorimotor synchronization (for extensive reviews see Repp, 2005 and Repp & Su, 2013) and the notion that tapping in synchrony with simple auditory rhythms is a natural task for human adults. And it is therefore characterized by higher accuracy and less variability compared to the synchronization with visual rhythms (Repp & Penel, 2002; Zarco et al., 2009).

Procrustes analysis gave a precision-independent measure of performance, by decomposing the distortions in tapping signals in quantifiable and noise-related components. Overall its results revealed how part of the detrimental effects observed using precision measures were partially explained by the presence of a systematic bias and a rescaling of the pattern. In all three studies there was a general tendency to slow down and slightly delay the local deviant sequence, in particular during Continuation phase when the external reference was absent. This task phase-specificity is partly explained by the absence of error correction mechanisms which are more commonly employed when tapping with an external metronome. And it is probably due to the fact that participants rely more on an internal representation of the pattern which in turn may give rise to characteristic "drifts" in the tapping frequency spectrum (Chen et al., 2002). In general, the quantification of these reproduction's distortions allowed to reveal that the global structure of the sequences was acquired and expressed in the tapping behavior, even if in a rescaled and biased version. The scaling component, in particular, can be considered a signature of the acquisition of the overall structure of a temporal pattern. In line with this result, a study by Hardy & Buonomano (2018) showed that participants trained to reproduce a complex temporal pattern can also perform accurately rescaled versions of it.

Limitations and future directions. The three studies presented here have some theoretical and technical limitations. First, differently from the original version, our variant of the local-global paradigm employed multiple repetitions of the sequences within a trial, even when they were globally deviant, due to the structure of the SCT. These repetitions may have weakened their global status of "unexpectedness" and possible oddball-like behavioral signatures. Future adaptations of this paradigm to the time domain should consider this limitation. Partially related to this, the use of isochronous rhythms as global deviant sequences may be suboptimal for the aims of the local-global paradigm. The fact that they are by definition fully predictable prevents the need to create a representation of an overall structure. Concerning our analysis in the frequency domain, we employed a measure of precision of performance (MSC) that comes with some limitations: MSC is static and sensitive to phase shifts and amplitude differences between signals (Pereda et al., 2005), hence it would be useful to integrate it with a time-frequency analysis of tapping signals.

Finally, to add completeness to Procrustes analysis, a further specification of the noise component in the tapping signal would be necessary, to exclude the presence of potentially overlooked trends.

2.6 CONCLUSION

The current work sought to investigate how local changes within the elements of a temporal sequence or changes in its overall structure differently affect its reproduction, and how these behavioral effects vary according to the sensory modality and to the degree of complexity of sequences' structure. Results revealed that the precision of reproduction is strongly influenced by local regularities within the sequence but also by its overall structure (i.e. global level). These results also suggest that temporal patterns can be detected and learned even if not precisely reproduced, as highlighted by internal consistency measures of performance and by the use of more complex patterns. Overall these experiments suggest that the reproduction of a temporal pattern seems to require an interplay between individual constitutive durations and overall structure, and each of these two levels is weighted differently depending on the complexity of the pattern and the sensory modality in which it is experienced.

Chapter 3
Structure transition in temporal pattern learning

ABSTRACT

The ability to detect and learn temporal regularities across different timescales features an endless number of our sensory and motor experiences, and represents an essential skill for a successful interaction with our environment. Several of these experiences also require to promptly adapt to changes in the temporal structure of the events, from being able to learn new temporal associations between them, to quickly adjusting our actions to a sudden drop or increase in the temporal predictability of the environment.

Previous research on rhythm perception and sensorimotor synchronization has shown not only humans' ability to perceive, learn and coordinate with different levels of temporal regularities, but also their ability to detect violations of these regularities or adapt their motor behavior to changes in them. Despite several studies have investigated such abilities at the behavioral and brain level, little is known about how humans learn new temporal structures.

In the current study we addressed this question by focusing on how humans learn to switch from a known to a novel temporal pattern, made by the same elements but having different structures, and how this switch is influenced by trial predictability. To this aim we conceived a novel experimental design in which two symmetrical temporal patterns occurred within an auditory stream according to complementary probability values that parametrically changed across trials, generating a gradual transition from the full presence of one pattern to the full presence of the other. The auditory streams were presented in a synchronization task, in which participants were asked to tap their finger in synchrony with the external stream.

Behavioral data were analyzed in the form of standard measures of tapping performance (e.g. inter tap interval) and modelled to estimate the parameters underlying participants' learning process and their transition from one structure to the other, both at the trial and single tap level. More specifically, a first simple mixture model was able to describe subjects' performances trial by trial in terms of probability of reproducing one pattern or the other, compared to their actual proportions embedded in the external stimuli. A second model, by estimating the proportion of the two patterns in the tapping, as well as the tapping variability and the learning rates governing the update of patterns' probabilities, allowed to increase the temporal resolution in describing the evolution of the learning process. The models implemented were able to retrieve the learning transition that our participants experienced during the task and to categorize their performance and learning strategies. Specifically, a group of "fast learners" quickly learned to synchronize with the first unknown pattern and showed higher sensitivity to changes between patterns across trials; and a group of "slow learners" slowly learned to synchronize with the first pattern and slowly switched

to the second one. Overall these results highlight the effect of trial predictability in the learning of new temporal structures and its interaction with subject-specific strategies employed in the task.

3.1 INTRODUCTION

Being able to detect and enjoy a rhythmic change while listening to a jazz performance, or the capacity of rapidly switching from one motor sequence to another to successful shoot a basketball are only few examples of humans abilities in detecting and learning temporal regularities at multiple timescales to successfully interact with the environment. Part of these experiences also require to promptly adapt to changes in the temporal structure of the events, for example being able to learn new temporal associations between them, or deciding whether new associations are stable or volatile and adjust our actions accordingly.

Such abilities have their most explicit expression in the perception and production of rhythms, in which different combinations of the same set of intervals can lead to very different percepts, like a highly regular rhythm, a syncopated or a more complex one. Rhythm perception studies, indeed, have extensively explored the conditions and features that facilitate the task of learning and reproducing a temporal pattern. In particular the presence of hierarchical levels of regularities (i.e. meter), of an underlying regular pulse in the stimuli (i.e. beat) and the presence of durations related by small integer ratio (all being fundamental properties of musical rhythms) are the combination of features that best led to the perception of a rhythmical structure in external stimuli (Essen & Povel, 1985; Grahn, 2012).

Closely related to rhythm perception, sensorimotor synchronization studies focus on the mechanisms underlying the ability of humans and non-human primates to synchronize motor actions with rhythmic stimuli (Repp, 2005; Merchant et., 2015). Synchronization studies with humans, in particular, have frequently investigated individuals' abilities to detect changes in an external rhythm, mainly induced by perturbation of its period or phase (e.g., Repp & Keller, 2004, 2008), and adjust their synchronization to it. Although studied in humans the simpler ability to generalize a learned rhythm to different tempi has been found also in monkeys (Gámez et al., 2018). At the perceptual level, in rhythm detection tasks, some studies have focused on the brain responses associated with changes in different features of a rhythmic structure. For example, in a fMRI study Grahn & Rowe (2009) investigated the differences in the degree of activation of the striatal network linked to either the discovery of a beat, its continuation or the switch to a different beat rate. The authors of this work found an increase of putamen activity in the beat continuation condition, which required an internal generation and prediction of the common beat underlying different patterns. Focusing on other features of rhythms, a study by Geiser and colleagues (2009) examined the neurophysiological signatures associated with violations of meter and rhythm itself (i.e., the perceptual grouping of a series of notes), finding similar ERP correlates (negative deflection) for both violations. Another fMRI study (Notter et al., 2019), of particular interest for the purposes of the current work, compared pattern of brain activity associated with the

presentation of a specific rhythmic structure in contrast to another which was made of the same set of durations. Structure-specific patterns of activation were found in the temporoparietal junction bilaterally and in the right inferior frontal cortex, also when the absolute base duration changed (i.e. only the ratio between single intervals was kept the same). The task used in the study, though, was a passive exposure to the rhythms followed by a yes/no question that could or could not refer to the temporal features of the stimuli. So, the task didn't require an active, ongoing detection of change from one structure to the other, or a reproduction of them. In this sense, even if the pattern of brain activity could discriminate these structures as two different perceptual entities, it remains unclear how they were encoded, learned and at which point an internal representation of them arose.

Beside the focus of the outlined research in rhythm perception and sensorimotor synchronization, little is known about how humans discover and learn new temporal structures in contexts where a familiar structure is already present. In the current study we addressed this question, more specifically we asked how humans switch from a known temporal structure to a new one which shares the same constitutive elements, and how this transition is influenced by the predictability of the environment. To this aim we parameterized a gradual transition between two mirroring temporal patterns, by manipulating their probability of occurrence across trials. The two patterns were embedded in auditory streams and presented to our participants during a synchronization task. Behavioral and model-based analysis were performed on tapping data, with the latter aiming at defining and estimating the parameters underlying the learning transition from one structure to the other in participants' performance.

3.2 METHODS

3.2.1 Participants

Fourteen participants (4 males; 3 left-handed; mean age 25 ± 3.2; age range 21-32) were recruited for the present study. Three participants reported some years of formal musical training (5 years on average).

Before starting the experiment, each subject gave written informed consent to take part to the study, which was approved by the SISSA Ethics Committee (protocol number 11175-III/13) and complied with the Declaration of Helsinki. All recruited participants were naïve to the purposes of the study and received a fixed monetary compensation after completing the task.

3.2.2 Experimental design and stimuli

Given the focus of the current study on the transition from a learned temporal structure to a new one, we designed our temporal patterns as made of the same elements but having different structures, and we parameterized the transition from one to the other by manipulating their probability of being presented within and across trials.

Stimuli. We employed two patterns (A and B, **Figure 3.1B**) made of empty intervals marked by brief beeps (33 ms pure tones of 1000 Hz, 2-ms rise and fall time filter), created in Matlab (The MathWorks Inc.) using a custom code. The empty durations used were the same for both patterns (D_1 = 250 ms, D_2 = 500 ms, D_3 = 750 ms), related by small integer ratio (1:2:3), and only their order changed between patterns. The structures of the two patterns mirrored each other, with pattern A being 500-250-250-750 (2-1-1-3) and pattern B being 750-250-250-500 (3-1-1-2). These specific structures were selected to ensure comparable complexity, they were easy enough to be learned although they could not be metrically organized in a simple 4/4 bar. This ensured that the resulting trial streams would not feature a highly regular overarching metrical structure with accents perceived always on the expected beats (Povel & Essens 1985; Grahn & Brett, 2007).

Transition. Within a trial the two patterns were presented several times in random order, with a relative frequency that varied parametrically across trials following a logistic function with midpoint between the first and the second half of the trials (at the 16th trial) and slope equal to 0.1. As shown in **Figure 3.1A**, this formalization entailed a gradual shift from one pattern to the other, whereby pattern A was mostly experienced in the first half of the trials and its presence gradually decreased with the increase of the trials number, and it was replaced by pattern B, that was mostly present in the second half of the trials. Both the probability of the patterns and the random order in which they appeared within a trial of the first half (1-15) were mirrored in an equivalent trial of the second half (17-31): this meant, for example, that in the 17th trial the proportion of pattern A and B was reversed compared to the 15th trial, and the order in which they alternated mirrored that of the 15th trial (an example of trials' mirror randomization is shown in **Appendix B**, **Figure B1A**).

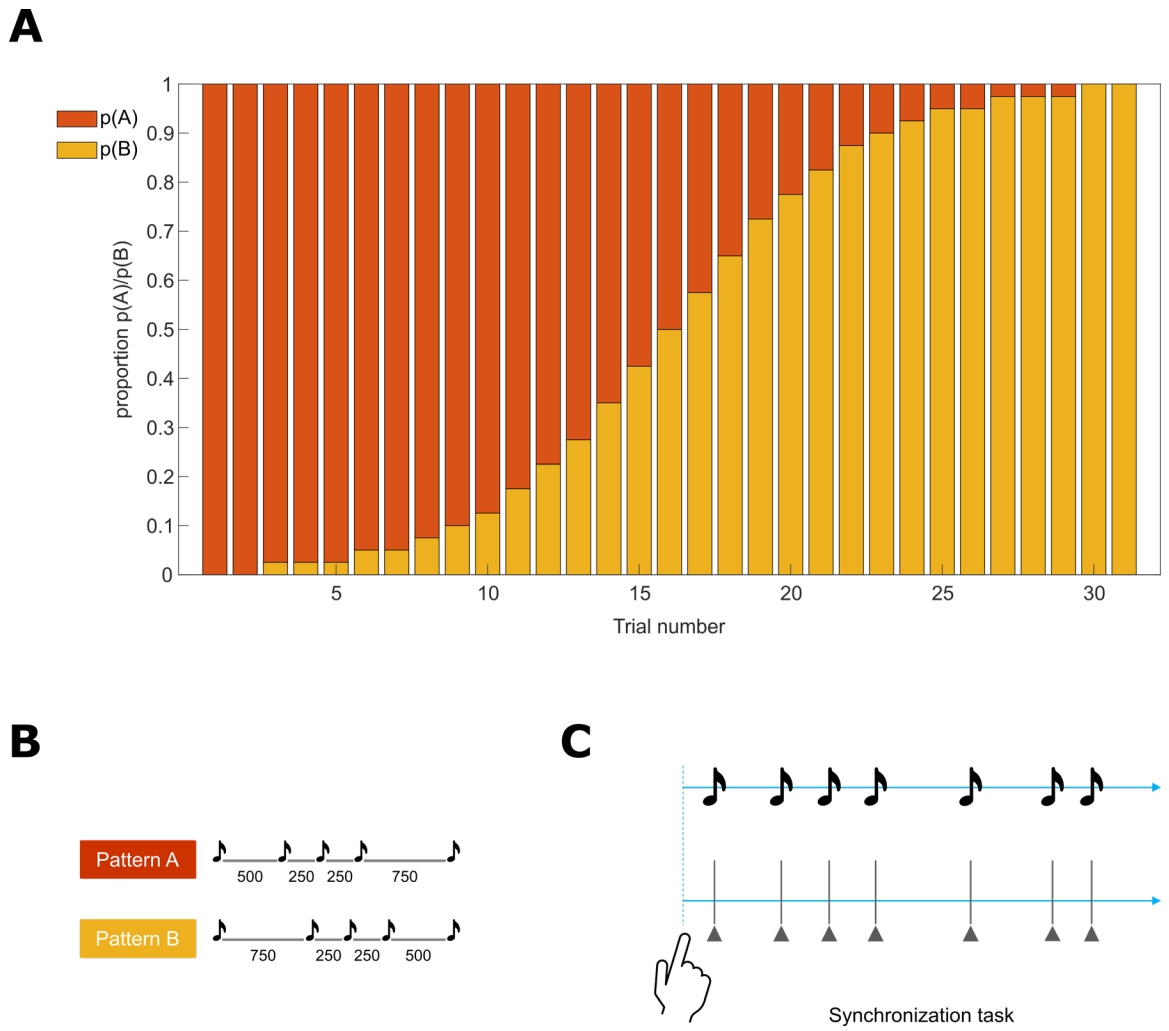


Figure 3.1. **Experimental design and task**. (A) The barplot shows the transition between patterns parameterized across trials as a gradual change in the probability distributions of pattern A and B according to a logistic function (color code: pattern A = dark red, pattern B = yellow). (B) Patterns structure: pattern A and B are made of the same set of empty durations, marked by brief tones. The durations are related by small integer ratio (1:2:3). The two structures are symmetrical. (C) Task: participants tap their finger in synchrony with the auditory stimulus.

3.2.3 Apparatus

The same custom-made set up (www.cynexo.com) was used to deliver the stimuli series and collect participants' tapping. The system included two microcontrollers, an Arduino Due (www.arduino.cc), connected to a response pad and a photodiode, and an Arduino Uno paired to a shield for sound delivery (Adafruit Wave Shield for Arduino, www.adafruit.com), into which noise-cancelling headphones were directly plugged.

Each trial event series was generated using Matlab and Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007), where each stimulus onset was coded as a visual trigger appearing on the

screen (hidden from the participant) and read by the photodiode. The instantaneous change of status of the photodiode triggered in turn the two microcontrollers serially to play the beep sound, marking the onset of each event. The response pad was equipped with a pressure sensor (SingleTact CS15-4.5N, www.singletact.com), attached underneath a resin-made button, so that the signal collected (sampled at 100 Hz) was analogic and reflected the force exerted on the button (range of sensitivity 0-4.5 Newton, represented in an arbitrary integer scale spanning from 0 to 1022). The resin-made button allowed us to get rid of the click or sensory feedback often associated with a key press on a plastic button. Preliminary technical tests showed a constant delay of 23 ms in sound delivery from the onset of the visual trigger to the headphones, with a negligible standard deviation (in the microsecond scale). This custom-made set up allowed us to have a single common recording and timescale for all events during the trial.

3.2.4 General procedure and task

Participants sat comfortably in front of a computer screen and wore noise-cancelling headphones. After receiving general instructions on the experimental procedure, they performed a short training to familiarize with the synchronization task (**Figure 3.1C**) and the response pad. Training consisted of two trials, each being a stream of 44 events (event = beep + empty interval), resulting in 11 repetitions of the same temporal pattern, to which participants had to synchronize their finger movement by tapping on the response pad. The training pattern was made of three durations, different from those used in the main experiment, related by small integer ratio (270, 540 and 810 ms, 1:2:3), with the following structure: 810-540-270-540 (3-2-1-2).

After the training participants started the main experiment, during which they performed 31 trials. Each trial was a stream of 164 events, corresponding to 41 presentations of the two patterns, lasting approximately 80 seconds. The stream was preceded by a brief random interval (0.9-1.7 sec), after which a message appeared on the screen concurrently with the onset of the first auditory marker of the first presentation. The message cued the participant to start tapping the finger on the response pad button in synchrony with the beep. The first presentation of a pattern was a copy of the second presentation as its function was only to establish participants' synchronization, whereas the following 40 were the experimental presentations of the pattern. Within a trial the probability of experiencing pattern A or pattern B was fixed and depended upon the parameterization defined in the experimental design. Whereas the position within a trial at which the patterns were presented within the stream was randomized across subjects, with the constraint of the mirror-configuration described above. The start of a new trial was self-paced to allow short breaks if necessary.

3.3 DATA ANALYSIS AND MODELLING

3.3.1 Pre-processing

Raw signals recorded by our delivery-response apparatus were pre-processed to extract stimuli and presses' onsets. All pre-processing steps were performed using custom codes written in Matlab (The MathWorks Inc.) and applied to each trial of each participant separately.

A stimulus onset was identified as a change of status of the photodiode, whose signal was recorded as a time series of 0 and 1, with 1 corresponding to the appearance of the visual trigger on the screen. Each positive derivative was classified as an onset and stored with its timestamp, to which a delay of 23 ms was added (as estimated by previous technical tests on sound delivery timing, see paragraph 3.2.3).

Concerning participants' tapping, the signal obtained from each trial was a time series of force values recorded from the beginning to the end of it (an example of tapping signal from one trial is shown in **Appendix B**, **Figure B1B**). Presses' onset detection was performed through a series of hierarchical steps. First, in order to get rid of mechanical noise-related fluctuations, the raw signal from each trial was smoothed using a 3rd order polynomial Savitzky-Golay filter (Savitzky & Golay, 1964), with a window of 15 samples (i.e. corresponding to roughly 150 ms). The standard deviation of the first 50 samples of each trial was calculated to obtain a baseline-noise standard deviation. As a second step, we calculated the mean force of the 150 highest values of the signal and, based on it, categorized it either as "Strong", "Medium" or "Weak" and assigned a trial-specific derivative-threshold (i.e. a positive difference between consecutive force samples representing a "jump" in the signal, large enough to be caused by a press). Based on this categorization, we defined also a category-specific standard deviation threshold (std-threshold) as a multiple of the baseline standard deviation, representing the minimum amount of increase in the standard deviation of neighboring samples indicative of a "press" event. These tailored thresholds allowed us to take into account the great variability observed between participants in the amount of force they exerted on the response pad.

Based on these threshold values, the third step comprised a series of inclusion/exclusion criteria to define the presence or the absence of a press onset. Specifically, samples were first filtered based on the presence of either a suprathreshold jump in the signal (i.e. a derivative greater than the derivative-threshold), or the cumulative presence of around-threshold derivatives in neighboring samples. For this second subgroup of potential onsets, the std-threshold was used as additional criterion to filter the candidate onsets. After this, all samples following a potential onset within a window of 100 ms were further excluded (i.e. they were considered either as signals' artifacts or "false" onsets, e.g. double taps). The final criterion was the presence of large enough negative derivatives between two candidate onsets, signaling a press' offset (i.e. return to baseline of force values).

As a result of the pre-processing stage, four participants had to be excluded due to technical issues in filtering their signal from baseline noise and reliably detecting presses' onsets in more than 25% of their trials. Hence the resultant sample used for the analysis is N=10.

From the pre-processed series, all stimuli and presses' onsets detected earlier than 350ms before the 5th stimulus onset of the Synchronization phase were removed, as the first pattern of each trial was a copy of the actual first experimental pattern and was added only to prompt the synchronization of the tapping. Tapping and stimuli onset series were then aligned by means of Dynamic Time Warping (DTW), a pattern matching algorithm that allows to pair time series of unequal length by computing the best warping path to minimize their distance (Sakoe & Chiba, 1978). Being the task a synchronization, this alignment allowed us to obtain an optimal one-to-one pairing between taps and stimuli and, consequently, detect and discard possible double taps (taps matched with the same stimulus) or missing taps (stimuli with no match).

The paired series were used to compute Inter Stimulus Intervals[1] (ISI), Inter Tap Intervals (ITI), ITI error and asynchronies. ISI and ITI were calculated as the intervals between two consecutive stimuli or taps' onsets, respectively. ITI error was calculated as the difference between matched ITI - ISI, with positive values indicating over-reproduction and negative values under-reproduction. Asynchrony was calculated as the positive or negative difference between paired stimulus - tap onsets, with negative values indicating an anticipated tap and a positive value a delayed tap (Repp & Su, 2013).

3.3.2 Behavioral analysis

As a preliminary assessment of participants' performance, we analyzed tapping data focusing on the first interval of all presentations of the patterns within a trial's stream. This choice was motivated by the fact that the 1st interval of our patterns is the most informative one from a learner perspective, as it allows us to discriminate between the two structures. Being the first interval of the pattern, its identity cannot be inferred from the intervals that preceded it. Therefore its reproduction is supposed to be mostly influenced by internal predictions about what will be the next pattern, to be more susceptible to errors caused by wrong predictions, and also to evolve according to the level of uncertainty of the trial. Hence the way the 1st interval is reproduced is indicative of the presence or absence of an ongoing learning and update.

To this purpose, we analyzed the error of the 1st ITI and the asynchrony of the 2nd tap onset, as they both can be referred to the reproduction of the first interval of the pattern (e.g. a positive asynchrony in the 2nd tap is the other side of an over-reproduction of the 1st ITI). To investigate how these measures differed between pattern A and B and across trial history and overcome the unbalanced occurrences of the two patterns within a single trial, we decided to bin the observations

[1] Throughout the chapter the term Inter Stimulus Interval (ISI) is used as a synonym of the term Inter Onset Interval (IOI), more commonly used in finger tapping and rhythm perception literature.

according to the number of times a given pattern had been presented, irrespective of the trial number (one bin = 31 presentations of a given pattern; e.g. bin 2 includes presentations from the 32nd to 62nd presentations of pattern A and B).

We then calculated bin-wise mean asynchronies and mean ITI errors for each participant and pattern. We tested pattern type and bin history effects by fitting two separate Linear Mixed Effects (LME) models, with asynchrony or ITI error as dependent variables, pattern identity and bin number as categorical and continuous predictors respectively, and participants as random intercept. LME models were fitted in R (R Core Team, 2020), using the packages *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017), and *MuMIn* package (Barton, 2020) for calculating R^2 statistics. Overall effects were tested using Type III Analysis of Variance with Satterthwaite's method for degrees of freedom. Interaction effects were tested using *emmeans* package (Lenth, 2020), using Kenward-Roger's method for degrees of freedom. Alpha level was always set at 0.05.

3.3.3 Model-based analysis

The general goal of the current model-based analysis is to describe and estimate the parameters governing how our participants learn to switch from a known to a novel temporal structure and how this transition is influenced by the predictability of the individual trial. Here we present two models that represent the first stage of a work still in progress that ultimately aims at a hierarchical and dynamic modelling of these parameters. In both models the series of observations whose likelihood is computed, consists only of the 1st ITI of the patterns, as this measure directly refers to the most informative interval of both structures.

Mixture model. The first model describes participants' behavior from a macroscopic and superficial perspective. It estimates, for each subject and each trial, the parameter θ that defines in which proportion pattern A and B are present in the reproduced series. The model is a simple mixture of two gaussians, with θ_A and θ_B (i.e., $\theta_B = 1 - \theta_A$) as mixture weight parameters. The likelihood is defined as:

$$p(y) = \theta_A \cdot \mathcal{N}(y \mid \mu_A, \sigma_A) + \theta_B \cdot \mathcal{N}(y \mid \mu_B, \sigma_B) \quad (1)$$

where y is the series of ITI, μ and σ define means and standard deviations of the two stimuli distributions (i.e., only the 1st ISI), and θ represents the proportion in which the distribution (i.e. pattern) with mean μ_i and standard deviation σ_i is present in the mixture. The posterior over each θ is estimated having as a prior a beta distribution with scale hyperparameters a and b, the estimation of which gives information about the uncertainty in the posterior of θ. This model was also run on ISI series as a sanity check to test whether the parameterized transition was accurately retrieved.

Static learning rate model. The second model increases the level of detail in the estimation of the proportions of the two patterns, as it is done at the single tap level (parameter p(A) and its complementary {1 - p(A)}). The model also increases the precision in describing the learning process, as it introduces two learning rate parameters (α_A and α_B) governing the evolution of the estimated proportions. A mixture of two gaussians defines each tap's likelihood formula:

$$p(y) = p(A) \cdot \mathcal{N}(y \mid \mu_A, \sigma_A) + (1 - p(A)) \cdot \mathcal{N}(y \mid \mu_B, \sigma_B) \quad (2)$$

where y is the nth tap, μ_A and μ_B are the known means of the two stimuli distributions (i.e., 1st ISI of the patterns), σ_A and σ_B are their unknown standard deviations and they are estimated based on an exponential prior with scale hyperparameter λ. The parameter p(A) represents the probability of the nth tap of being pattern A, whereas {1 - p(A)} corresponds to its counterpart for pattern B. The posterior over each $p(A)_n$ is estimated following a Rescorla-Wagner update equation:

$$p(A)_n = \begin{cases} p(A)_{n-1} + \alpha_A (1 - p(A)_{n-1}) & \text{if pattern}_{n-1} = \text{A} \\ p(A)_{n-1} - \alpha_B p(A)_{n-1} & \text{if pattern}_{n-1} = \text{B} \end{cases} \quad (3)$$

where $p(A)_n$ is the probability of the nth ITI of being pattern A; $p(A)_{n-1}$ is the previous ITI estimated probability; α_A and α_B are the learning rates ruling the update. The lower their values, the more static will be p(A) estimation over time; the higher their values, the more frequently p(A) will be updated, based on the identity of the previous pattern. While the posterior over p(A) is estimated at the tap level, the two learning rates are assumed as fixed across trials, hence the model estimates only one α_A and one α_B for each participant.

To test whether learning rates correlate with simple behavioral measures, we calculated a subject-wise slope coefficient for both patterns describing the evolution of the 1st ITI error across trials and computed its correlation with estimated α_A and α_B parameters.

As part of the ongoing model optimization process, we performed a subject-wise parameter recovery to evaluate the adequacy of this model in describing our participants' performance and learning. To this aim, several synthetic datasets per participant were generated by the model using his/her specific estimated parameters. From the 2500 simulated dataset we randomly picked 20 per participant and fed each of them to the model to retrieve new parameters estimates, that were compared with the generative parameters.

Both models were implemented in Stan using R as the interface (Stan Development Team, 2020, 2021, http://mc-stan.org; Carpenter et al., 2017). Stan uses a Markov Chain Monte Carlo sampling method to estimate posterior distributions over model parameters. For each participant (and each trial, in the case of the first model) we ran the models with four independent chains and 5000 iterations (2500 warm-up). Convergence of sampling chains was estimated through the Gelman-Rubin $\hat{R}$ statistic (Gelman & Rubin, 1992), keeping as acceptable threshold $\hat{R}$ values smaller than 1.01. Correlation analysis on behavioral measures and learning rate estimates were performed in

R (R Core Team, 2020), using package lme4 (Bates et al., 2015) to estimate slope coefficients and R built-in function for Spearman's rank correlation.

3.4 RESULTS

3.4.1 Behavioral analysis

Preliminary analysis on tapping measures were conducted to assess simple effects of our experimental manipulations on performances. In particular, we analyzed each pattern's mean ITI error of the 1st interval (**Figure 3.2**) and mean asynchrony of the 2nd tap onset (**Appendix B**, **Figure B2**), in order to focus on the most informative part of the patterns.

We used the measures as dependent variables in two LME models, with pattern identity and presentation bin history as independent variables, and participants as random intercept (model formula: measure ~ pattern identity * bin number + (1 | subject)). Goodness of fit of both models was evaluated using R^2 statistic ($R^2_{marginal}$ = 0.70, $R^2_{conditional}$ = 0.76 for ITI error fit; $R^2_{marginal}$ = 0.33, $R^2_{conditional}$ = 0.75 for asynchrony fit). Type III ANOVA on the two models' estimates showed overlapping results: significant main effects of pattern identity ($F(1, 387) = 526.208$, $p<0.0001$ for ITI error; $F(1, 387) = 236.459$, $p<0.0001$ for asynchrony) and bin history ($F(1, 387) = 92.683$, $p<0.0001$ for ITI error; $F(1, 387) = 103.93$, $p<0.0001$ for asynchrony), as well as a significant interaction between pattern and bin history ($F(1, 387) = 71.434$, $p<0.0001$ for ITI error; $F(1, 387) = 45.903$, $p<0.0001$ for asynchrony). In terms of behavioral effects, these results indicate that ITI error values differed significantly between patterns, with a tendency to over-reproduce the 1st interval of pattern A and under-reproduce the 1st of pattern B (see **Figure 3.2**). This difference slowly decreased across bin history, mainly because ITI error in pattern B decreased significantly throughout trials and reached values closer to 0 towards the 20th bin. Whereas the over-reproduction of pattern A remained almost constant across bins. This last result could explain the difference in slope coefficients for the interaction term of the model (A x bin: slope = 0.000593; B x bin: slope = 0.00912; A - B = - 0.00853, t.ratio(387) = - 8.452., $p<0.0001$). Concerning asynchrony values, the effects reported can be summarized as a general greater negative asynchrony for pattern B compared to A and a common trend towards a reduction of the asynchrony across bin history. This trend, though, was significantly stronger for pattern B than A, as revealed by the difference in slope coefficients for the interaction term of the model (A x bin: slope = 0.00145; B x bin: slope = 0.0072; A - B = - 0.00575, t.ratio(387) = - 6.775., $p<0.0001$).

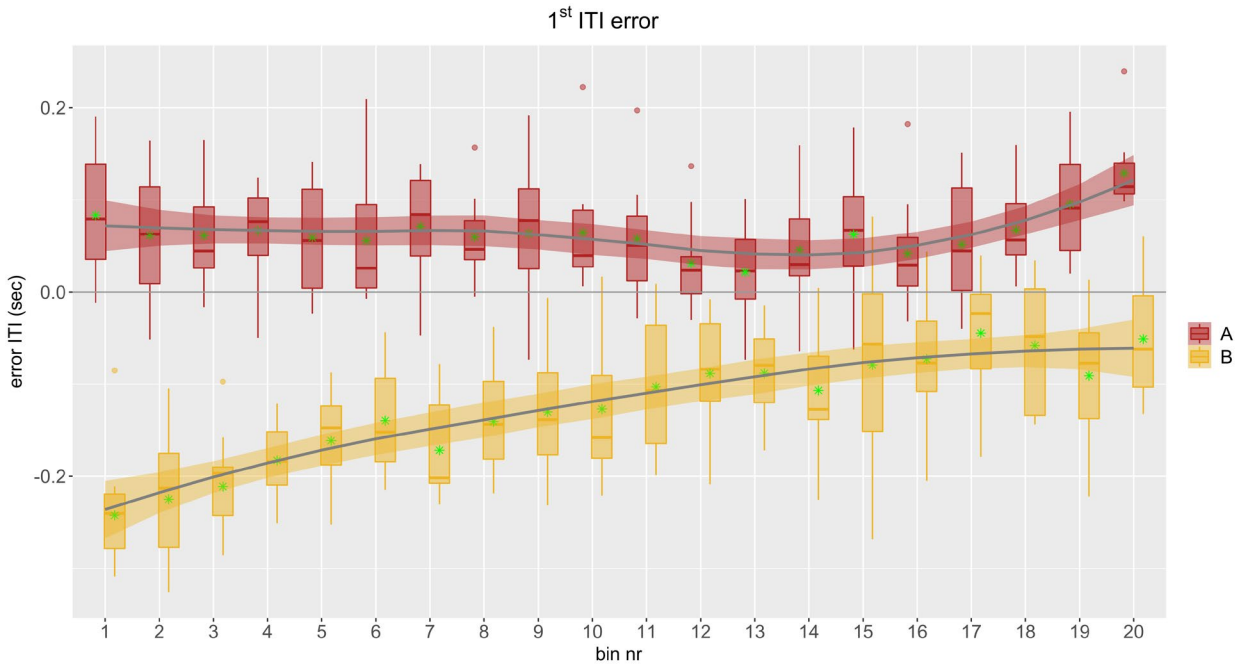


Figure 3.2. **Mean ITI error of 1st interval of the patterns**. The plot shows the evolution across trials of the ITI error of the 1st interval of each pattern (color code: pattern A = dark red, pattern B = yellow). Pattern A is characterized by a mild overreproduction, stable across trials, whereas pattern B shows a large underreproduction that progressively reduces. Statistical analysis confirmed the presence of a trial history effect for pattern B as a steeper slope coefficient (here graphically represented using LOESS method).

Overall these results reveal that participants quickly learned to synchronize with pattern A, as suggested by smaller ITI error and smaller negative asynchrony than in pattern B throughout the task. As its probability of appearing became lower in the last group of trials, synchronization with pattern A was characterized by a mild worsening (greater over-reproduction of its 1st interval and near above-zero asynchronies). On the contrary, synchronization with pattern B was characterized by greater ITI error in the opposite direction (under-reproduction) and large negative asynchronies not only in the trials where its presence was unlikely, but also in subsequent trials where its probability of appearing was the same as pattern A. Nevertheless, pattern B showed a constant steep improvement in error and asynchrony throughout all trials, and reached comparable levels of accuracy of pattern A in the second part of the task. This improvement was probably due to the fact that throughout the trials participants learned its structure and expected its presence.

3.4.2 Model-based analysis

As outlined in the method section, the first and the second models implemented in the present study are the preliminary stages of an ongoing work aimed at formalizing and estimating the parameters underlying the learning of new temporal structures, defined as the transition from a known to a novel one, and the role of the volatility of the environment in modulating this process. Here we present the results from these two models.

Mixture model. The first model is also the simplest: it gives a general, superficial picture of participants' performance by quantifying trial by trial in which proportions the two patterns are present. This quantification is estimated by means of a mixture of two gaussians, with parameters θ_A and θ_B representing the mixture proportions of pattern A and pattern B, respectively. A measure of uncertainty in this estimation is given by a beta prior on each θ, whose scale hyperparameters a and b are estimated by the model as well.

A sanity check of estimation accuracy was done by feeding the model with each trial's stimuli series (i.e. 1st ISI of the patterns). Results showed that the parameterized transition embedded in the series was accurately retrieved by the model and the trial-specific level of predictability was reflected in the shape of the two θ prior distributions, obtained from the estimation of hyperparameters a and b (see **Appendix B**, **Figure B3A and B3B**).

Concerning tapping performances, parameters estimation was done subject- and trial-wise using the 1st ITI of each occurrence of the patterns. A representative example of model results from one participant is shown in **Figure 3.3A** for the θ parameter and in **Figure 3.3B** for the beta priors over θ. The evolution of estimated θ_A and θ_B reveals that this participant was capturing with a good approximation the distributions of the two patterns across trials, but the switch between the two θ happened later compared to the stimuli, meaning that his tapping was more likely to be synchronized with pattern A even when its probability to appear was equal to pattern B. The estimation of the beta priors over θ highlights this delay even more and it is characterized by greater uncertainty compared to the stimuli, as revealed by the large overlap between the two distributions almost until the last trials. Results at the group level confirm the trend observed for this participant. **Figure 3.4** shows a summary plot of θ_A and θ_B estimation, where the delayed transition to pattern B is very clear. This delay may represent either a carryover effect of pattern A on performance, or the result of middle trial's low predictability which may bias the tapping towards the more familiar pattern. The prolonged overlap in θ_A and θ_B estimated prior distributions is present at the group level too (**Appendix B**, **Figure B3C**), thus confirming the greater uncertainty compared to the distributions estimated from the stimuli series.

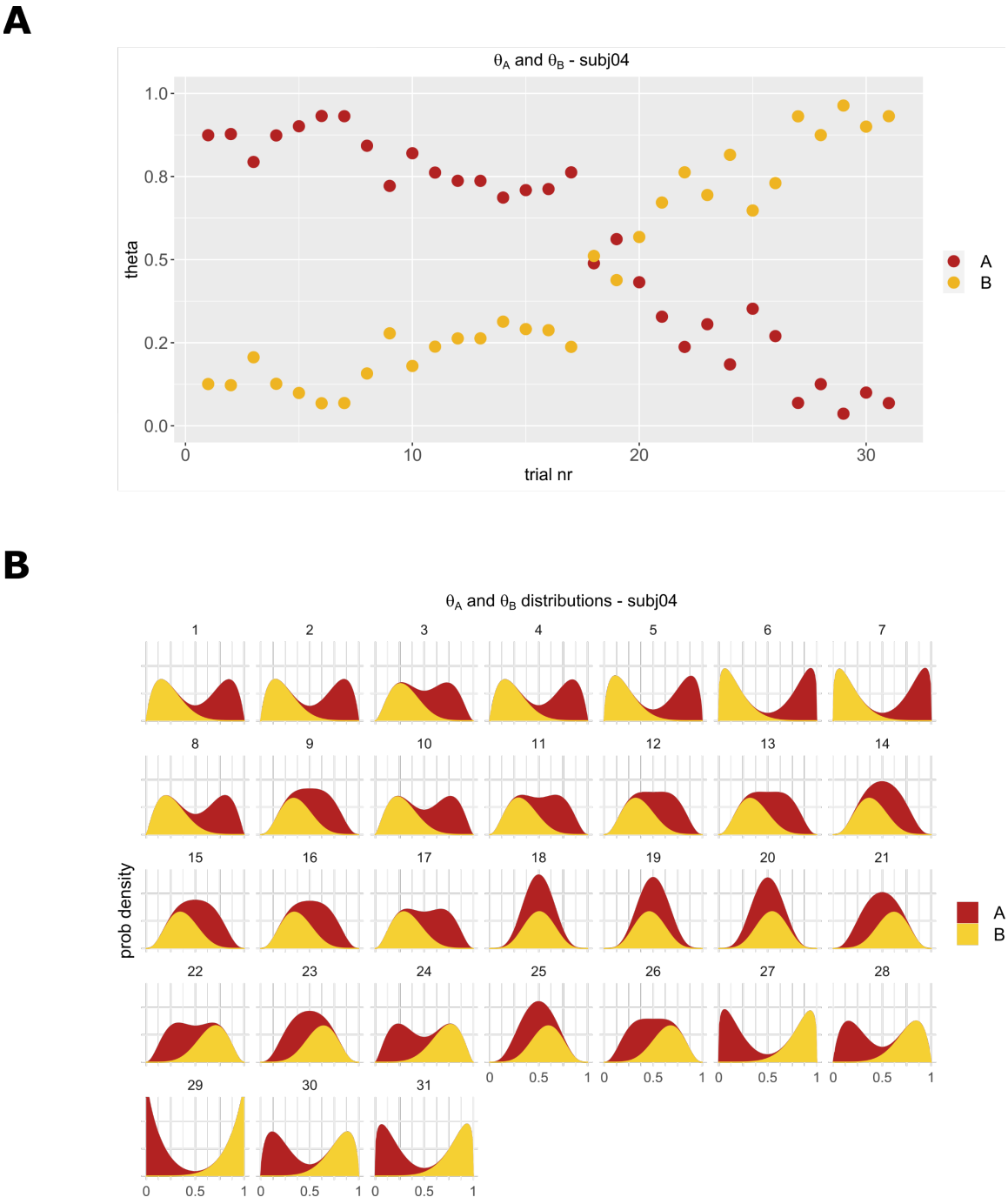


Figure 3.3. **Mixture model: parameters estimation**. (A) Estimated θ_A and θ_B (color code: θ_A = dark red, θ_B = yellow) across trials for participant 04: the values represent the proportions of pattern A and B across trials in the reproduction of this participant, as estimated by the first model. (B) Beta priors on θ_A and θ_B across trials for participant 04, as the result of the estimation of scale hyperparameters a and b for each θ, color-coded as in (A). The overlap between beta distributions is an indication of the level of uncertainty in the estimation of parameter θ.

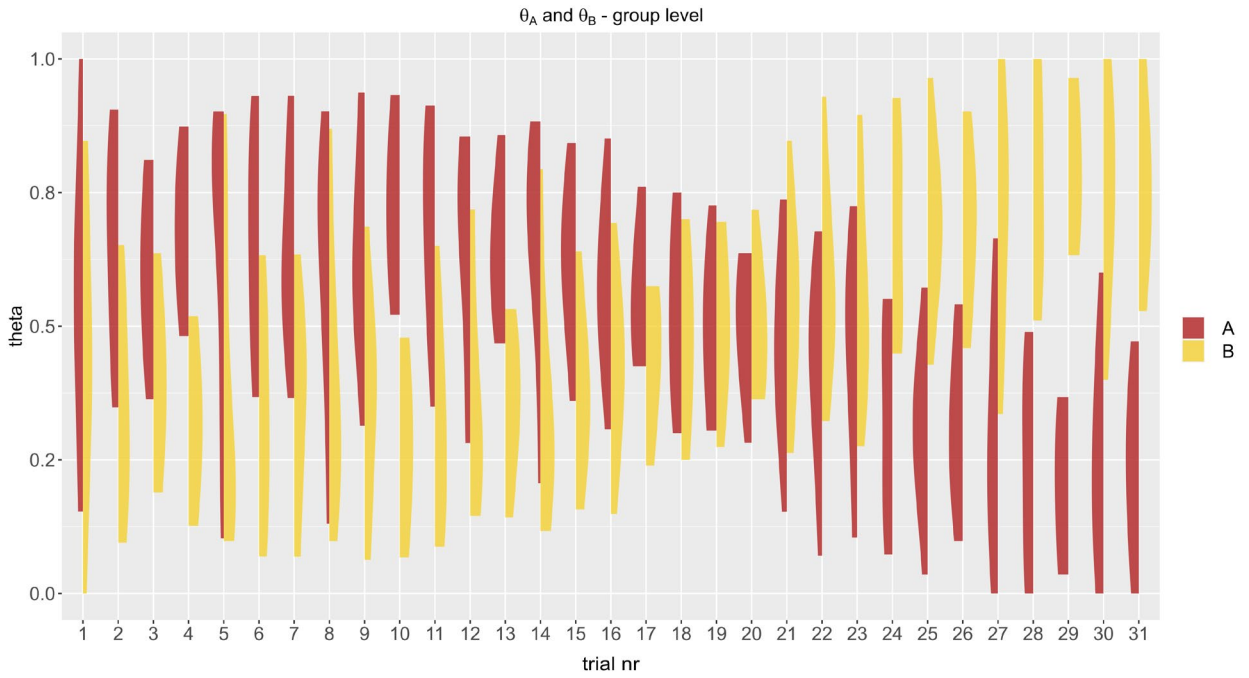


Figure 3.4. **Group-level plot on estimated parameter θ.** Estimated θ_A and θ_B (color code: θ_A = dark red, θ_B = yellow) across trials at the group level: the plot represents a graphical summary of all theta values estimated at the subject-level, grouped by pattern identity and trial number.

Static learning rate model. The second model deepens the estimation of the proportions of pattern A and B in participants' performance both in terms of temporal resolution and of definition of the underlying process governing their evolution in time. The model, in brief, defines each tap performed as a mixture of two gaussians, whose weight parameters (p(A) and {1 - p(A)}) are updated based on static learning rates (α_A and α_B). Parameters estimation was done subject-wise, using the 1st ITI of each pattern presentation throughout the trials. Mixture weight parameters were estimated tap-wise, whereas a unique value was estimated for their respective learning rates and for the standard deviations of the two gaussians.

Results for parameter p(A) estimation are shown for all participants in **Figure 3.5**. In general, the estimation follows the delayed-switch trend already seen in the first model's estimation of parameter θ. Participants tended to produce more often pattern A, even after the point of transition in the stimuli, and switched to pattern B only in trials in which its presence was more frequent. Graphical inspection of p(A) estimation suggests the presence of two subgroups of performers: "fast learners" and "slow learners". In the first group (e.g. subject 7) p(A) estimates quickly saturate to 1 and remain stable, except for little perturbations most likely linked to the appearance of the other pattern. Values of p(A) start to drop around the time zone of the switch, and reach very low probability values only towards the tail of the tap series. In the "slow learners" group (e.g. subject 9) estimates of p(A) take quite a large number of samples to reach high probability values,

meaning that during all the first part of the task the tapping was either very noisy or synchronized with some form of internal reference. Once the estimates approach high values (at least higher than chance level), they tend to drop more slowly compared to the other group, and never reach 0. In general, the estimates for this group appear less sensitive to the appearance of pattern B (absence of perturbations in the estimates) and to the actual proportions of pattern A and B changing trial by trial. This consideration is supported by the differences in estimated learning rates among the two subgroups (see **Appendix B**, **Figure B6B,** where α_A and α_B are plotted against estimations obtained from parameter recovery): "slow learners" have also very low learning rates for both patterns, compared to the other group. This means that the estimation of p(A) was relying most of the time on previous taps and was very weakly influenced by the two learning rates.

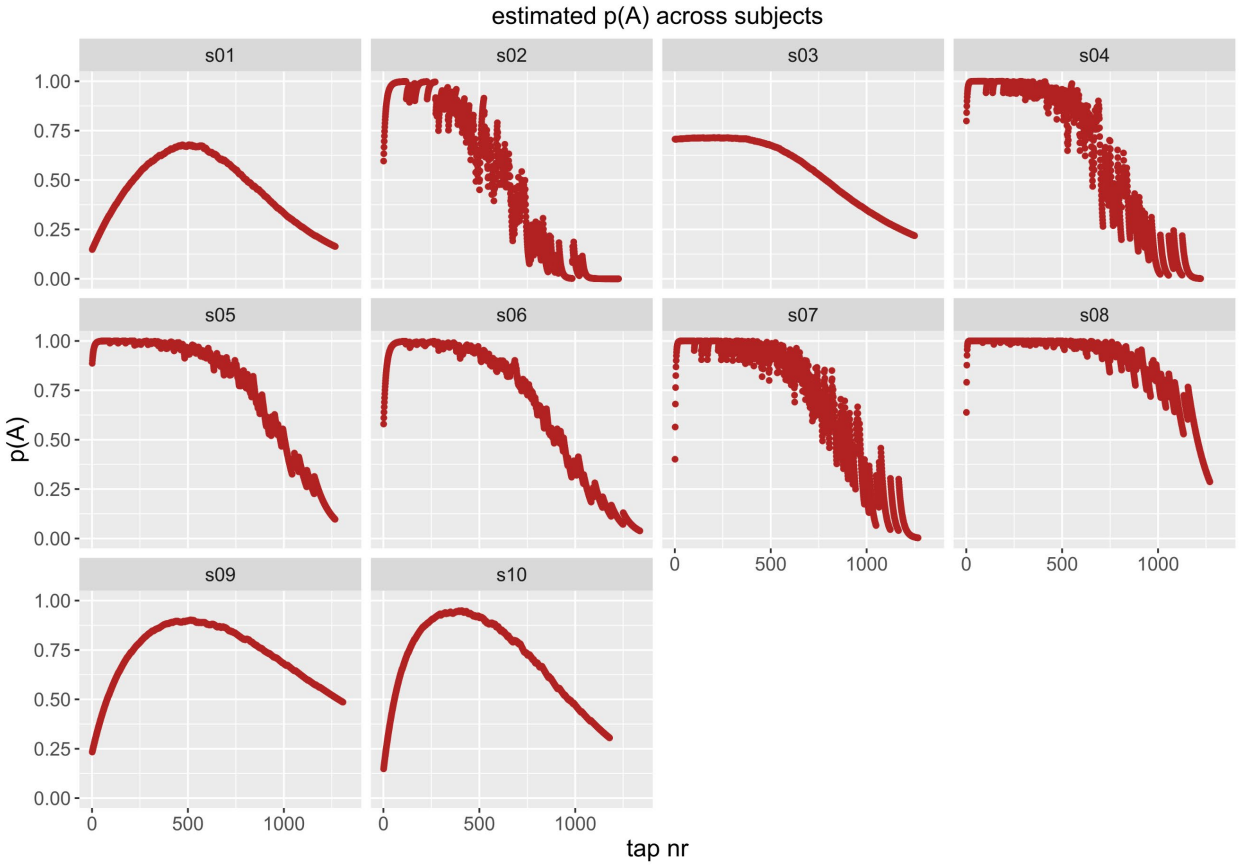


Figure 3.5. **Static learning rate model: parameter p(A) estimation**. Each plot shows the subject-wise estimation of parameter p(A), representing the probability of reproducing pattern A tap by tap (i.e. for pattern B {1 - p(A)}). The two subgroups of participants described as "fast learners" and "slow learners" can be spotted by graphical inspection, in terms of point in time when p(A) tends to saturate, steepness of the drop in its probability values in the second part of the tap series, small perturbations in estimation values related to the appearance of pattern B.

To assess whether model estimates on α_A and α_B were related to the ITI error measure derived from the 1st interval of the patterns (i.e. see behavioral analysis), we calculated a slope coefficient describing its evolution across trial history for each participant and pattern (fitting formula: ITI error ~ 0 + pattern identity * bin number + (0 + pattern identity: bin number | subject) + (1 | subject)). We then computed Spearman's rank correlation between slope values and learning rate estimates (**Appendix B**, **Figure B4**). A positive correlation was found only between pattern B slope coefficients and α_B (α_A - error A: $r(8) = -0.285$, $p = 0.427$; α_B - error B: $r(8) = 0.83$, $p < 0.01$). This result suggests that steeper reductions over time in ITI error for pattern B are associated with higher learning rates. The absence of correlation between slope coefficients for pattern A and α_A was expected, given the almost flat slopes obtained for all subjects. Considering also the small sample size, this correlation must be confirmed by further analysis.

Parameter recovery. A subject-wise parameter recovery was performed to evaluate how well the static model parameters describe our participants' behavior and can be accurately retrieved from simulated data. For each subject we randomly selected 20 synthetic datasets among those generated by the model using the parameters estimated from the original data. Each synthetic dataset was fed to the model to obtain new estimates of the parameters. For the portion of participants labelled as "fast learners", the model was able to generate synthetic data that correlated well with the original performances, whereas for the other subgroup the correlation was very low (**Appendix B**, **Figure B5**). However, in all simulated performances there was a substantial reduction in variability compared to the real data. Results from simulation-based estimation of p(A) showed overall a good consistency with the original values (**Appendix B**, **Figure B6A**). However, estimation of learning rates α_A and α_B and mixtures' standard deviation σ_A and σ_B from the synthetic data revealed some important differences compared to original parameters' estimates (**Appendix B**, **Figure B6B-C**). For the subgroup of "slow learners" learning rate values were extremely low as the generative parameters, but for the other group the values were highly variable (more for α_A than α_B), and in some cases they diverged to a great extent from the original estimation. Concerning σ_A and σ_B estimated from simulated data, their values were overall lower than the original parameters for almost all subjects, in accordance with the lack of variability already spotted in the synthetic data. Parameter recovery results showed estimation differences that appear mostly related to two main limitations of the static model. The first concerns its static formalization of learning rate parameters: given the shape of the transition embedded in the stimuli and the different level of predictability that trials have, a dynamic formalization of α_A and α_B seems a reasonable way to optimize the model. The second limit has to do with the large amount of variance that the model is not able to capture, thus representing another aspect to be optimized. These limitations are further examined in the discussion session.

3.5 DISCUSSION

In the present study we investigated how humans learn new temporal structures, in particular the process by which they learn to switch from a known to a novel temporal pattern, made of the same elements but having a different structure, and how this switch is influenced by trial predictability. To this aim we embedded a gradual transition between two symmetrical auditory temporal patterns, by manipulating their probability of occurrence across trials, and presented them to our participants in a synchronization task. We carried out a model-based analysis aimed at defining and estimating the parameters describing the learning process underlying this transition in participants' tapping performance. The results presented here are from the first two models implemented in this analysis. The first model gives a trial by trial estimate of the proportions in which the two patterns were present in tapping performances, whereas the second model brings this estimation at the single tap-level and includes a weighted update of the proportions of the two patterns according to static learning rates.

Despite both models represent only incomplete steps towards the learning process we aim to fully describe and parameterize, their results already offer meaningful insights on it, also in combination with the behavioral analysis we conducted. In particular, behavioral measures and parameters estimation from the first model suggest that participants relatively quickly learned to synchronize to pattern A. This was expressed as lower ITI error and better asynchrony compared to pattern B throughout the task, as well as in the estimation of its proportion across trials (i.e. parameter θ). Pattern B was instead performed with lower accuracy even beyond middle trials, and θ estimation from tapping performances showed a delayed switch in patterns' proportions compared to the stimuli (i.e. the carry-over effect). Combined together, these results seem to suggest that pattern A became a sort of baseline knowledge on what to expect from a trial and what to refer to when the stimuli heard were too unpredictable. Once the presence of pattern B became frequent and stable enough within the trial, θ estimation showed the switch in patterns' proportions and, in parallel, ITI error and asynchronies of pattern B improved. The greater level of temporal detail in the estimation of the second model allowed us to recognize two different types of performers within our sample of subjects. One group ("fast learners") quickly learned to synchronize to pattern A (i.e. immediate saturation of p(A) estimates values). This group appeared more sensitive to the ongoing changes in patterns' proportions (i.e. higher learning rates) and they showed a sharper transition to pattern B (i.e. clear drop in p(A) values and higher slope coefficient in ITI error decrease), even if delayed compared to the stimuli. A second group ("slow learners") seemed to need on average more time to learn to synchronize to pattern A (i.e. slower increase in p(A) estimates), tended to be less sensitive to trials' variability and stick to pattern A longer than the other group (i.e. slower decrease of p(A) estimates). This last group seemed to rely more on previous taps (i.e. very low learning rates) rather than to the pattern's changes across trials. This behavior thus resulted in a less steep and on average slower transition to pattern B. It is important to emphasize here that the sample size used for the analysis was relatively small. It is therefore

likely that with a bigger subjects' sample, the two subgroups described here may turn out to be more than two, or instead to be part of the same distribution of performances.

Even though the focus of this study is on temporal structures' learning in general, either rhythmic or not, our preliminary results can be partially linked to a group of experimental works and theoretical accounts of rhythm processing in the context of predictive coding framework (Vuust et al., 2009; Vuust & Witek, 2014; Koelsch et al., 2019). According to this view, tracking and synchronizing with rhythms can be considered a highly predictive process that is hierarchical in nature, and can be described in terms of optimal Bayesian inference based on a continuous comparison between prior expectations (i.e. metrical structure, top-down) and sensory experience (i.e. rhythmic patterns, bottom-up). A study that is noteworthy (Elliott et al., 2014), for example, applied a mixture model with flexible number of components (either one or two) to synchronization data from a task in which participants had to tap along with two simultaneously presented metronomes of equal mean period, but slightly differing either in phase or in standard deviation of the period. Essentially, to "solve" the task participants had to either separate the cues or integrate them. The authors showed that asynchronies data were better accounted for by a Bayesian inference-based mixture model that switches between separation (two mixture components) and integration (one component) of cues according to the degree of phase-offset between metronomes. And they showed that the point of switch between separation and integration is subject-specific. We could hypothesize that in our task a similar flexible tradeoff between relying on the external stimulus and on an internal representation is required. The additional effort for our participant is to adjust this internal reference as soon as the presence of the new pattern becomes more stable.

Limitations and model optimization. Parameter recovery results shed light on some important limitations of the second model that are currently being optimized. Specifically, the model estimation of learning rates α_A and α_B from synthetic data showed large variability and didn't approximate the values of the generative parameters, except when the original estimates were close to zero. The static formalization of learning rates appears problematic also considering the dynamic nature of the experimental design: across trials the two patterns' proportions vary following a logistic shape (i.e. in a non-linear fashion), hence an ideal performer would need to rely on a likewise flexible and dynamic update of his internal representation of patterns' proportions to synchronize to the stimuli. Model optimization must include this flexibility by implementing dynamic learning rates that would reflect trial-specific level of uncertainty.

Another issue of the second model is represented by its inability to capture an important amount of variance of the original data, as was evident in the calibration plots. This may be due to several reasons, the first of which has to deal with the estimation of σ_A and σ_B. Parameter recovery showed that simulation-based estimates were always lower compared to the original parameters, and for some participants the difference between them was also quite big. A better formalization of σ_A and σ_B should consider models of motor timing variability, starting from the simplest and most known, i.e. the generalized form of Weber's law (Wing & Kristofferson, 1973; Getty, 1975; Ivry &

Hazeltine, 1995), or more recent development of its original form (Laje et al., 2011; García-Garibay et al., 2016).

In relation to the issue of modelling tapping variability, a confound and additional source of variability neglected by our models, comes from the specific task used in this study. When synchronizing with an external reference, humans typically adopt error-correction mechanisms to re-adjust to the stimulus and compensate the inherent variability of their motor actions. This phenomenon has been extensively studied in sensorimotor synchronization literature (for a review, see Repp, 2005; Repp & Su, 2013), and different models in the field have been proposed to account for it. The most popular being linear, event-based models (e.g., Schultze et al., 2005), or non-linear, dynamic system models (e.g., Large & Jones, 1999). These models are usually tested in tasks where error-correction is elicited by phase or period perturbations of a regular base-rhythm, as already mentioned in the introduction of this chapter. The peculiarity of our experimental manipulation, i.e. the transition from one temporal pattern to another across trials, resulted in a quite atypical synchronization task whose stimuli are very different from those generally used in sensorimotor synchronization studies. Incorporating such error correction accounts to optimize the model may represent a difficult endeavor, though it should be considered carefully.

Finally, another possible improvement to the model may come from including in the analysis the full dataset, i.e. all the ITI of the patterns. Even if our choice of focusing on the first interval confirmed to be a highly informative selection, the other three intervals may still add useful information to understand how our participants were tracking trial events and adapting to them.

3.6 CONCLUSION AND FUTURE DIRECTIONS

In the current study we asked how humans switch from a known temporal pattern to a new one which shares the same constitutive elements but differs in its structure, and how this transition is influenced by trial predictability. To this aim we parameterized a transition between two mirroring auditory temporal patterns, by manipulating their probability of occurrence across trials, and presented them to human participants in a synchronization task.

Model-based analysis on tapping data allowed to estimate participants' probability of reproducing one pattern or the other both at the trial- (first model) and single tap-level (second model). The second model also increased the depth of description of the ongoing learning process by estimating tapping variability and the learning rates governing the update of patterns' proportions. Despite the acknowledged limitations, the models implemented were able to retrieve the learning transition that our participants experienced during the task and to categorize different types of performance. These results were supported by behavioral analysis on tapping errors and asynchronies. Model optimization will include the presence of dynamic learning rates, estimated trial by trial, and a more precise formalization of tapping variability.

Chapter 4

Learning the statistical properties of temporal patterns

ABSTRACT

Series of discrete highly regular sensorimotor events are often experienced as temporal patterns. Studies on rhythm perception, sensorimotor learning and predictive coding in the auditory domain have shown that humans learn, are highly sensitive to and form expectations about the temporal regularities of the environment. A complete understanding of the basic mechanism supporting these abilities is far from clear.

Here we investigated the process through which humans learn the statistical properties of temporal sequences of auditory stimuli (empty intervals marked by brief tones). Specifically, we generated sequences using three first-order Markov processes, whose possible states were represented by three durations. According to the underlying generative process (i.e. experimental condition) these durations could be presented following random or predictable transition probabilities, leading to different mean duration and entropy rate of each process' stationary distribution. The generated sequences were presented to human participants in a delayed temporal reproduction task. Behavioral performances were analyzed combining Bayesian modelling and data-driven approaches, with the aim of retrieving the statistical rules underlying participants' learning and comparing them with stimuli generative processes.

Results showed that participants' performance was linked to the set of statistical rules defined by our experimental conditions, but also reflected other types of low- and high-level statistics, namely central tendency and block history, along with a shrinkage of the durations-states space. These results suggest that the general human ability of detecting statistical regularities in the environment can serve as a basic mechanism for the learning of a broad range of temporal patterns, either partially or fully predictable, like musical rhythms.

4.1 INTRODUCTION

Our experience of the world is constantly marked by many levels of temporal regularities, spanning different timescales and involving different kinds of temporal features, like the order of the events, their durations, and how these durations relate to each other. A successful and adaptive interaction with the environment requires flexible and sharp abilities in detecting and learning such regularities, no matter whether you are a musician playing in a jam session, a bird recognizing the pattern of the call of a conspecific, an athlete performing temporally fine-tuned motor sequences, or simply a tourist trying to extract some meaning from the temporal features in the prosody of foreign linguistic sounds.

Humans' ability to detect, learn and exploit temporal regularities has been investigated from different perspectives. Temporal regularities constitute an essential knowledge our brain uses to anticipate events and have been investigated, for example, in the form of temporal expectations (for a review see Nobre & van Ede, 2017). A related perspective comes from studies on auditory scene processing, where it has been shown that the human brain is highly sensitive to both unexpected omissions in ongoing temporally regular sound sequences and to temporal structures arising from randomness (Andreou et al., 2015; Barascud et al., 2016).

A special case of temporal regularities is represented by musical rhythms, and rhythm-like patterns in general, which are characterized by fixed and fully predictable relationships between intervals (the uncertainty is close to zero), by many levels of hierarchically organized regularities (i.e., rhythm, beat, meter, for a review see Grahn, 2012) and they are heavily influenced by culturally determined priors. To this regard, a couple of studies demonstrated how these rhythmic priors can easily be triggered even in the complete absence of temporal structure (Ravignani et al., 2016; Jacoby & McDermott, 2017). In particular, Ravignani and colleagues (2016) used a transmission chain paradigm in which first-generation participants were exposed to a random series of intervals and their reproduction of the series served as training for second-generation participants, and so on until the last generation. Their results showed that the series reproduced featured well known rhythmic priors (i.e., rhythmic universals), despite the complete absence of structure in the original stimuli.

Beside musical rhythms and similar highly deterministic patterns, there are other types of temporal structures that come with some degree of uncertainty, where for example the temporal identity of an event can make the next one more likely than another one, but not certain. In this framework, temporal pattern learning can be conceived, in its most general and basic form, as the learning of statistical regularities embedded within a series of durations. These regularities can be for example the mean of the intervals experienced, their frequency distributions, their entropy rate, their long-range correlations, the probability of an interval to be followed by another. Such probabilistic framework has been proved successful in accounting for many forms of sequence learning in humans (e.g., Dehaene et al., 2015; Frost et al., 2015) and it has also been tested adopting Bayesian-inference approaches (e.g., Meyniel et al., 2016; Maheu et al., 2022). The study by Maheu and colleagues (2022), in particular, tested experimentally and computationally whether humans employ different statistical inference processes according to the structure of the sequences they are tracking: one characterized by probabilistic relationships between events and the other one following deterministic rules. The authors of this work were able to show that humans can adaptively switch between these two structures while processing sequences and participants' behavior can be accounted for by a unique hierarchical Bayesian inference mechanism.

The effectiveness of such probabilistic approach in describing the basic mechanisms of temporal pattern learning is relatively unexplored. In the current study we adopted this approach to investigate whether and how humans extract and learn low- and high-level statistical properties underlying sequences of temporal intervals and modelled this learning process in terms of optimal

Bayesian inference. More specifically, instead of having fixed temporal patterns, we defined three generative processes based on which we created multiple sequences of temporal intervals. These generative processes (i.e. our experimental conditions) were described by specific sets of transition probabilities and led to different mean duration and entropy rate of their respective stationary distributions. We presented the generated sequences to human participants in a delayed temporal reproduction task and used an ideal observer model to retrieve the parameters defining the statistical properties participants learned.

4.2 METHODS

4.2.1 Participants

Twenty-four participants (10 males; 2 left-handed; mean age 26 ± 4.3; age range 19-37) were recruited for the present study. Nine participants reported some form of musical expertise, either as amateur players (N=6, 4.6 years on average) or as a result of formal musical training (N=3, 11 years on average). One of the participants turned out to be on psychiatric medication at the moment of the experiment, hence was excluded from the sample.

Before starting the experiment, each participant gave written informed consent to take part in the study, which was approved by the SISSA Ethics Committee (protocol number 11175-III/13), and complied with the Declaration of Helsinki. All recruited participants were naïve to the purposes of the study and received a fixed monetary compensation after completing the task.

4.2.2 Experimental Design

The experimental conditions (see **Figure 4.1C**) are three 1st order Markov processes, each described by a 3-by-3 Transition Probability Matrix (TPM) and in which the possible states are represented by three durations (D_1, D_2, D_3). In the what we called *Random* condition the transitions from one duration to the other are equally likely (for example, $p(D_1|D_1) = p(D_2|D_1) = p(D_3|D_1)$), the three durations are equally distributed so that the general mean of the Markov Chain (MC) stationary distribution lies around the center of the three durations (μ = 550 ms), and its entropy rate is maximal (H = 1.58 bit). Mutual information is equal to zero. In what we called *Pattern 1* and *Pattern 2* conditions instead there are more likely transitions between two of the three durations, in a mirrored way. In Pattern 1 there are more likely alternations between D_2 and D_3 (i.e. $p(D_3|D_2) > \{p(D_1|D_2) = p(D_2|D_2)\}$ and $p(D_2|D_3) > \{p(D_1|D_3) = p(D_3|D_3)\}$) and in Pattern 2 there are more likely alternations between D_1 and D_2 (i.e. $p(D_2|D_1) > \{p(D_1|D_1) = p(D_3|D_1)\}$ and $p(D_1|D_2) > \{p(D_2|D_2) = p(D_3|D_2)\}$). This leads to a bias in the frequency distributions of the three durations and a shift in the general mean of their respective MC stationary distributions, i.e. a greater mean in Pattern 1 (μ = 624 ms) and a smaller mean in Pattern 2 (μ = 464 ms). Since Pattern 1 and 2 mirror each other, the two Pattern conditions have equal entropy rate (H=1.40 bit), which

is instead lower than the entropy rate of the Random condition. In both Patterns 1 and 2 mutual information is greater than zero.

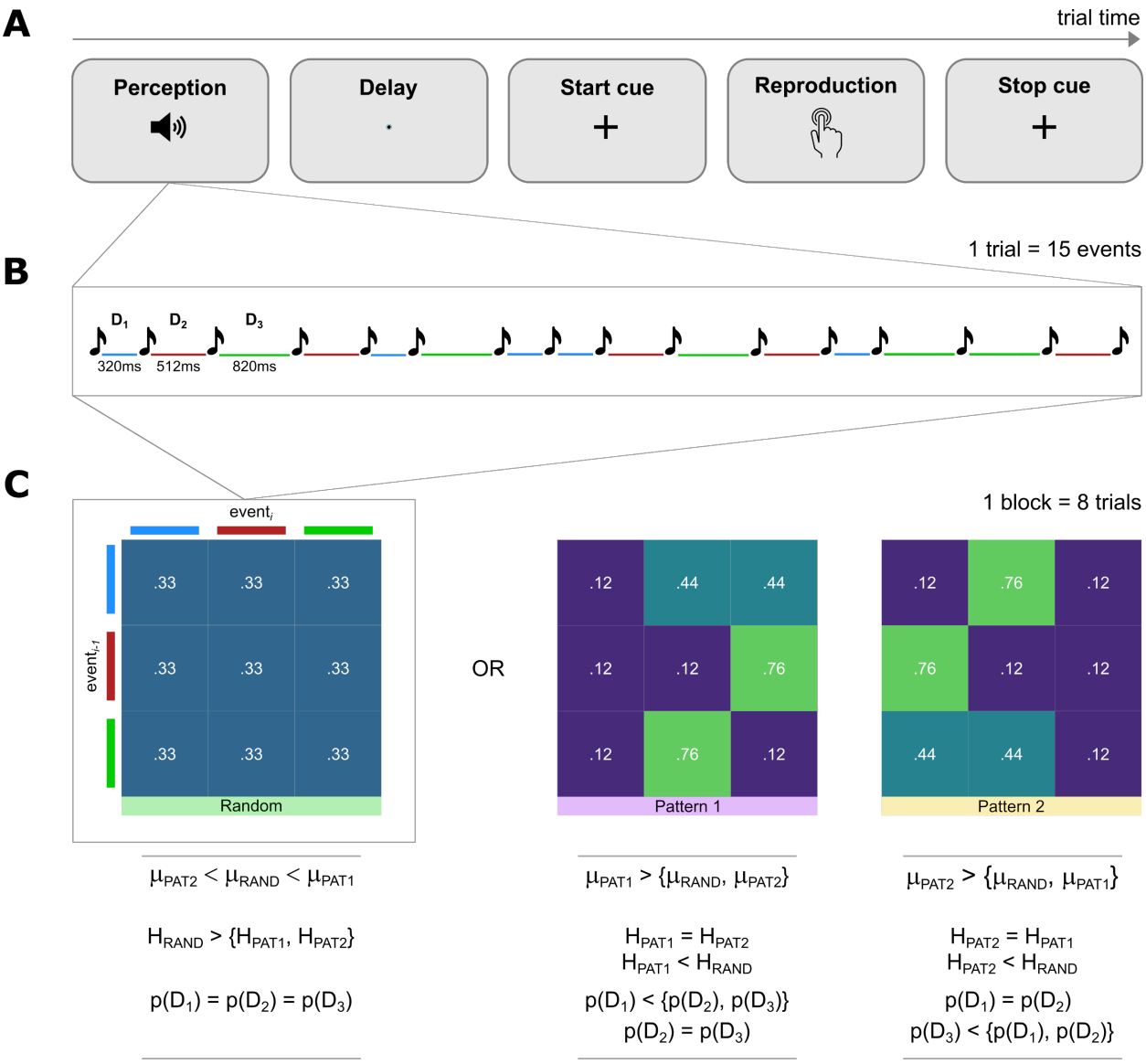


Figure 4.1. **Experimental conditions and trial structure**. (A) Delayed temporal reproduction task: in each trial a series of temporal intervals is presented and, after a brief random delay, a Start cue alerts the participant to start the reproduction of the series. A Stop cue signals the end of the trial. (B) In each trial a subset of 15 events/durations of a generated sequence is presented to the participant. The entire sequence contains 120 events and constitutes one block. Each block-sequence is split in 8 trials. The durations alternating in the sequence are three empty intervals (color coded as blue = 320 ms, red = 512 ms, green = 820 ms) marked by brief tones. (C) Each block-sequence is generated from one of the three 1st order Markov processes that represent the experimental conditions (color coded as green = Random, purple = Pattern 1, yellow = Pattern 2). Each process differs from the other two in terms of the transition matrix ruling the alternations between durations, with mirroring configurations for Pattern 1 and Pattern 2 conditions. Each matrix leads to a specific mean duration, entropy rate (equal for Pattern 1 and 2) and item frequency of its respective process' stationary distribution.

The choice of using mirroring TPM for the two Pattern conditions allows us to have a comparable level of structural complexity in the generated sequences, while maintaining a difference in the mean duration between conditions.

For each experimental block, a sequence of 120 durations was created based on one of these three conditions/generative processes. Each sequence was generated ensuring that its statistical properties (i.e., mean, entropy rate, transition probabilities) were the closest to those parameterized in the respective generative experimental condition.

4.2.3 Apparatus

The same system was used to deliver the stimuli and collect participants' response: a custom-made set up (www.cynexo.com) including two microcontrollers, an Arduino Due (www.arduino.cc) connected to a response pad and a photodiode, and an Arduino Uno paired to a shield for sound delivery (Adafruit Wave Shield for Arduino, www.adafruit.com), into which noise-cancelling headphones were directly plugged.

Each trial event series was generated using Matlab and Psychtoolbox (Brainard, 1997; Kleiner et al., 2007), where each stimulus onset was coded as a visual trigger appearing on the screen (hidden from the participant) and read by the photodiode. The change of status of the photodiode triggered in turn the two microcontrollers serially to play the beep sound, marking the onset of each event. The response pad was equipped with a pressure sensor (SingleTact CS15-4.5N, www.singletact.com), attached underneath a resin-made button, so that the signal collected (sampled at 100 Hz) was analogic and reflected the force exerted on the button (range of sensitivity 0-4.5 Newton, represented in an arbitrary integer scale spanning from 0 to 1022). The resin-made button allowed us to get rid of the click or sensory feedback often associated with a key press on a plastic button. Preliminary technical tests showed a constant delay of 23 ms in sound delivery from the onset of the visual trigger to the headphones, with a negligible standard deviation (in the microsecond scale). This custom-made set up allowed us to have a single common recording and timescale for all events during the trial.

4.2.4 Stimuli and Task Procedure

Stimuli. Stimuli (the events of a generated sequence) were empty intervals marked by brief beeps (33 ms pure tones of 1000 Hz, 2-ms rise and fall time filter), created using a custom code in Matlab (MathWorks). The three durations used were 320 ms (D_1), 512 ms (D_2) and 820 ms (D_3). The length of each duration was a constant proportion (i.e., 60%) of the closest one. This strategy ensured a sufficient level of discriminability among the durations and a non-integer ratio relationship between them. The non-integer ratio choice was motivated to avoid potential rhythm-like confounds. Namely, the possibility of accidentally generating beat-inducing portions of the sequences and/or imposing overarching metrical structures (which would act as higher order

statistics). These confounds would have muddled or even shadowed the effects on tapping behavior of the simpler set of statistical rules defined in our manipulation.

Training. Participants sat comfortably in front of a computer screen and were given noise-cancelling headphones. After receiving general instructions on the experimental procedure, they performed a short training to familiarize with the task. To avoid any transfer of temporal learning to the main experiment, training blocks were generated from simpler Markov processes, i.e. two 2-by-2 Transition Probability Matrices (a Random TPM with p=0.5 for all transitions, and a Pattern TPM with p={0.1, 0.9, 0.1, 0.9}), in which different durations/states were alternating (350 ms and 770 ms). Each of the two training TPM generated a sequence of 45 events, split in 3 trials of 15 events each, for a total of 6 training trials. Training trials had the same structure of experimental trials (described later in the paragraph). At the end of each training block participants received feedback about the tapping pressure exerted on the button and the level of accuracy of the reproduction i.e. whether the tapping was too fast or too slow compared to the stimuli and/or whether there were long gaps within the reproduced sequence. After completing the training, participants started the main experiment.

Main experiment. Each of the three Markov processes defining the 3 experimental conditions (i.e., Random, Pattern 1 and 2) was used to generate 5 sequences of 120 events each (1 event = beep + empty interval), for a total of 15 blocks. Each block was presented to the participant divided in short trials, each comprising a subset of 15 events of the sequence, for a total of 8 trials per block (**Figure 4.1B**). The length of the subset was chosen to be long enough to prevent a "verbatim" reproduction of the subset, but at the same time short enough for its statistical properties to be learned. The presentation order of the blocks was randomized across subjects, with the only constraint of having a Random Condition as the first block. Within a block, trials were presented one after the other, only divided by a brief Inter Trial Interval, in order to have a continuous stream of perception-reproduction for the whole block sequence. Instead, the start of a new block was self-paced to allow short breaks if needed.

Trial structure. Each trial (**Figure 4.1A**) started with a "Listen" message on the screen, immediately followed by the presentation of the 15 events (*Perception phase*). Once the sequence ended, after a brief delay (jittered between 1.6-2.8 sec), a white fixation cross and a "Start" message on the screen cued the beginning of the *Reproduction phase*. In this phase the participant had to reproduce the sequence presented as accurately as possible, according to what was retained and learned, by tapping the finger on the response pad button. Reproduction time was randomly chosen to last within 1.3-1.8x the length of the sequence presented, after which a "Stop" message on the screen signaled to the participant the end of the trial. A brief Inter Trial Interval (jittered between 1.6-2.8 sec) separated the current trial from the next.

4.3 Data Analysis and Modelling

4.3.1 Pre-processing

Raw signals recorded by our stimulus-response apparatus were pre-processed to extract responses' onsets. All pre-processing steps were performed using custom codes written in Matlab (MathWorks). Each step described in this paragraph was applied to each trial of each participant separately.

Each stimulus' onset was identified as a change of status of the photodiode, whose signal was recorded as a time series of 0 and 1, with 1 corresponding to the appearance of the visual trigger on the screen. The actual timing of each beep's delivery was calculated by adding a constant delay of 23 ms to each detected stimulus onset. Concerning participants' response, the signal coming from each trial was a time series of force values recorded from the beginning to the end of it.

Presses' onset detection was performed through a series of hierarchical steps. First, in order to get rid of mechanical noise-related fluctuations, the raw signal from each trial was smoothed using a 3rd order polynomial Savitzky-Golay filter (Savitzky & Golay, 1964), with a window of 15 samples (i.e. corresponding to roughly 150 ms). The standard deviation of the first 50 samples of each trial was also calculated to obtain a baseline-noise standard deviation. As a second step, the mean force of the 20 highest values of the signal was calculated and, based on it, the trial was categorized either "Strong", "Medium" or "Weak" and assigned a trial-specific derivative-threshold (i.e. a positive difference between consecutive force values representing a "jump" in the signal, large enough to be caused by a press). Based on this categorization, also a category-specific standard deviation threshold (std-threshold) was defined as a multiple of the baseline standard deviation, representing the minimum amount of increase in the standard deviation of neighboring samples indicative of a "press" event. These tailored thresholds allowed us to take into account the great variability observed not only between participants, but also between trials performed by the same subject.

Based on these threshold values, the third step comprised a series of inclusion/exclusion criteria to define the presence or the absence of a press onset. Specifically, samples were first filtered based on the presence of either a suprathreshold jump in the signal (i.e. a derivative greater than the derivative-threshold), or a cumulative adding up of around-threshold derivatives in neighboring samples. For this second subgroup of potential onsets, the std-threshold was used as additional criterion to filter the candidate onsets. After this, all samples following a potential onset within a window of 180 ms were further excluded (i.e. they were considered either as signals' artifacts or "false" onsets, e.g. double taps). The final criterion was the presence of large enough negative derivatives between two candidate onsets, signaling a press offset (i.e. return to baseline of force values).

Stimuli onsets and presses' onsets were used to compute Inter Stimulus Intervals[2] (ISI) and Inter Tap Intervals (ITI), respectively. ISI and ITI are defined as the intervals between two consecutive stimuli onsets and two taps' onsets, respectively.

As a result of the pre-processing stage, two participants had to be excluded due to technical issues in cleaning their signals and detecting presses' onsets in more than 25% of their trials. Hence the resulting sample size used for the analysis is 21.

4.3.2 Model-based analysis

Our model assumes that participants are learning the Markovian generative process underlying the sequences of duration presented. It is an "ideal observer" model that, for each block and for each participant, estimates the parameters defining the statistical properties of the observed sequence (i.e. series of ITI or ISI). These parameters were the mean and standard deviation of the three performed durations' distributions, their proportions within the sequence and the transition probabilities relating them. Parameters estimation was done through a two step-procedure, that goes backward compared to the process generating the stimuli: first, the model identifies the three durations/states by estimating mean and standard deviation of the their distributions, and second, it estimates the transition probability matrix describing the generative process.

The first step of the model is formalized as a mixture of gaussians:

$$p(y) = \sum_{k=1}^{K} \theta_k \cdot \mathcal{N}(y \mid \mu_k, \sigma_k) \tag{4}$$

where y represents the series of observations whose likelihood is computed, K is the number of the sequence's possible states (i.e. durations), given as input along with y, and each θ_k represents the proportion in which each duration's distribution with mean μ_k and standard deviation σ_k is present in the sequence of observations. The model also receives a gaussian prior on the three durations' distributions, whose means are equal to the stimuli durations and whose standard deviation scalarly increases with duration's length.

Each parameter σ_k is estimated using the Generalized Weber's law formalization of σ^2 (Getty,1975; Ivry & Hazeltine, 1995):

$$\sigma_{tot}^2 = w \cdot T^2 + \sigma_m^2 \tag{5}$$

where the total variance of the reproduced duration is explained by a time-dependent and a motor-dependent component. The motor-dependent component σ_m^2 is assumed as constant variability in

[2] Throughout the chapter and in the supplementary figures in Appendix C the term Inter Stimulus Interval (ISI) is used as a synonym of the term Inter Onset Interval (IOI), more commonly used in finger tapping and rhythm perception literature.

the motor response that is duration-independent, whereas the time-dependent component *w* scales as the reproduced duration increases.

The estimated location parameters μ_k and scale parameters σ_k are used as priors in the second step of the model for estimating the parameter Φ, a k-by-k transition matrix describing the underlying generative process of the sequence observed. This estimation is performed by computing the likelihood of the data through the forward algorithm (Baum & Petrie, 1966), which calculates the most likely state at each point in time, given the history of the observations:

$$p(z_t, y_{1:t}) = \alpha_t(z_t) = p(y_t \mid z_t) \sum_{z_{t-1}} p(z_t \mid z_{t-1})\, \alpha_{t-1}(z_{t-1}) \quad (3)$$

The forward algorithm takes advantage of Markov chains' properties that the current observation y_t is only dependent on the hidden state z_t and z_t is only dependent on the previous state z_{t-1} . Based on these properties, it recursively marginalizes over all past observations and in turn estimates the joint probability of being in a specific hidden state z_t and the evidence so far $y_{1:t}$.

For each subject and each block we calculated the entropy rate and the conditional entropy of the estimated parameter Φ_{ITI} (i.e., estimated transition probabilities underlying the series of ITI reproduced), as well as its mean squared error (MSE) calculated with respect to different reference Φ. These references were the condition-specific transition matrix as it was parameterized in the design (Φ_{ITI} - Φ_{PRM}), the transition matrix estimated by the model for current block-ISI (Φ_{ITI} - Φ_{ISI}), for previous block-ISI (Φ_{ITI} - Φ_{ISI-1}) and for previous block-ITI (Φ_{ITI} - Φ_{ITI-1}). These measures allowed us to have group-level metrics to investigate the extent to which participants learned the statistical properties underlying the sequences presented, as well as assess learning differences across experimental conditions and exploring history effects. To this aim, we tested the effects of Experimental Condition (3 levels factor: Random, Pattern 1 and Pattern 2) and Condition History (5 levels factor, each corresponding to n^{th} presentation of each condition across blocks) on each of these measures running a series of Linear Mixed Effects (LME) models, with participants as random intercept.

The model was implemented in Stan using R as the interface (Stan Development Team, http://mc-stan.org; Carpenter et al., 2017). Stan uses a Markov Chain Monte Carlo sampling method to estimate posterior distributions over model parameters. For each participant and each block ITI and respective ISI, we ran the model with four independent chains and 5000 iterations (2500 warm-up). Convergence of sampling chains was estimated through the Gelman-Rubin $\hat{R}$ statistic (Gelman & Rubin, 1992), keeping as acceptable threshold $\hat{R}$ values smaller than 1.01. Statistical analysis and LME model fitting were performed in R (R Core Team, 2020), using the packages and methods described in detail in the next paragraph.

4.3.3 Data-driven analysis

To assess the effect of our experimental manipulations directly on participants' tapping behavior and identify in a model-agnostic way which types of low- and high-level statistics (i.e. which mean duration and transition probabilities) participants performance reflected, we adopted also a data-driven approach.

ITI mean and ITI error analysis. ITI mean and ITI error were calculated at the block-level for each participant. Different ITI errors were computed using as reference either a global or local mean. These error indexes were: *glob-error* was computed as the difference between block ITI mean and all blocks' ISI mean (irrespective of the experimental condition); *cond-error* was the difference between block ITI mean and the mean of each condition's ISI; ISI_b*-error* was the difference between block ITI mean and current block ISI mean; ISI_{b-1}*-error* was the difference between block ITI mean and previous block ISI mean; ITI_{b-1}*-error* was the difference between block ITI mean and previous block ITI mean. For each of these measures, the effects of Experimental Condition (3 levels factor: Random, Pattern 1 or Pattern 2) and Condition History (5 levels factor, each corresponding to n^{th} presentation of each condition across blocks) were assessed using LME models, with participants as random intercept.

K-means cluster analysis. To obtain a model-agnostic estimation of how many durations and which were reproduced by our participants, for each individual we performed k-means clustering analysis on the raw ITI data of each block. Specifically, for each series of ITI we ran k-means algorithm looping from K=1 to K=10, with n=25 starting random centroids per K. A silhouette score, or coefficient, was obtained for each K, representing the mean degree of similarity of data-points assigned to the same cluster compared to other clusters (Rousseeuw, 1987). Silhouette scores range from -1 to 1, with the latter representing the best possible match between cohesion within cluster and separation among clusters. Based on silhouette analysis results, we reduced the K-range between 2-to-3 clusters and carried out two parallel analyses aimed at comparing model-based and k-means results across experimental conditions.

For the first analysis, we selected for each block the K showing the highest silhouette score among K=2 or K=3 (i.e. the preferred K) and, according to it, categorized each raw ITI with its cluster's label (e.g. for K=2 as preferred clustering, each data point was categorized either as 1 or 2). As a result, we obtained a series of discrete values for each block, equivalent to a series of "known" states. Each transformed series was used to feed a naïve model to estimate the parameter Φ, a k-by-k matrix describing the transition probabilities relating the states, where K corresponds to the block-specific preferred K. Parameter fitting was performed by looping over the likelihood of each of the series' state z_t , which follows a categorical distribution parameterized by $\Phi_{z_{t-1}}$:

$$z_t \sim categorical\ (\Phi_{z_{t-1}})$$

(4)

The second analysis directly assessed the relationship between the K preference and the degree of overlap between model-estimated posterior distributions over μ. Specifically, for each participant

and each block we calculated the overlap between μ_1 and μ_2 distributions and the one between μ_2 and μ_3 distributions. This measure was then used as dependent variable in a LME model, with participants as random intercept and a series of potential modulatory factors as fixed effects: specifically, experimental condition (3 levels factor: Random, Pattern 1 or Pattern 2), identity of the overlapping μ (i.e. either the overlap between μ_1-μ_2 or the one between μ_2-μ_3), the block-specific k-means preference (2 levels factor: K=2 or K=3).

Data-driven analyses were all performed in R (R Core Team, 2020). LME models were fitted using the packages *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017), and *MuMIn* package (Barton, 2020) was used for calculating R^2 statistics. Overall effects were tested using Type III Analysis of Variance with Satterthwaite's method for degrees of freedom. Post-hoc comparisons were conducted using *emmeans* package (Lenth, 2020), using Kenward-Roger's method for degrees of freedom and Tukey's method for p-value adjustment. Alpha level was always set at 0.05. The R package *cluster* (Maechler et al., 2019) was used for K-means clustering analysis. The naïve model was implemented in Stan using R as the interface (Stan Development Team, http://mc-stan.org; Carpenter et al., 2017), using four independent chains and 5000 iterations (2500 warm-up).

4.4 RESULTS

4.4.1 Model-based analysis

The normative model proposed in this study represents an ideal observer model, which tracks and learns through Bayesian inference the statistical properties of the Markovian generative processes underlying the sequences of durations presented. These statistical properties are expressed by a set of parameters estimated by the model for each block. The parameters are: mean and standard deviation of each of the three durations alternating in the sequences, the proportions in which they are present within each sequence, and the transition probabilities ruling the alternations between durations.

Parameters estimation was done block-by-block for each participant, for both the specific sequences of stimuli perceived (i.e. ISI, Inter Stimulus Intervals) and for the series of durations reproduced (i.e. ITI, Inter Tap Intervals). The double estimation had a twofold goal. First, to have a posteriori sanity check that the model's estimation would mimic the statistical properties parameterized in the experimental design for each condition. Second, to obtain a block-specific estimation "reference" and make direct comparisons between model's results for ISI and ITI.

As a representative example of the results obtained from each step of model estimation, **Figure 4.2A** shows estimated posterior probabilities on location parameters μ and scale parameters σ for each block from one participant, and **Figure 4.2B** shows for the same subject the results of the θ parameter estimation, both for ISI sequences (i.e. θ_{ISI}) and ITI sequences (i.e. θ_{ITI}). The combination of parameters μ_{ITI}, σ_{ITI} and θ_{ITI} defines, for each block, the mixture of gaussians estimated by the model, reflecting the durations' means, standard deviations and frequency

distributions learned by this specific participant. This information is used by the model to estimate the transition probabilities underlying this subject's performance (i.e. parameter Φ). **Figure 4.3** shows the results of this estimation, both for the sequences of ITI reproduced (i.e. $\mathbf{\Phi}_{ITI}$ in **4.3A**) and the sequences of ISI presented (i.e. $\mathbf{\Phi}_{ISI}$ in **4.3B**).

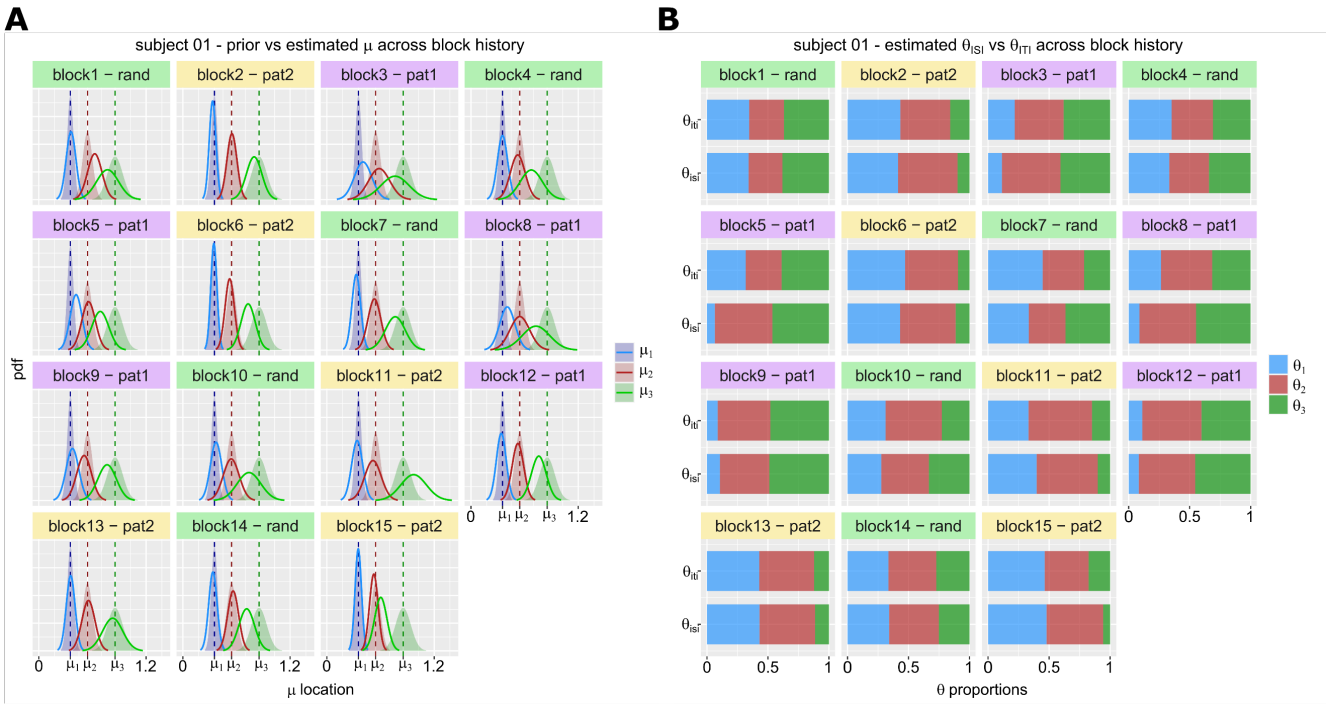


Figure 4.2. **Estimated parameters μ and θ**. Results of μ and θ estimations are shown for subject 01. These parameters define, block-by-block, the mixture of gaussians of the three durations performed by subject 01. Each facet's title indicates block number and block-specific experimental condition (color coded as green for Random, purple for Pattern 1, yellow for Pattern 2 condition). (A) Prior and posterior durations' distributions: the shaded areas on each facet represent the prior distributions over the three μ (blue for μ_1=320 ms, red for μ_2=512 ms, green μ_3=820 ms), whereas the thick lines are the subject-specific durations' posterior distributions estimated by the model. (B) Estimated θ_{ISI} and θ_{ITI}: Each facet's lower barplot depicts the block-specific estimation of the three μ_{ISI} proportions within the sequence of ISI (i.e. θ_{ISI}) that subject 01 perceived. The upper barplot in each facet represents block-specific estimation of the three μ_{ITI} proportions within the sequence of ITI (i.e. θ_{ITI}) that subject 01 reproduced. Each duration's identity is color-coded as in (A). ISI = Inter Stimulus Interval; ITI = Inter Trial Interval.

As **Figure 4.3B** shows, the model was able to accurately retrieve the transition matrices specific of each experimental condition, confirming that it was acting as an optimal Bayesian observer and it was able to track the ongoing statistical properties of the sequences of stimuli. Results from parameter Φ_{ITI} estimation, instead, inform on the extent to which participants were learning condition-specific higher order statistics (i.e., the set of transition probabilities) and how they differed from those used to design the stimuli series.

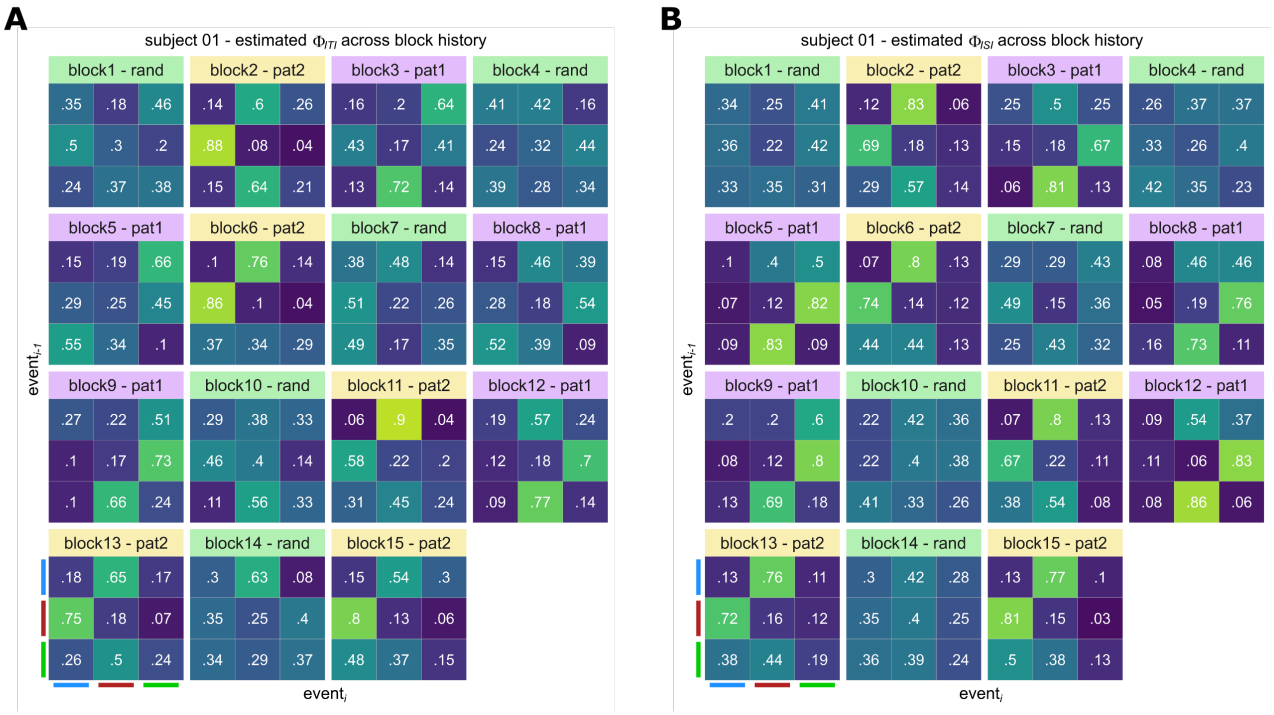


Figure 4.3. **Estimated parameters Φ_{ITI} and Φ_{ISI}.** The figure shows the results of parameter Φ estimation for subject 01. For each block the model estimates Φ, representing the transition matrix describing the generative process of ISI sequences (Φ_{ISI}) and ITI sequences (Φ_{ITI}). Each facet's title indicates block number and block-specific experimental condition (color coded as green for Random, purple for Pattern 1, yellow for Pattern 2 condition). The color hue and the number inside each tile indicate the block-specific conditional probability of the current $event_i$ being one of the three durations, given the identity of the previous duration (i.e. $event_{i-1}$). The three durations' identities are depicted on the bottom left corner matrix of each plot (blue for μ_1=320 ms, red for μ_2=512 ms, green μ_3=820 ms). (A) Estimation of Φ_{ITI} underlying the series of durations performed by subject 01. (B) Estimation of Φ_{ISI} underlying the series of durations perceived by subject 01. ISI = Inter Stimulus Interval; ITI = Inter Trial Interval.

Entropy measures. To evaluate at the group level learning effects across conditions and their history, we calculated entropy rate and conditional entropy of estimated Φ_{ITI} for each block and each participant. For each of the two measures we fit a LME model with entropy as the dependent variable, experimental condition and condition history as independent variables in interaction, and participants as random intercept (model formula: entropy ~ experimental condition * condition history + (1 | subject)). Goodness of fit of both models was evaluated using R^2 statistic ($R^2_{marginal}$ = 0.148, $R^2_{conditional}$ = 0.186 for entropy rate fit; $R^2_{marginal}$ = 0.238, $R^2_{conditional}$ = 0.36 for conditional entropy fit). Type III ANOVA on the model estimates showed in both cases a significant main effect of experimental condition ($F(2, 280) = 22.091$, $p<0.0001$ for entropy rate; $F(2, 280) = 56.5049$, $p<0.0001$ for conditional entropy) (**Figure 4.4A and 4.4B**). Post-hoc comparisons for LME model on entropy rate revealed that Pattern 2 condition had significantly lower entropy values compared to the other two conditions (pat1–pat2 = 0.042, t.ratio(280) = 4.425, $p<0.0001$; rand–pat2 = 0.0617, t.ratio(280) = 6.508, $p<0.0001$), whereas Pattern 1 condition's entropy rate,

although still lower, did not differ significantly from that of Random condition (rand–pat1 = 0.0198, t.ratio(280) = 2.083, p=0.0952). Concerning post-hoc comparisons for LME model on conditional entropy instead, both Pattern conditions significantly showed lower entropy from Random condition (rand-pat1 = 0.1043, t.ratio(280) = 6.443, p<0.0001; rand-pat2 = 0.1707, t.ratio(280) = 10.544, p<0.0001), with Pattern 2 conditional entropy values being significantly lower also with respect to Pattern 1 (pat1–pat2 = 0.0664, t.ratio(280) = 4.101, p<0.001). For both entropy measures, there were neither main effects of condition history ($F(4, 280) = 0.9646$, $p=0.427$ for entropy rate; $F(4, 280) = 0.1686$, $p = 0.954$ for conditional entropy), nor interaction effects of condition type and history ($F(8, 280) = 1.1692$, $p = 0.3177$ for entropy rate; $F(8, 280) = 0.3785$, $p = 0.9316$ for conditional entropy), indicating that the learned statistical properties were stable across blocks and didn't show any form of improvement or deterioration (see **Appendix C**, **Figure C1**). Entropy measures results suggest that participants were learning the underlying statistical properties of each experimental condition, to a sufficient extent to be able to discriminate and learn the condition-specific set of conditional probabilities relating the durations. Nevertheless, this learning effects appear stronger for Pattern 2 compared to Pattern 1 condition. The absence of condition history effects indicates that the learning remained stable across blocks.

Mean Squared Error measures. A similar set of LME models was used to evaluate condition- and history-dependent learning effects expressed as differences in MSE values, calculated with respect to the set of reference Φ as described in the Method section (i.e., $\Phi_{ITI} - \Phi_{PRM}$, $\Phi_{ITI} - \Phi_{ISI}$, $\Phi_{ITI} - \Phi_{ISI\text{-}1}$, $\Phi_{ITI} - \Phi_{ITI\text{-}1}$, all represented in **Figure 4.4C**). Each log-transformed MSE (i.e. since MSE values are positively-skewed distributed, their log-transformation allows to approximate normality and homogeneity of residuals) was used as dependent variable, whereas experimental condition and condition history were set as independent variables in interaction, along with participants as random intercept (model formula: MSE ~ experimental condition * condition history + (1 | subject)). Goodness of fit of each model was evaluated using R^2 statistic ($R^2_{marginal} = 0.173$, $R^2_{conditional} = 0.326$ for $\Phi_{ITI} - \Phi_{PRM}$ MSE fit; $R^2_{marginal} = 0.156$, $R^2_{conditional} = 0.284$ for $\Phi_{ITI} - \Phi_{ISI}$ MSE fit; $R^2_{marginal} = 0.026$, $R^2_{conditional} = 0.046$ for $\Phi_{ITI} - \Phi_{ISI\text{-}1}$ MSE fit; $R^2_{marginal} = 0.038$, $R^2_{conditional} = 0.068$ for $\Phi_{ITI} - \Phi_{ITI\text{-}1}$ MSE fit), showing overall poorer fit for the two models on MSE values calculated in respect to $\Phi_{ISI\text{-}1}$ and $\Phi_{ITI\text{-}1}$.

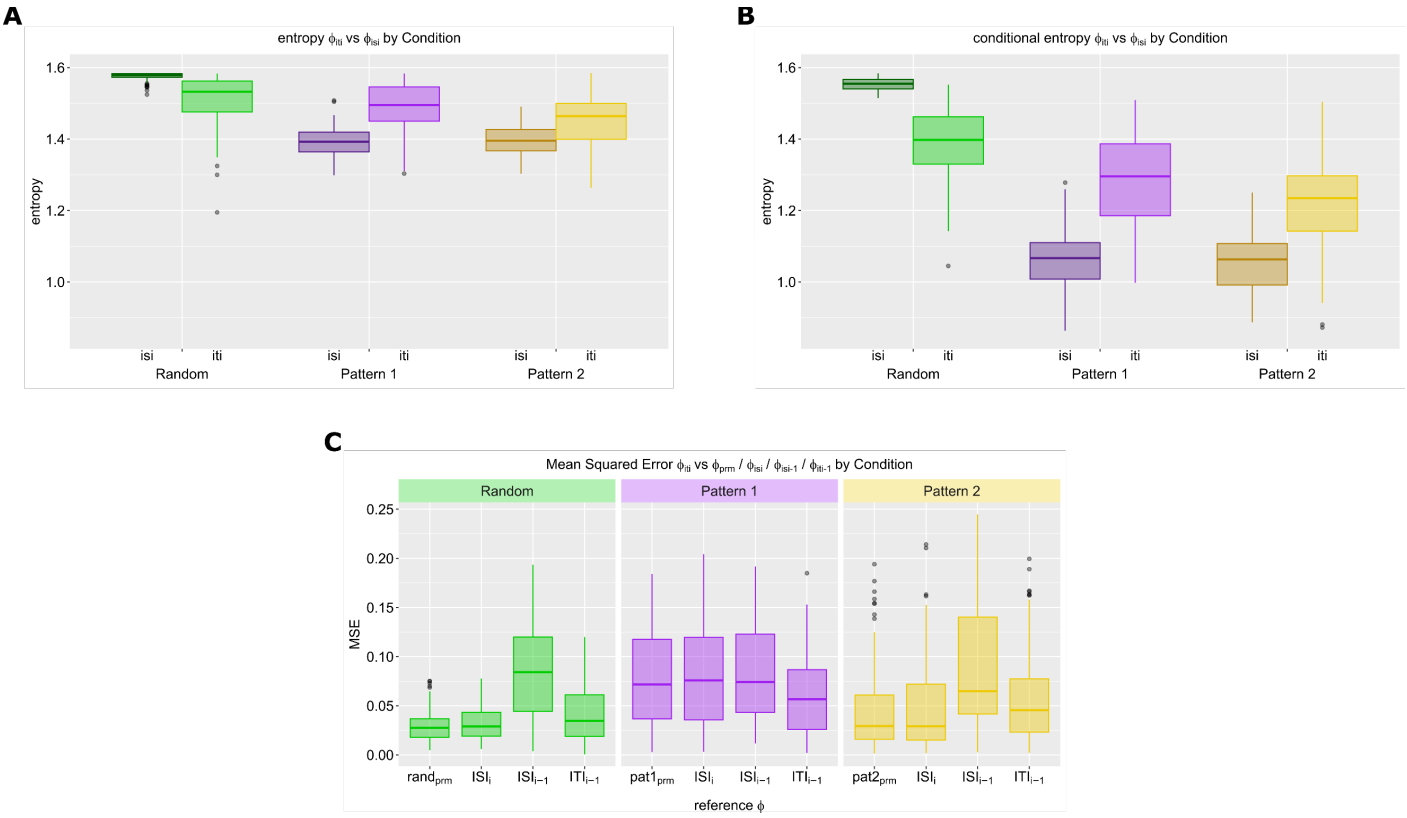


Figure 4.4 **Entropy rate, conditional entropy and Mean Squared Error for estimated parameter Φ_{ITI}.** (A) Entropy rate values calculated from estimated parameters Φ_{ISI} and Φ_{ITI} grouped by experimental condition (color coded as green for Random, purple for Pattern 1, yellow for Pattern 2 condition). Darker boxplot represents entropy values for Φ_{ISI}. (B) Conditional entropy values calculated from estimated parameters Φ_{ISI} and Φ_{ITI} grouped by experimental condition (color coded as in A). Darker boxplot represents entropy values for Φ_{ISI}. (C) Mean Squared Error between Φ_{ITI} and four different reference Φ: Φ_{prm} (defined in the respective experimental condition), Φ_{ISI} (estimated from current block ISI series), $\Phi_{ISI\text{-}1}$ (estimated from previous block ISI series), $\Phi_{ITI\text{-}1}$ (estimated from previous block ITI series), grouped by experimental condition (color coded as in A). MSE = Mean Squared Error; ISI = Inter Stimulus Interval; ITI = Inter Tap Interval.

Since the reference Φ is almost identical, type III ANOVA on Φ_{ITI} - Φ_{PRM} and Φ_{ITI} - Φ_{ISI} MSE models estimates showed overlapping results as expected: only a significant main effect of experimental condition emerged ($F(2, 280) = 34.5201$, $p<0.0001$ for Φ_{ITI} - Φ_{PRM}; $F(2, 280) = 29.7713$, $p<0.0001$ for Φ_{ITI} - Φ_{ISI}), which in both cases was due to overall significantly higher MSE values for Pattern 1 condition compared to Random and Pattern 2 conditions, as revealed by post-hoc comparisons (for Φ_{ITI} - Φ_{PRM}: rand–pat1 = -0.845, t.ratio(280) = -7.771, $p<0.0001$; pat1–pat2 = 0.699, t.ratio(280) = 6.434, $p<0.0001$; for Φ_{ITI} – Φ_{ISI}: rand–pat1 = -0.778, t.ratio(280) = -7.220, $p<0.0001$; pat1–pat2 = 0.643, t.ratio(280) = 5.969, $p<0.0001$), and no MSE differences between Random and Pattern 2 conditions (for Φ_{ITI} - Φ_{PRM}: rand–pat2 = -0.145, t.ratio(280) = -1.337, $p = 0.376$; for Φ_{ITI} – Φ_{ISI}: rand–pat2 = -0.135, t.ratio(280) = -1.251, $p = 0.4242$). In both models no effect of condition history on MSE values was observed ($F(4, 280) = 0.8086$, $p = 0.5205$ for Φ_{ITI} - Φ_{PRM}; $F(4, 280) = 0.5974$, $p = 0.6648$ for Φ_{ITI} - Φ_{ISI}), nor any interaction with condition type (F(8,

280) = 1.0284, p = 0.4147 for Φ_{ITI} - Φ_{PRM}; F(8, 280) = 0.8408, p = 0.5674 for Φ_{ITI} - Φ_{ISI}) (see **Appendix C**, **Figure C1**). The two models' overlapping results, in summary, indicate that Pattern 1 condition shows the greatest dissimilarity between the transition probabilities underlying the ISI series and those estimated by the model for ITI series, and this distance is constant across condition history. On the contrary, Pattern 2 and Random conditions show in general low dissimilarity between stimuli's and performances' underlying transition probabilities.

Concerning Φ_{ITI} - Φ_{ISI-1} and Φ_{ITI} - Φ_{ITI-1} MSE fit, type III ANOVA on both models estimates revealed no significant effects of experimental condition (F(2, 220) = 1.015, p = 0.3641 for Φ_{ITI} - Φ_{ISI-1}; F(2, 220) = 1.9853, p = 0.1398 for Φ_{ITI} - Φ_{ITI-1}), or condition history (F(3, 220) = 1.0468, p = 0.3727 for Φ_{ITI} – Φ_{ISI-1}; F(3, 220) = 0.6878, p = 0.5603 for Φ_{ITI} - Φ_{ITI-1}), nor interaction effects (F(6, 220) = 0.3152, p = 0.9287 for Φ_{ITI} – Φ_{ISI-1}; F(2, 280) = 0.7251, p = 0.6298 for Φ_{ITI} - Φ_{ITI-1}) (**Figure 4.4C**). Overall these two models' results highlight the absence of any effect of recent ISI and ITI history in the ongoing learning process.

Summary of model results and possible scenarios. Statistical comparisons on the measures derived from estimated parameter Φ showed that participants were able to discriminate and, to a significant extent, learn high-level statistical properties defining each experimental condition, although with differences between Pattern 1 and 2, despite their mirroring properties. Nevertheless, as can be noticed in **Figure 4.2A**, in some case the estimated posterior distributions of the performed durations diverged from the distribution of the stimuli, and in some other there was a certain amount of overlap between them. Given the informative priors the model has (i.e. the fixed number of durations within the series, as well as location and scale of their distributions within the series of stimuli) and the 2-step procedure it follows for parameters estimation, it is clear that low-level statistics parameters can influence those defined as high-level statistics. In other words, the level of uncertainty in estimating the location and shape of the three durations performed by the subject (i.e. the parameters μ_{ITI} and σ_{ITI}) can bias and distort to a great extent the estimation of the transition probabilities describing how the three durations alternate. This, in turn, can conceal the possibility in which the reproduced series are quite accurate from the high-level-statistics-point of view (i.e. the transition probabilities underlying performed sequences reflect those designed in each experimental condition), but are imprecise with respect to low-level statistics. The uncertainty in the estimation of the three durations performed can indeed conceal different possible scenarios of participants' performance (**Figure 4.5**). One of this (**Figure 4.5B**) may be characterized by an "overall shift" of the durations performed. The participant is reproducing the correct number of durations (i.e. 3), but the durations are under-reproduced to the point that the longest of the three is confused with the medium duration by the model, due to the priors on their location. Another possibility is the "shrinkage" scenario (**Figure 4.5C**): the participant is somehow merging the three durations into two, but since the model has a prior also on the number of durations, it will still estimate three distributions characterized though by a great degree of overlap. And this in turn will confuse the estimation of the transition probabilities underlying the series of ITI. To able to disclose those possible scenarios and understand in a more agnostic, uninformed way how many

and which durations participants were learning and performing, we carried out a series of data-driven analysis on raw ITI data.

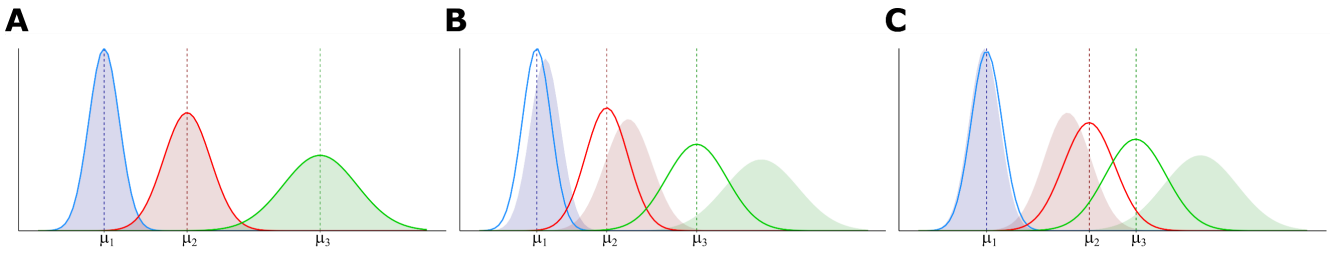


Figure 4.5. **Hypothesized μ distortions in participants' reproduction**. Shaded curves represent prior distributions over the three μ that the model receives, whereas the thick lines mimic possible distortions in the reproduced durations' distributions (color code: blue for μ_1=320 ms, red for μ_2=512 ms, green μ_3=820 ms). (A) "No-distortions" scenario: reproduced durations are consistent in number and locations with those of the sequence of stimuli. (B) "Overall shift" scenario: the durations reproduced by the participant are actually three as in the stimuli series, but there is a general shift in their means, i.e. a general underreproduction of all three durations. (C) "Shrinkage" scenario: the number of reproduced durations is two instead of three, hence two of the three means are located very close to each other because they reflect essentially only one distribution.

4.4.2 Data-driven analysis: ITI mean and ITI error

We analyzed block-ITI mean and error addressing a series of related questions: a) quantify, independently of model assumptions, which type of low-level statistics (i.e. mean durations) participants were extracting and reproducing; b) defining which type of perceptual distortions (i.e., over-reproduction or under-reproduction) corresponded to the shifts and overlaps observed in the three durations' posterior distributions estimated by the model; c) assessing the changes of these distortions across experimental conditions and condition history, and d) understanding whether these distortions were influenced by recent block/trial history.

ITI mean. We calculated the mean of raw ITI series at the block-level (**Figure 4.6A**) and fit a LME model using each of the two measures as dependent variable, with experimental condition and condition history as independent variables in interaction, and participants as random intercept (model formula: mean ITI ~ experimental condition * condition history + (1 | subject)). Goodness of fit of the model was evaluated using R^2 statistic ($R^2_{marginal}$ = 0.499, $R^2_{conditional}$ = 0.775). Type III ANOVA on the model estimates showed a significant main effect of experimental condition ($F(2, 280) = 313.0976$, $p<0.0001$), indicating that participants were learning condition-specific mean durations, whether these means were close to the real condition means or not. Post-hoc comparisons further clarified these differences as a higher block-ITI mean for Pattern 1 condition compared to Random (rand–pat1 = -0.0510, t.ratio(280) = -12.131, $p<0.0001$) and Pattern 2 condition (pat1–pat2 = 0.1053, t.ratio(280) = 25.020, $p<0.0001$). We also observed a higher block-ITI mean for Random compared to Pattern 2 condition (rand–pat2 = 0.0542, t.ratio(280) = 12.889, $p<0.0001$). A main effect of condition history emerged as well ($F(4, 280) = 14.2717$, $p<0.0001$),

with no interaction with condition type ($F(8, 280) = 1.5799$, $p=0.1305$). The main effect of condition history was due to a general and constant decreasing trend in block-ITI mean across condition repetitions (rep1-rep2 = 0.01755, t.ratio(280) = 3.231, p=0.012; rep1-rep3 = 0.02604, t.ratio(280) = 4.795, p<0.0001; rep1-rep4 = 0.02937, t.ratio(280) = 5.409, p<0.0001; rep1-rep5 = 0.03836, t.ratio(280) = 7.064, p<0.0001; rep2-rep3 = 0.00849, t.ratio(280) = 1.564, p=0.5219; rep2-rep4 = 0.01183, t.ratio(280) = 2.178, p=0.1912; rep2-rep5 = 0.02081, t.ratio(280) = 3.832, p=0.0015; rep3-rep4 = 0.00333, t.ratio(280) = 0.614, p=0.9728; rep3-rep5 = 0.01232, t.ratio(280) = 2.269, p=0.1582; rep4-rep5 = 0.00899, t.ratio(280) = 1.655, p=0.4637).

ITI error. A second, related set of LME models was used to assess the effects of condition type and history on block-ITI error values, calculated with respect to different global and local reference means as described in the Method section (i.e., *glob-error*, *cond-err*, *ISI_b-error*, *ISI_{b-1}-error* and *ITI_{b-1}-error*, all represented in **Figure 4.6B**). Each ITI error was entered in the model as dependent variable, whereas experimental condition and condition history (according to the type of error, either 4 or 5 levels factor, each corresponding to n^{th} presentation of each condition across blocks) were set as independent variables in interaction, along with participants as random intercept (model formula: ITI error ~ experimental condition * condition history + (1 | subject)). Goodness of fit of each model was evaluated using R^2 statistic ($R^2_{marginal} = 0.502$, $R^2_{conditional} = 0.773$ for *glob-error* fit; $R^2_{marginal} = 0.36$, $R^2_{conditional} = 0.708$ for *cond-error* fit; $R^2_{marginal} = 0.353$, $R^2_{conditional} = 0.70$ for *ISI_b-error* fit; $R^2_{marginal} = 0.196$, $R^2_{conditional} = 0.254$ for *ISI_{b-1}-error* fit; $R^2_{marginal} = 0.418$, $R^2_{conditional} = 0.418$ for *ITI_{b-1}-error* fit).

Results from LME models on *glob-error*, *cond-error* and *ISI_b-error* shed light on potential central tendency effects combined with condition-specific distortions in participants' performance. If some kind of regression toward a general mean duration is influencing the performance, condition-specific effects would be present in the form of either positive-signed ITI errors (i.e. overreproduction) or negative-signed ITI errors (i.e. underreproduction) with respect to each condition's mean duration. If a central tendency effect is absent, *cond-error* and *ISI_b-error* would be both close to zero and without any trend, whereas *glob-error* would be highly condition-specific. The three models' results overlap for all effects, hence are presented together. Type III ANOVA on the three models estimates showed a significant main effect of experimental condition ($F(2, 280) = 313.0976$, $p<0.0001$ for *glob-error*; $F(2, 280) = 159.2176$, $p<0.0001$ for *cond-error*; $F(2, 280) = 154.6293$, $p<0.0001$ for *ISI_b-error*), indicating that participants' errors were both influenced by the global mean and by each condition's mean duration, hence confirming a condition-specific regression toward the mean. Post-hoc comparisons specified further these effects, revealing that Pattern 1 and Random *cond-errors* and *ISI_b-errors* resulted in a general under-reproduction of the durations presented, and this effect was significantly more pronounced for Pattern 1 compared to Random condition (rand–pat1 = 0.0321, t.ratio(280) = 7.635, p<0.0001 for *cond-error*; rand–pat1 = 0.0321, t.ratio(280) = 7.526, p<0.0001 for *ISI_b-error*), whereas Pattern 2 errors were the result of an over-reproduction of the durations presented, leading to significant differences between this and the other two conditions (for *cond-error*: pat1–pat2 = -0.0747, t.ratio(280) = -17.786, p<0.0001; rand–pat2 = -0.0427, t.ratio(280) = -10.151, p<0.0001; for *ISI_b-*

error: pat1–pat2 = -0.0747, t.ratio(280) = -17.528, p<0.0001; rand–pat2 = -0.0426, t.ratio(280) = -10.002, p<0.0001).

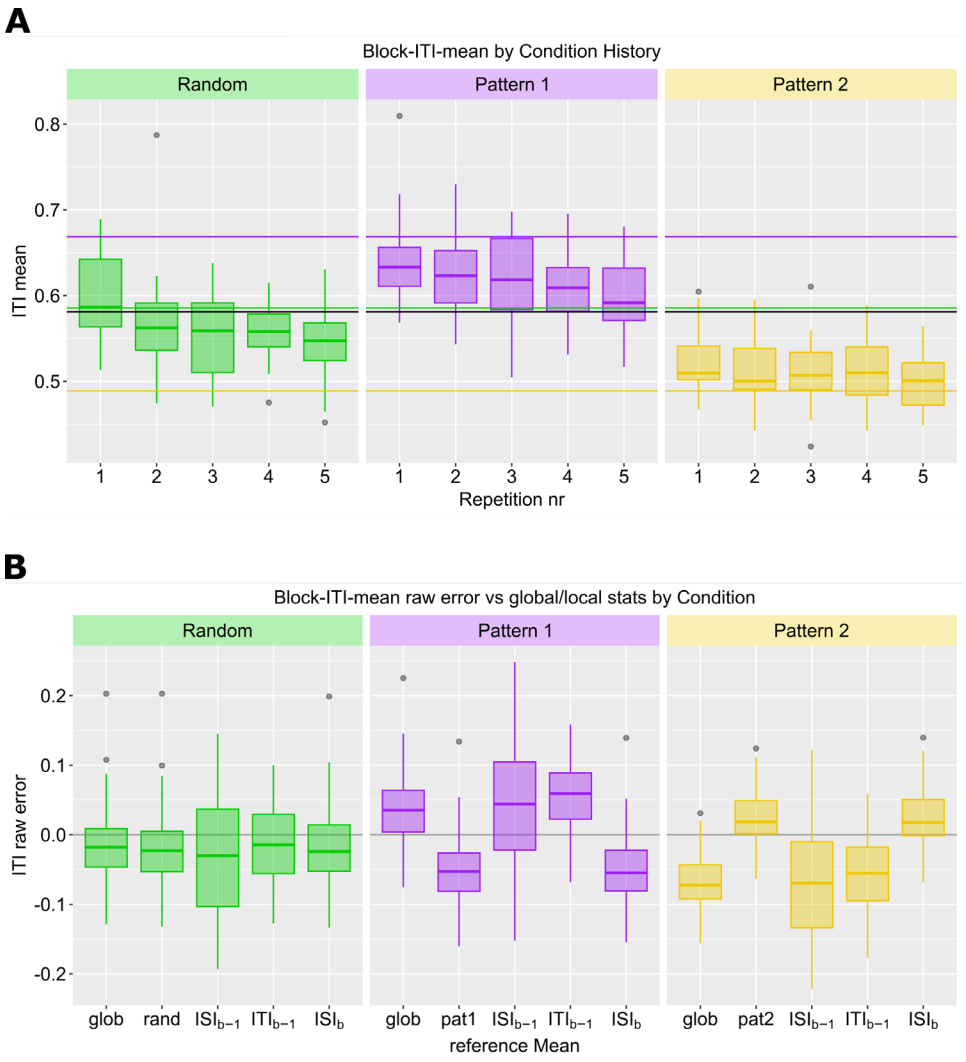


Figure 4.6. **Block-mean ITI and block-ITI error.** (A) ITI mean calculated at the block level grouped by experimental condition (color coded as green for Random, purple for Pattern 1, yellow for Pattern 2 condition) and plotted by condition history (i.e. number of repetition of each condition across blocks). Horizontal colored lines indicate each experimental condition's mean duration, as calculated from ISI (color coded by condition), whereas the black line represents all stimuli global mean duration. A main effect of condition history is visible as a constant decreasing ITI mean across all conditions' repetitions. (B) Block-ITI raw error calculated in respect of different global- or local-level reference means (see Method section for details), grouped by experimental condition (color coded as in A). Negative error values indicate underreproduction in respect to the reference mean duration, whereas positive error values indicate overreproduction compared to the reference mean duration. Inter Stimulus Interval; ITI = Inter Tap Interval.

Post-hoc comparisons on *glob-error*, as a consequence, mirrored the two other errors' trends: ITI errors were significantly more positive for Pattern 1 compared to Random condition (rand–pat1 =

-0.0510, t.ratio(280) = -12.131, p<0.0001) and Pattern 2 condition (pat1–pat2 = 0.1053, t.ratio(280) = 25.020, p<0.0001), and they were more negative for Pattern 2 compared to Random condition (rand–pat2 = 0.0542, t.ratio(280) = 12.889, p<0.0001). For all three models also a main effect of condition history emerged (F(4, 280) = 14.2717, p<0.0001 for *glob-error*; F(4, 280) = 14.3029, p<0.0001 for *cond-error*; F(4, 280) = 12.5106, p<0.0001 for ISI_b-*error*), with no interaction with condition type (F(8, 280) = 1.5799, p=0.1305 for *glob-error*; F(8, 280) = 1.5833, p=0.1295 for *cond-error*; F(8, 280) = 1.4695, p=0.1681 for ISI_b-*error*), due to a general drift towards negative values for ITI errors across condition repetitions (see **Appendix C**, **Figure C2** and its caption for post-hoc comparisons details).

Results from LME models on ISI_{b-1}-*error and* ITI_{b-1}-*error* inform on potential effects of recent block-history and how these effects may evolve across block repetitions. If there is any influence of previously perceived or performed durations, this would result in a substantial decrease in ITI error values for these two types of indexes. Results reveled no clear recent history effects. Type III ANOVA on the two models estimates showed again, a significant main effect of experimental condition (F(2, 220) = 29.7026, p<0.0001 for ISI_{b-1}-*error*; F(2, 240) = 86.1733, p<0.0001 for ITI_{b-1}-*error*), due to the fact that ISI_{b-1}-*error* and ITI_{b-1}-*error* values were influenced only by the condition-specific mean and by the global mean, thus excluding recent history effects. Post-hoc comparisons showed that, for both error types, differences among experimental conditions were replicating those found for *glob-error* results. Specifically we found significantly more positive ITI errors for Pattern 1 compared to Random condition (rand–pat1 = -0.0672, t.ratio(220) = -5.266, p<0.0001 for ISI_{b-1}-*error*; rand–pat1 = -0.0682, t.ratio(220) = -8.479, p<0.0001 for ITI_{b-1}-*error*) and Pattern 2 condition (pat1–pat2 = 0.0957, t.ratio(220) = 7.507, p<0.0001 for ISI_{b-1}-*error*; pat1–pat2 = 0.1039, t.ratio(220) = 12.919, p<0.0001 for ITI_{b-1}-*error*;), And more negative ITI errors for Pattern 2 compared to Random condition (rand–pat2 = 0.0286, t.ratio(220) = 2.241, p=0.066 for ISI_{b-1}-*error*; rand–pat2 = 0.0357, t.ratio(220) = 4.440, p<0.0001 for ITI_{b-1}-*error*). For both models no main effect of condition history was found (F(3, 220) = 1.397, p=0.2441 for ISI_{b-1}-*error*; F(3, 240) = 0.5960, p=0.6182 for ITI_{b-1}-*error*), and no interaction with condition type (F(6, 220) = 0.425, p=0.8619 for ISI_{b-1}-*error*; F(6, 240) = 1.0919, p=0.3678 for ITI_{b-1}-*error*), thus excluding also potential build-up of recent history effects across block repetitions (see **Appendix C**, **Figure C2**).

Overall results from the analysis on ITI mean and error revealed that participants' performance reflected an underlying learning of low-level statistical properties of the series of durations presented, i.e. the mean duration of each experimental condition, as was evidenced by the ubiquitous main effect of experimental condition in all LME models results. Nevertheless, ITI error analysis allowed to spot a central tendency effect, modulated by each condition's mean duration. Specifically, there is a general under-reproduction of the durations experienced in Pattern 1 condition, and an over-reproduction of those experienced in Pattern 2. In Random condition a mild under-reproduction is still present, but, due to the overlap between condition mean and global mean, it is not as pronounced as in Pattern 1 condition. As reflected by the condition history effect found in almost all LME models, there is a generic trend toward a reduction in the mean of the

reproduced durations across block history. This effect appears to be condition-independent and may be due to the aforementioned central tendency effect in a sort of "moving average-central tendency effect".

4.4.3 Data-driven analysis: k-means clustering

To see whether the shifts and overlaps observed in the three durations' posterior distributions estimated by the model reflected either an "overall shift" or a "shrinkage" of the durations reproduced (as hypothesized in paragraph 3.4.2), we performed a second data-driven analysis. This analysis aimed at obtaining an agnostic estimate of the number and of the identity of the durations performed by our participants. To this aim, we performed k-means clustering analysis on raw ITI data of each block for each subject, looping from K=1 to K=10 clusters, and comparing the resulting silhouette scores of each K. This silhouette analysis revealed consistently highest scores for K=2 clusters across experimental conditions (**Figure 4.7**) and a slow decrease in the score as the number of K increases. This pattern of results was stable across conditions and block history (**Appendix C**, **Figure C3**). An ANOVA on silhouette scores values obtained for K=2 and K=3 clusters (formula: silhouette values ~ experimental condition * block number * number of K), confirmed the consistency of silhouette analysis results, showing only a main effect of experimental condition ($F(2) = 47.9181$, $p<0.0001$) and of preference for K=2 clusters ($F(1) = 43.7264$, $p<0.0001$), and no effect of block number ($F(14) = 1.2763$, $p=0.217$) or any type of interaction.

Based on silhouette scores results, two parallel analyses were carried out after reducing the number of possible clusters to K=2 and K=3, with the general goal of comparing directly model-based and k-means results across experimental conditions.

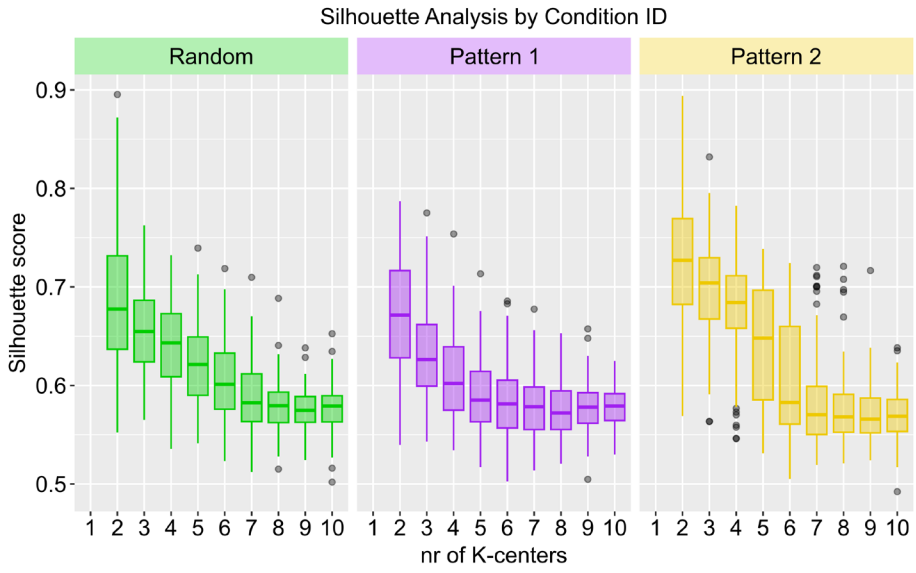


Figure 4.7. **K-means silhouette analysis on raw ITI**. Silhouette scores are plotted for each K iteration, grouped by experimental condition (color coded as green for Random, purple for Pattern 1, yellow for Pattern 2 condition). Results of the analysis showed that the highest silhouette scores are consistently obtained for K=2 preference across experimental conditions, with a slow decrease in the score as K increases until a plateau towards K=6.

Naïve model. For each participant and block, each raw ITI was categorized with its cluster's label, based on the block-specific k-preference either for 2 or 3 clusters (e.g. for K=2 as preferred clustering, each data point was categorized either as 1 or 2). The result of this procedure was a series of labels for each block, that was treated as a series of "known" states and was fed to a naïve model to estimate the transition probabilities ruling the states' alternations (parameter Φ). Results of parameter Φ estimation for one participant are shown in **Figure 4.8A**. The flexibility of k-means in defining the number of durations performed by the subject allows, in turn, a more accurate estimation of the transition probabilities underlying the series reproduced, and it reveals that participants were learning to the same extent each experimental condition's high-level statistics. This is evident when comparing the matrices estimated by the original model (**Figure 4.3A**) and those estimated by the naïve model for Pattern 1 condition in the same subject. Once the number of states/durations is reduced to 2, the transition probabilities approximate those governing the stimuli series. Based on the block-preferred number of clusters used for classifying ITI series, we reconstructed a correspondent discrete series of ISI states for each block. This was done by labelling raw ISI according to the closest centers among those found by k-means. For both ITI and ISI transformed series, we calculated the proportion of each cluster within the series and obtained a k-means-based reconstruction of parameter θ (**Figure 4.8B**). This qualitative analysis further confirmed that participants were learning the underlying statistical properties of each experimental condition. Given that the estimated matrices can vary in dimension across blocks and conditions,

no general metric was extracted to carry out statistical comparisons of naïve model results at the group level.

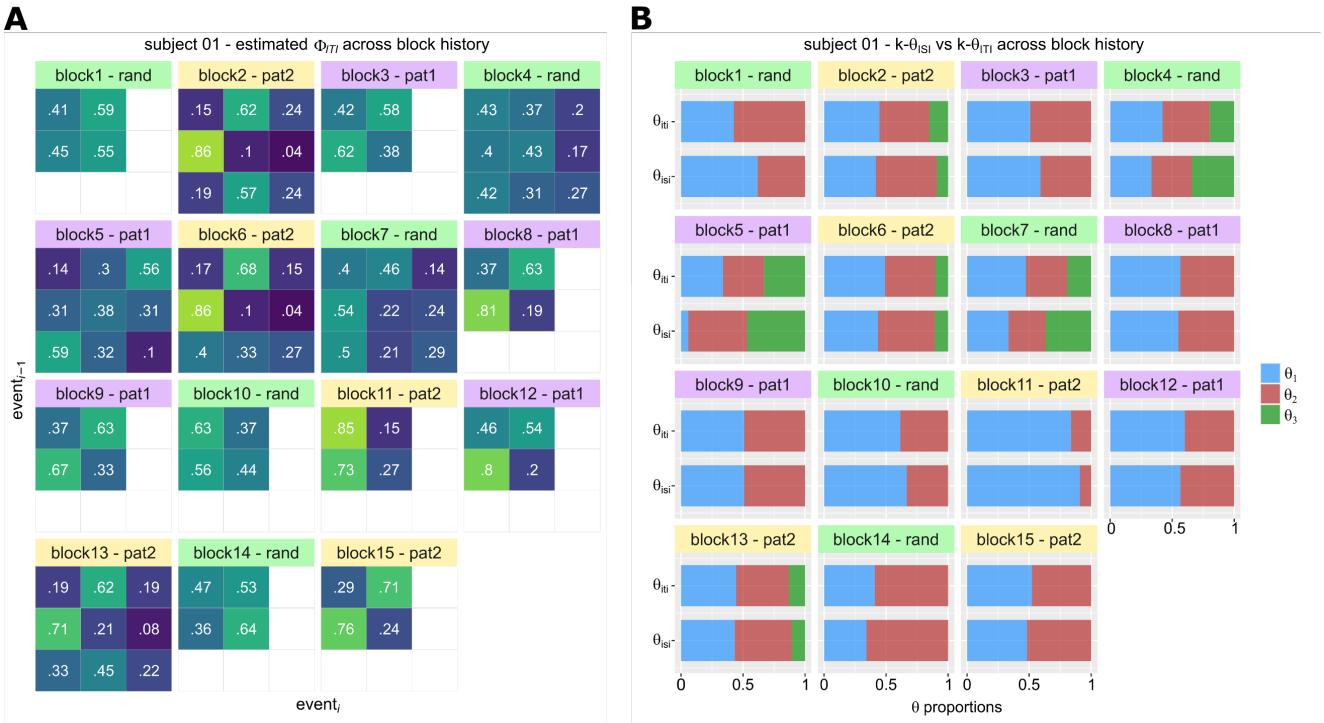


Figure 4.8. **Naïve model parameter Φ_{ITI} estimation and k-means-derived θ_{ISI} and θ_{ITI}.** (A) The plot shows results of parameter Φ_{ITI} estimation for subject 01: for each block the naïve model estimates a k-by-k transition matrix describing the set of conditional probabilities underlying the known state sequence, where k is the preferred number of clusters for the series of durations reproduced in that specific block. Each facet's title indicates block number and block-specific experimental condition (color coded as green for Random, purple for Pattern 1, yellow for Pattern 2 condition). The color hue and the number inside each tile indicate the block-specific conditional probability of the current $event_i$ being one of the two/three durations, given the identity of the previous duration (i.e. $event_{i-1}$). (B) The plot represents θ_{ISI} (lower barplot) and θ_{ITI} (upper barplot) reconstruction based on k-means clustering results (i.e. each cluster's proportions within the series of ISI and ITI, respectively). Each color here represents one cluster (blue = 1st cluster, red = 2nd cluster, green = 3rd cluster). ISI = Inter Stimulus Interval; ITI = Inter Trial Interval.

4.4.4 Comparison between modelling approach and data-driven approach

To directly compare results obtained by model-based analysis and by k-means analysis, we calculated the degree of overlap between the posterior distributions over the three estimated μ i.e. the overlap between $\mathcal{N}(\mu_1, \sigma_1)$ and $\mathcal{N}(\mu_2, \sigma_2)$ distributions (hereafter overlap μ_1-μ_2) and the overlap between $\mathcal{N}(\mu_2, \sigma_2)$ and $\mathcal{N}(\mu_3, \sigma_3)$ distributions (hereafter overlap μ_2-μ_3). This procedure resulted in two μ-overlap values for each block of each participant.

To assess how consistently, across blocks, these values were associated to a preference for K=2 clusters by k-means and how they were modulated by each experimental condition, we fit a LME model, with the two μ-overlap values as dependent variables, and experimental condition, overlap type (μ_1-μ_2 or μ_2-μ_3), and k-means preference (2 or 3) as independent variables in interaction, along with participants as random intercept (model formula: overlap values ~ overlap type * experimental condition * K-preference + (1 | subject)). Goodness of fit of the model was evaluated using R^2 statistic ($R^2_{marginal}$ = 0.33, $R^2_{conditional}$ = 0.45). Type III ANOVA on model estimates showed significant main effects of all three predictors, as well as two-way and three-way interactions. Specifically, ANOVA results revealed a main effect of overlap type (F(1, 597.99) = 86.585, p<0.0001), because across blocks the posterior distributions of μ_2 and μ_3 have a greater overlap compared to the posterior distributions of μ_1 and μ_2 (**Figure 4.9A**). The main effect of experimental condition (F(2, 600.24) = 31.318, p<0.0001), was due to the overall greater μ-overlap values for Pattern 1 condition, compared to Random (rand–pat1 = -0.0452, t.ratio(601) = -3.121, p<0.01) and Pattern 2 condition (pat1–pat2 = 0.1105, t.ratio(601) = 7.664, p<0.0001), and greater μ-overlap values for Random compared to Pattern 2 condition (rand–pat2 = 0.0652, t.ratio(599) = 5.111, p<0.0001) (**Figure 4.9B**). The main effect of K-preference (F(1, 615.7) = 34.497, p<0.0001), was due to the fact that blocks that are best clustered by K=2 have higher μ-overlap values compared to blocks that are best clustered by K=3 (**Figure 4.9C**). This last main effect points towards the "shrinkage" hypothesis.

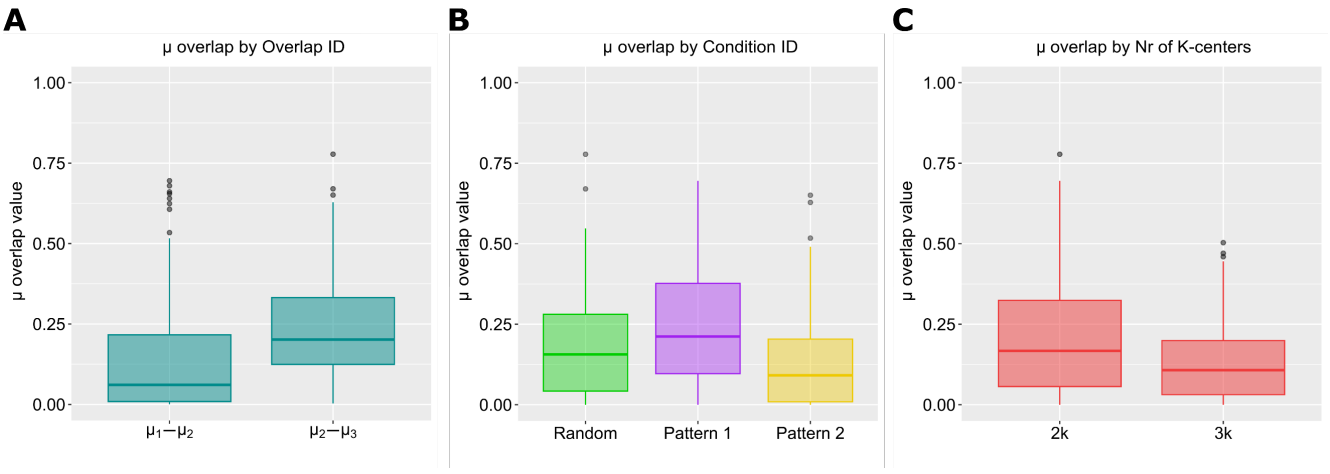


Figure 4.9. **Effects of μ-overlap type, experimental condition and K-preference on μ overlap values**. (A) Overlap values grouped by μ-overlap type: overlap μ_2-μ_3 is overall significantly higher than overlap μ_1-μ_2. (B) Overlap values grouped by experimental condition: Pattern 1 has the greatest degree of μ-overlap between estimated posterior distributions compared to the other condition, whereas Pattern 2 condition has the lowest degree of μ-overlap. (C) Overlap values grouped by K-preference: blocks in which k-means preference is for 2 clusters are consistently associated to a greater degree of overlap between estimated posterior distributions by the model.

ANOVA results also showed a significant 2-way interaction between overlap type and experimental condition ($F(2, 597.99) = 16.931$, $p<0.0001$), with condition-dependent differences in μ-overlap values (**Figure 4.10A**). This interaction was due to a significantly higher overlap μ_2-μ_3 compared to the overlap μ_1-μ_2 both for Random (μ_1-μ_2 vs μ_2-μ_3 = -0.17046, t.ratio(598) = -9.405, $p<0.0001$) and Pattern 2 condition (μ_1-μ_2 vs μ_2-μ_3 = -0.13671, t.ratio(598) = -7.660, $p<0.0001$), and comparable values of overlap μ_2-μ_3 and μ_1-μ_2 for Pattern 1 condition (μ_1-μ_2 vs μ_2-μ_3 = -0.00776, t.ratio(598) = -0.347, $p=0.999$). This interaction highlights an important difference between conditions. For Random and Pattern 2 conditions the posterior distributions of μ_1 and μ_2 have overlap values close to zero. This means that in most of the blocks of these conditions the estimated distributions of μ_1 and μ_2 are clearly separated, whereas the distributions of μ_2 and μ_3 show high overlap, meaning that they are estimated as closer. In Pattern 1 condition, instead, high overlaps are equally present both between μ_1 and μ_2 distributions and between μ_2 and μ_3 distributions, and this means that in all blocks of this condition both pairs are estimated as close.

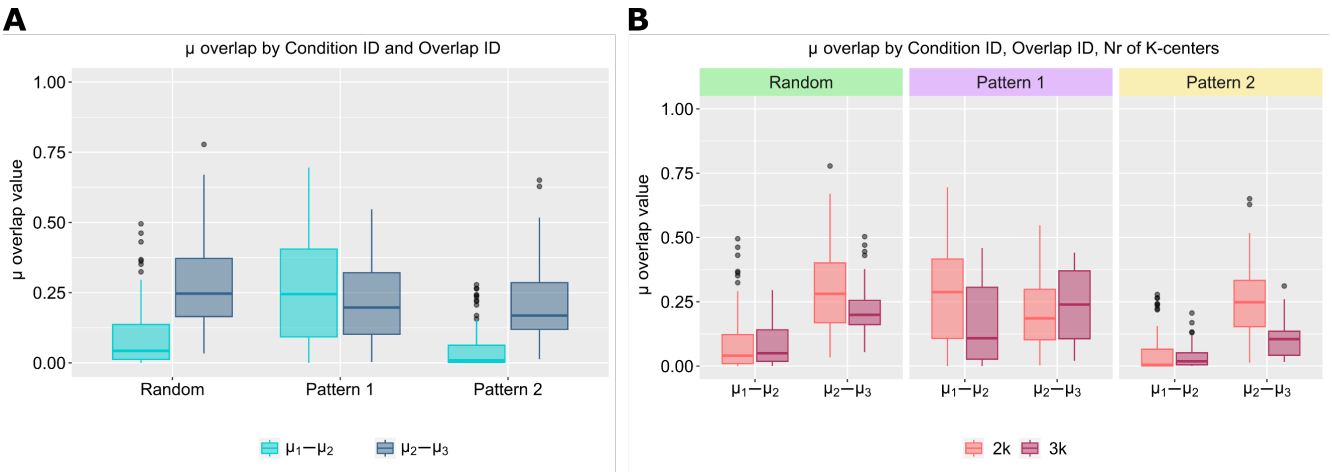


Figure 4.10 **Interaction effects on μ overlap values**. (A) The plot shows overlap values grouped by μ-overlap type and experimental condition: a 2-way interaction was found between these two variables, whereby both Random and Pattern 2 conditions have significantly higher values for overlap μ_2-μ_3 compared to overlap μ_1-μ_2, whereas Pattern 1 condition has comparable levels of overlap between distributions. (B) This summary plot represents overlap values grouped by μ-overlap type, K-preference (color coded as light red for K=2 and dark red for K=3) and experimental condition (color coded as green for Random, purple for Pattern 1 and yellow for Pattern 2 condition): a 3-way interaction emerged from statistical comparisons of overlap values, whereby in Pattern 1 condition a greater degree of overlap μ_1-μ_2 is associated with k-means preference for 2 clusters, and in Pattern 2 condition the k-means preference for 2 clusters is consistently associated with overlap μ_2-μ_3.

Finally, ANOVA on model fit also showed a 3-way interaction between overlap type, experimental condition and K-preference ($F(2, 597.99) = 12.126$, $p<0.0001$), which is represented in **Figure 4.10B**. The 3-way interaction revealed the presence of a greater overlap μ_2-μ_3 consistently associated with a preference for K=2 clusters in Pattern 2 only (K2/μ_2-μ_3 vs K3/μ_2-μ_3 = 0.1658, t.ratio(607) = 6.401, $p<0.0001$). And in a totally opposite direction, we found the presence of a

greater overlap μ_1-μ_2 consistently associated with a preference for K=2 clusters in Pattern 1 condition only (K2/μ_1-μ_2 vs K3/μ_1-μ_2 = 0.1316, t.ratio(603) = 4.104, p=0.0027). The higher overlap μ_2-μ_3 found in the Random condition is not significantly associated with any of the two K-preference (K2/μ_2-μ_3 vs K3/μ_2-μ_3 = 0.0832, t.ratio(607) = 3.166, p=0.07).

This 3-way interaction helps to clarify the main effects discussed above. Concerning the overlap μ_2-μ_3, whereas it was found generally higher across all experimental conditions, compared to the overlap μ_1-μ_2, only in Pattern 2 condition this overlap was consistently associated with a preference for K=2 clusters. Concerning Pattern 1 condition instead, we observed a general greater degree of μ-overlap compared to Pattern 2 and Random condition However, only for the overlap μ_1-μ_2 an interaction with a preference for K=2 clusters was found. This last result means that the overlap between μ_1 and μ_2 distributions in Pattern 1 condition was systematically associated with a 2-durations clustering by k-means, whereas this clear association was not present for the overlap between μ_2 and μ_3 distributions.

Overall, this last analysis shows the convergence between model-based and data-driven analysis. This analysis reveals that the degree of overlap between posterior distributions over the three μ is linked to the k-means clustering results and they both suggest a potential shrinkage of the "duration space" defining our participants' performance in the two Pattern conditions. In Pattern 2 condition this shrinkage penalizes the longest duration of the three, whereas in Pattern 1 condition it affects mainly the shortest duration.

4.5 DISCUSSION

In the current study we sought to investigate the process by which humans learn statistical regularities embedded in temporal patterns whose structure is not predetermined or fixed, but defined by a set of statistical rules. To this aim we employed three 1st order Markov processes (i.e. our experimental conditions) to generate multiple sequences in which three durations (i.e. states) alternated according to each process' statistical properties. The generated sequences were presented to human participants in a delayed temporal reproduction task. The reproduced series fed an ideal observer model, which estimated the parameters defining the statistical properties learned by participants. These parameters were the mean and standard deviation of the three durations reproduced, their proportions within the series and the transition probabilities ruling their alternations.

Results from model estimation showed that participants were able to extract and learn the statistical rules underlying the series of durations presented, and could discriminate between more and less predictable generative processes. This was evidenced by the comparison between entropy measures derived from estimated parameter Φ and, in particular, from the conditional entropy of the estimated matrices. Entropy values indicated more predictable alternations in the two Pattern conditions than the random one. On the contrary, statistical comparisons on MSE measures suggested a greater variability for Pattern 1 condition compared to the other two. This measure,

though, informs only on the distance between the stimuli transition matrices and those estimated from participants' performance. In this sense, it is a measure of accuracy with respect to an external reference, not a measure of how much the reproduction is pattern-biased. Overall, modelling results confirmed that humans deal with non-deterministic temporal patterns by exploiting their low- and high-level statistical properties, as they do for other types of sequences (e.g., Hunt & Aslin, 2001; Kirkham et al., 2002; Koelsch et al., 2016), and that this learning process can be modelled in terms of Bayesian inference (Meyniel et al., 2016; Maheu et al., 2022).

The combination of model-based and data-driven analysis revealed that the series reproduced by our participants were also characterized by additional statistical properties along with those featured in the stimuli series: namely a "shrinkage" of the space of possible durations-states performed, central tendency and history effects. These effects were present to a greater extent in the more predictable experimental conditions (Pattern 1 and 2). In particular, analysis on raw Inter Tap Intervals mean and error highlighted how durations were either under-reproduced or over-reproduced, in Pattern 1 and 2 conditions respectively. The results of this analysis revealed a diffuse regression toward the mean combined with a negative linear trend across block history. Central tendency is a well-studied effect in interval timing processes (Jazayeri & Shadlen, 2010). The regression toward the mean of the distribution of intervals experienced is a strategy to optimize performance, in particular when the sensory representation of the duration perceived is noisy (Cicchini et al., 2012). The fact that our sequences were not fixed but characterized by a high level of uncertainty may have prompted the extraction of an even more general low-level statistics than the condition-specific mean, i.e. the overall mean of the global temporal context, to optimize the reproduction.

Results from k-means analysis, on the other hand, showed how raw data were overall best clustered using two duration categories instead of three: this reduction in the number of possible durations was often associated with a more accurate rendering of the transition probabilities defining both Pattern conditions, as revealed by our naïve model. This finding was linked to the results obtained in the model estimation of the posterior distribution of the durations performed. A greater degree of overlap between durations' distribution was found to be significantly associated to a preferred clustering of two durations by the k-means analysis. This association was pattern-specific: in the case of Pattern 1 condition, where the medium and the long durations appear and alternate more frequently, those blocks showed a great degree of overlap with the posterior distributions of μ_1-μ_2 (i.e. small and medium durations) clustered with two durations instead of three by k-means. The opposite was true for Pattern 2 condition, where the small and medium durations appear and alternate more often. Blocks with high overlap between posterior distribution of μ_2-μ_3 (i.e. medium and long durations) were most of the time best clustered using two categories by k-means. No such association was found for the Random condition.

These mirroring results suggest that, whenever participants were catching the presence of a pattern in the series presented, their reproduction reflected a simplified version of the statistical properties underlying the series (e.g., performing only two of the most frequent intervals). This effect may

be due to the tendency to perceive the intervals of a rhythmic pattern as discrete categories and not on a continuous scale (Desain & Honing, 2003). In the two Pattern conditions of our experiment it is possible that participants reduced the number of durations performed to the most salient categories. Another closely related explanation lies on the effect of rhythmic priors in situations in which the temporal structure of the stimuli is characterized by a high level of uncertainty: when humans are asked to reproduce a random stream of temporal intervals, they end up imposing a rhythmical structure on it, and this structure reflects cultural bias on rhythms (Ravignani et al., 2016; Jacoby & McDermott, 2017). One of these priors is represented by western populations' preference for simple duple and triple meters, compared to other rhythmic structures (Savage et al., 2015). In the specific case of our study, such a rhythmic prior may have interacted with the statistics learned from the stimuli and this interaction might have contributed in creating a simplified representation of the two Pattern conditions.. It is unclear, though, the potential effect of a duple meter prior in the reproduced series in the Random condition. In fact in this condition we didn't observe a clear association between a 2-durations clustering and the overlap between estimated durations' distribution.

Model optimization. The model proposed in the current study has a good potential in describing how humans learn the statistical properties of non-deterministic temporal sequences. The combination with an agnostic, data-driven approach, however, demonstrated how this model can be optimized in several ways in order to be fully predictive of humans' behavior. More specifically, it should incorporate the same flexibility used by k-means in estimating the number of state/durations in the series. The priors used should combine information coming from physical stimuli with perceptual (i.e., central tendency) and cultural (i.e. rhythmic priors) biases that adult participants are expected to show. Finally, the model should be able to accurately simulate series of reproduced durations that mimic the statistical properties of those performed by real participants.

4.6 CONCLUSION

In this study we aimed at investigating the process by which human learn the underlying statistical properties of non-deterministic temporal patterns. By employing an ideal observer Bayesian model, we were able to retrieve the parameters describing which statistics our participants were extracting and expressing in their reproduction and compare them to the generative processes underlying our stimuli.

The combination of model-based and model-agnostic analysis revealed that participants performance reflected the low- and high-level statistics embedded in the series of temporal intervals they perceived. In addition it revealed that in a more predictable conditions, participants' reproduction was often characterized by a reduced set of durations and influenced by a central tendency effect. The proposed model can be improved by incorporating greater flexibility in the estimation of the number and identity of the durations reproduced, but also by weighting different types of prior knowledge in duration reproduction.

In summary, the present study offers a relatively novel perspective to study how humans learn temporal regularities in their environment, in particular when these regularities are not fully predictable.

Chapter 5
General Discussion

The research presented in the current thesis is centered on the general topic of the processing and learning of temporal patterns. Despite the lively interest in this topic and the numerous studies coming from multiple research fields, several questions concerning the processing and learning of temporal patterns are still lacking exhaustive answers. In the current work I tried to tackle some of these questions by adopting a "crossroad" perspective: I drew on timing research and sequence processing studies to operationalize my research questions; I adapted existing designs to the temporal domain and developed novel ones; I combined in each study multiple analytical approaches and techniques to answer our questions. Indeed, one of the aims of this work is to offer new (or less traditional) combinations of tools to study temporal patterns. In the next paragraphs a summary of the main results is given for each of the studies conducted, along with a general discussion of the main findings across studies.

5.1 Summary of the findings

In the first group of experiments (*chapter 2*) we investigated how the reproduction of a temporal sequence is affected by different levels of violation of its regularity, namely a violation within its elements (i.e. local deviance) and of its overall structure (i.e. global deviance). Additionally, we asked whether the behavioral signatures of these violations would change according to the complexity of the sequence's structure and to the sensory modality through which the sequence is experienced (i.e. visual or auditory). These questions were tackled by three behavioral experiments in which we adapted the local-global paradigm (Bekinschtein et al., 2009) to the temporal domain, specifically to a Synchronization Continuation task.

Analysis of participants' performances in the frequency-domain were carried using a measure of similarity between the power spectrum of the tapping series and that of the stimuli series. Overall results showed a general effect of local deviance in all three experiments: the precision in reproducing a sequence was invariably affected by its internal irregularity (i.e. local deviance), independent of whether the sequence was globally regular or not. This effect was greatest when using auditory sequences (2nd experiment). However, when using visual sequences, we observed also global deviance effects, which were modulated by the complexity of the structures used. In the case of simple structures (1st experiment), a worsening of performance emerged only when the violation of the sequence was both at the local and global level. When using more complex structures (3rd experiment) the global deviance effect was independent of the local status of the sequence, and it was specifically observed during the Synchronization phase of the task. In other words, the tapping was less precise when synchronizing with a complex unfamiliar sequence (global deviant), no matter if it was locally regular or deviant.

Analysis of participants' performances in the time-domain using Procrustes technique showed that in all three experiments the detrimental effects observed in the frequency-domain were partially explained by specific sources of distortions in the tapping, namely a systematic bias and a tendency to tap rescaled versions of the sequences. These distortions were prominently and consistently present during the Continuation phase of the task. Specifically, the reproduction of locally deviant sequences was always characterized by a delay and a slowing down of the tapping. The phase-specificity of this effect was probably due to the fact that during the Continuation phase participants' tapping relied only on an internal reference, and error correction mechanisms, typically observed during the Synchronization phase, were absent. This in turn might have led to characteristic "drifts" in the tapping frequency spectrum, an effect already reported in the literature (Repp, 2005).

The second study (*chapter 3*) focused on the learning of new temporal structures. Specifically, we asked how humans switch from a known to a novel temporal structure made by the same constitutive elements, and how this transition is influenced by trial predictability. To answer this question, we conceived a novel design in which we parameterized a gradual transition between two mirroring temporal patterns (i.e. pattern A and B), by manipulating their probability of occurrence across trials. The two patterns were embedded in auditory streams and presented in a synchronization task. We analyzed tapping data combining standard finger tapping analysis and model-based analysis. The general aim of the model-based analysis was the definition and estimation of the parameters describing the evolution of the learning process underlying the transition between temporal structures, as expressed in participants' performances. The two models presented in this thesis are meant to be considered the first steps toward this general goal. Specifically, the first model gave us a trial by trial estimation of the proportions in which the two patterns were present in tapping performances. The second model allowed us to increase the temporal resolution of the estimation of these proportions by computing it at the single tap-level. This model provided also an estimation of tapping variability and of two static learning rates which rules the update of the proportions of the two patterns.

Results from standard behavioral analysis and from both models showed that pattern A was quickly learned by participants, who tended to use it as a reference for the tapping in those trials in which the predictability was very low (i.e. middle trials). Indeed, both error and asynchrony measures remained lower compared to pattern B throughout the majority of the trials. The estimation of the two patterns' proportions across trials and across taps from the two models confirmed this result. In fact, the probability of performing pattern A was estimated as greater than pattern B even beyond the point of transition designed in our stimuli, suggestive of a carry-over effect of pattern A. This means that participants required to experience pattern B several times to be able to switch to it. The second model also allowed us to identify two types of performance among our subjects and link parameters' estimation to specific behavioral signatures. Specifically, a group of "fast learners" learned to synchronize with pattern A almost immediately and showed higher sensitivity to the ongoing changes in patterns' proportions across trials, as revealed by their higher learning rates. This group also showed a sharper transition to pattern B, expressed in its

estimated proportion, once its presence became more frequent. This "sharpness" was confirmed by the steeper decrease in pattern B reproduction error across trials. A second group of "slow learners" showed, instead, a slow increase in the estimated proportion of pattern A, meaning that they needed more trials to learn to synchronize with it. Likewise, for this group the transition to pattern B was very slow, as suggested both by its estimated proportion and the slow decrease of its error across trials. On average this group also showed less sensitivity to changes in the two patterns' proportions trial by trial, as revealed by the very small estimated learning rates. This suggests that these participants were relying more on their previous taps rather than on the external stimuli. On a final note, results from parameter recovery highlighted some important limitations of the second model that the current model optimization is trying to overcome. These limitations concern the need of a better estimation of tapping variability and of a dynamic estimation of learning rates to take into account trial-specific levels of uncertainty.

The third study (*chapter 4*) addressed the learning of temporal patterns from a different perspective compared to the second study. Specifically, we asked whether and how humans extract and learn statistical regularities embedded in sequences of temporal intervals, which do not have fixed structures, yet are defined by underlying statistical rules that can be learned. To this aim we adopted a probabilistic approach and created a novel experimental design in which we defined the generative processes (i.e. our experimental conditions) based on which we created multiple sequences of temporal intervals. These generative processes were defined by low- and high-level statistical properties, namely the mean and entropy rate of their stationary distribution and the transition probabilities ruling the alternation between the intervals in the sequences. We presented the generated sequences to human participants in a delayed temporal reproduction task and developed an ideal Bayesian observer model to estimate the parameters defining the statistical properties participants learned. Results from model estimation showed that participants were able to extract the statistical rules underlying the series of durations presented, and could discriminate between more and less predictable generative processes. This was expressed by condition-specific entropy and error measures derived from the estimated transition probabilities underlying the reproduced series.

Given the "informed" nature of our model-based analysis, we also analyzed participants' performances adopting an agnostic, data-driven approach. This second analysis revealed that the series reproduced featured additional statistical properties that interacted with those learned from the stimuli series. In particular, the analysis on mean reproduced duration across conditions showed the presence of a central tendency effect in the predictable conditions combined with a general history-dependent negative linear trend. K-means analysis revealed that raw data were in general best clustered into two duration categories, instead of the three used to generate the stimuli series, highlighting the presence of a "shrinkage" in the space of possible durations performed. These results were directly compared with the three durations' distributions estimated by our model. Overall, we found that a greater overlap between the estimated durations' distributions was consistently associated with a 2-durations clustering by k-means. This association was also

condition-specific. Specifically, in the predictable condition (in which the medium and long durations are alternating frequently) the 2-durations clustering was significantly associated with a greater overlap between the smallest and the medium durations' distributions. In the other predictable condition (in which the small and the medium durations alternate frequently), the association was reversed: the 2-durations clustering was significantly associated with a greater overlap between the medium and the long durations' distribution. In other words, whenever participants were detecting the presence of a pattern in the stimuli series, their reproduction reflected a reduction in the number of durations performed towards the two durations most represented in the series heard. The combination of model-based and data-driven analysis allowed to confirm that our participants were learning the low- and high-level statistics embedded in the series of temporal intervals they perceived. It also revealed, though, that in more predictable conditions their reproduction reflected a simplified version of these properties, both in terms of number of durations performed and regression toward the mean of the durations distribution.

5.2 Discussion

Overall results from the three experiments presented in chapter 2 showed how local and global violations of a temporal pattern differentially affect its reproduction, according to the sensory modality in which the pattern is experienced and the complexity of its structure. Specifically, the presence of both local and global violations effects in the third experiment seem to suggest that an accurate reproduction of complex patterns requires the processing and integration of both local elements and global structure. In fact, if the overall structure of such complex patterns had been ignored by our participants, we would have found only local deviance effects across all three experiments. The reproduction of simple patterns, instead, was always more affected by local violations. In this case the internal regularity of the pattern's elements was essential to perform a precise reproduction of it. It is important to stress that the simple patterns we employed were isochronous and nearly isochronous rhythms, which may be easily reproduced without the need to integrate single elements into a global structure. It is highly likely that reproducing such periodic patterns automatically recruited an entrainment-based mechanism (Repp, 2005; Cadena-Valencia et al., 2018). The recruitment of such mechanism is also considered at the base of the perception, production and synchronization with periodic, beat-inducing patterns, through the engagement of a tightly coupled network of auditory, supplementary motor, premotor and subcortical areas (for reviews see Grahn, 2012; Merchant et al., 2015). The stronger coupling between motor and auditory, compared to visual, regions is thought to support the auditory advantage on the visual modality when producing or synchronizing with a rhythm (Repp & Su, 2013). This greater accuracy and easiness with which humans synchronize with auditory rhythms, compared to visual ones, is a behavioral finding extensively reported in the literature (Repp & Penel, 2002; Zarco et al., 2009). Indeed, we found the greatest precision of tapping performance when using simple auditory patterns. It would be interesting, however, to test local and global violations effects using patterns with simple rhythmic (but not isochronous) structures. In this case it may be possible that

a certain degree of integration of local elements into a global structure is involved in processing the pattern.

An interesting finding from these three experiments is also the frequent presence of a scaling component in the reproduced patterns during the Continuation phase of the task, in particular in the locally deviant patterns. This effect is suggestive of the acquisition of their overall structure: it would be very difficult to reproduce rescaled versions of a pattern if its structure was not encoded. This observation is even more relevant in the case of complex, aperiodic patterns like those used in our third experiment. If the processing of these patterns was based only on the absolute duration of their elements, i.e. by using an interval-based mechanism, we wouldn't have observed any scaling effect. Indeed, a study by Hardy and colleagues (2018) demonstrated that complex aperiodic temporal patterns, once they are learned, can be reproduced in rescaled versions with a good level of accuracy.

Results from the study presented in chapter 3 highlighted the role of the predictability of the environment when learning novel temporal structures. Model-based analysis were also able to describe how such predictability interacts with subject-specific strategies employed in this learning process. Our findings suggest that when the external stimuli are highly unpredictable there is a general tendency among participants to rely on a learned temporal structure to perform the task. As a consequence of this tendency, even when a novel structure is experienced to the same extent as the familiar one, participants opt for sticking to what is known and expected based on the past trials, and ignore the novel structure until its presence is well established. This general behavior appears to be modulated by subject-specific strategies that differ in the weight given to the changes in the external stimuli and in the quickness in adapting the behavioral response to these changes. Adopting the broad framework of predictive coding and Bayesian inference, an optimal performer of our task would employ a flexible trade-off between relying on an internal reference and on external stimuli, and would be able to adjust the internal reference as soon as a new structure is consistently experienced. And indeed, a portion of our participants behave in such a way, as it was shown by models' results. A theoretical account in line with this rationale (Vuust & Witek, 2014; Koelsch et al., 2019) proposes that tracking and synchronizing with rhythms entail a constant weighing and update of our internal expectations about a rhythmic metrical structure (i.e. top-down prediction) based on the sensory stimulus we are experiencing (i.e. bottom-up information). In this sense, this continuous comparison is well described in terms of Bayesian inference and is consistent with the predictive coding framework. Some experimental evidence supporting this theoretical account comes from studies that have shown the typical brain signature of prediction error (i.e. Mismatch Negativity) in response to rhythmic incongruity (Vuust et al., 2009; Lumaca et al., 2019).

The role of the predictability of the environment, or conversely its level of uncertainty, is a key notion also to understand the complex set of results of the study presented in chapter 4. This study gives interesting insights on how humans deal with non-deterministic temporal patterns, i.e. patterns that feature some degree of unpredictability. Overall, we showed that participants were

able to learn the low- and high-level statistical properties of the sequences they experienced. But we also showed that additional statistical properties were at play and interacted with those underlying the physical stimuli. These additional effects emerged more clearly in the two predictable conditions, i.e. those in which there was a more frequent alternations between two out of the three durations employed.

One of these additional effects concerns the reduction in the number of categories of the durations performed, compared to those experienced in the stimuli. In the two predictable conditions, in particular, this reduction was clustered in a mirrored way, as if participants were performing only the two intervals that appeared more frequently, resulting in the reproduction of simplified versions of the series they heard. In general, this effect may have been prompted by a categorical representation of the temporal intervals experienced, reducing the number of intervals to the most salient categories. This categorical representation has been demonstrated in the perception and production of both rhythmic patterns (Desain & Honing, 2003) and random series of intervals (Ravignani et al., 2016; Jacoby & McDermott, 2017). In particular, these studies have shown how this clustered representation of temporal patterns tends to gravitate toward few duration categories which are usually related by small integer ratio (Jacoby & McDermott, 2017). A closely related explanation for our findings, indeed, concerns the potential effect of humans' rhythmic priors triggered by the high level of uncertainty present in our stimuli. Specifically, western populations' preference for simple duple meters (Savage et al., 2014) may have biased participants' reproduction, particularly in the two predictable conditions, resulting in a simplification of the statistical properties that the subjects were learning. Indeed, it has been shown how such priors can be elicited even by series of intervals which totally lack temporal structure (Ravignani et al., 2016). It is not completely clear, though, how such duple meter prior interacted with the statistical properties embedded in our Random condition. In this experimental condition, in fact, the reduction of the durations' categories was not clearly clustered towards two specific durations among the three we used.

A second important additional effect we found in participants' reproduction was a general regression toward the global mean of the durations presented. In fact, central tendency effects were prominent in the two predictable conditions, where the stimuli mean was either lower or higher than the global mean of the whole distribution of durations used. This difference with respect to the Random condition makes sense, given that its mean duration overlapped with the global mean. Central tendency effects in interval timing judgement and reproduction tasks have been successfully described in terms of Bayesian inference (Jazayeri & Shadlen, 2010; Cicchini et al., 2012). The extraction of the mean duration within a distribution of intervals acts as prior information to optimize task performance and this information is weighted differently according to the perceived noisiness of the temporal "environment". The less precise is the sensory representation of the durations experienced, the more the subject will use this general prior as a strategy to perform the task (Cicchini et al., 2012). In this sense the high level of uncertainty characterizing our sequences of stimuli may have pushed our subjects to perform the task relying

on the simpler statistic underlying the stimuli experienced, i.e. the overall mean of the global temporal context.

In light of these considerations, future implementations of the model we developed in the last study should incorporate flexible priors that combine global statistics on the physical stimuli experienced (i.e. central tendency) and cultural biases (i.e. rhythmic priors).

APPENDIX A – SUPPLEMENTARY MATERIAL CHAPTER 2

Contrast	Estimate	Std Error	z value	p-value
Reg A – Dev A	0.016520	0.041905	0.394	0.9792
Dev B – Dev A	-0.095564	0.041905	-2.281	0.1027
Reg B – Dev A	**0.011966**	**0.041905**	**0.286**	**0.9919**
Dev B – Reg A	**-0.112084**	**0.041905**	**-2.675**	**0.0375**
Reg B – Reg A	-0.004554	0.041905	-0.109	0.9995
Reg B – Dev B	0.107530	0.041905	2.566	0.0504

Table A1T. MSC analysis Exp1V. Pairwise post-hoc comparisons (interaction trial type x sequence type). Comparisons of interest are highlighted in bold.

Contrast	Estimate	Std Error	df	t.ratio	p-value
Dev A – Reg A	0.00847	0.00526	75	1.609	0.3799
Dev A – Dev B	0.00410	0.00526	75	0.779	0.9596
Dev A – Reg B	**-0.00261**	**0.00526**	**75**	**-0.497**	**0.9596**
Reg A – Dev B	**-0.00437**	**0.00526**	**75**	**-0.83**	**0.8399**
Reg A – Reg B	-0.01108	0.00526	75	-2.106	0.1607
Dev B – Reg B	-0.00671	0.00526	75	-1.275	0.5814

Table A2T. Procrustes analysis Exp1V. Pairwise post-hoc comparisons on scaling-synchronization (interaction trial type x sequence type). Comparisons of interest are highlighted in bold.

Contrast	Estimate	Std Error	df	t.ratio	p-value
Dev A – Reg A	0.01507	0.00676	75	2.228	0.1252
Dev A – Dev B	0.00963	0.00676	75	1.424	0.4885
Dev A – Reg B	**0.00235**	**0.00676**	**75**	**0.348**	**0.9854**
Reg A – Dev B	**-0.00544**	**0.00676**	**75**	**-0.803**	**0.8526**
Reg A – Reg B	-0.01272	0.00676	75	-1.88	0.2455
Dev B – Reg B	-0.00728	0.00676	75	-1.076	0.705

Table A3T. Procrustes analysis Exp1V. Pairwise post-hoc comparisons on scaling-continuation (interaction trial type x sequence type). Comparisons of interest are highlighted in bold.

Contrast	Estimate	Std Error	df	t.ratio	p-value
Dev A – Reg A	-0.116	0.0815	75	-1.423	0.4892
Dev A – Dev B	-0.0896	0.0815	75	-1.099	0.6913
Dev A – Reg B	**0.0129**	**0.0815**	**75**	**0.159**	**0.9986**
Reg A – Dev B	**0.0264**	**0.0815**	**75**	**0.324**	**0.9882**
Reg A – Reg B	0.1289	0.0815	75	1.582	0.3951
Dev B – Reg B	0.1025	0.0815	75	1.258	0.5925

Table A4T. Procrustes analysis Exp1V. Pairwise post-hoc comparisons on bias-synchronization (interaction trial type x sequence type). Comparisons of interest are highlighted in bold.

Contrast	Estimate	Std Error	df	t.ratio	p-value
Dev A – Reg A	-0.29796	0.144	75	-2.069	0.1728
Dev A – Dev B	-0.3008	0.144	75	-2.089	0.1663
Dev A – Reg B	**-0.15531**	**0.144**	**75**	**-1.078**	**0.7037**
Reg A – Dev B	**-0.00284**	**0.144**	**75**	**-0.02**	**1.000**
Reg A – Reg B	0.14264	0.144	75	0.99	0.7553
Dev B – Reg B	0.14549	0.144	75	1.01	0.744

Table A5T. Procrustes analysis Exp1V. Pairwise post-hoc comparisons on bias-continuation (interaction trial type x sequence type). Comparisons of interest are highlighted in bold.

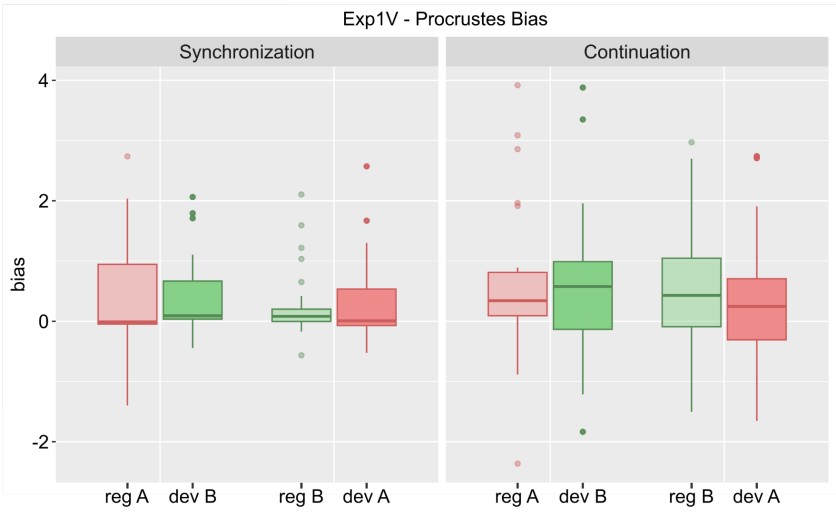
Contrast	Estimate	Std Error	df	t.ratio	p-value
Dev A – Reg A	0.008	0.00479	78	1.671	0.3458
Dev A – Dev B	0.0236	0.00479	78	4.932	< 0.0001
Dev A – Reg B	**0.00806**	**0.00479**	**78**	**1.684**	**0.3389**
Reg A – Dev B	**0.0156**	**0.00479**	**78**	**3.261**	**0.0088**
Reg A – Reg B	0.000062	0.00479	78	0.013	1.000
Dev B – Reg B	-0.0155	0.00479	78	-3.248	0.0091

Table A6T. Procrustes analysis Exp2A. Pairwise post-hoc comparisons on scaling-continuation (interaction trial type x sequence type). Comparisons of interest are highlighted in bold.

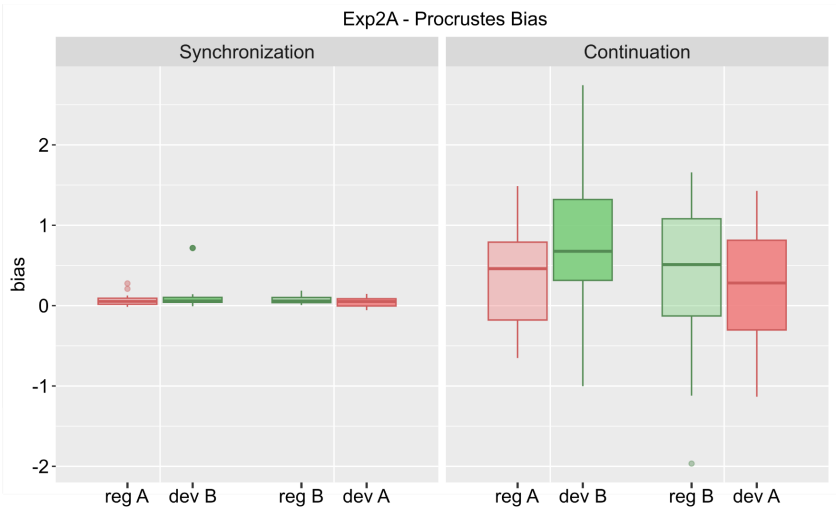
Contrast	Estimate	Std Error	df	t.ratio	p-value
Dev A – Reg A	-0.1903	0.12	78	-1.584	0.3936
Dev A – Dev B	-0.5812	0.12	78	-4.837	<0.0001
Dev A – Reg B	**-0.2041**	**0.12**	**78**	**-1.699**	**0.3311**
Reg A – Dev B	**-0.3909**	**0.12**	**78**	**-3.253**	**0.009**
Reg A – Reg B	-0.0138	0.12	78	-0.115	0.9995
Dev B – Reg B	0.377	0.12	78	3.138	0.0126

Table A7T. Procrustes analysis Exp2A. Pairwise post-hoc comparisons on bias-continuation (interaction trial type x sequence type). Comparisons of interest are highlighted in bold.

A


Exp1V - Procrustes Bias
Synchronization
Continuation
bias
4
2
0
-2
reg A
dev B
reg B
dev A

B


Exp2A - Procrustes Bias
Synchronization
Continuation
bias
2
1
0
-1
-2
reg A
dev B
reg B
dev A

C

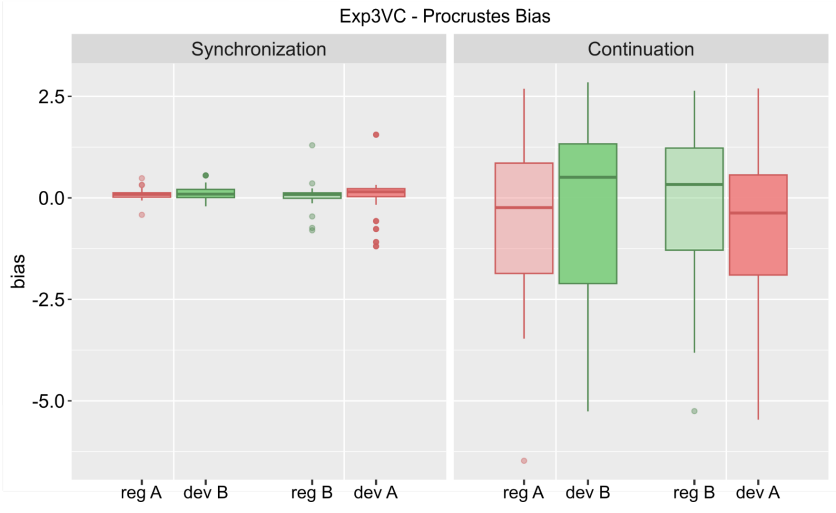

Exp3VC - Procrustes Bias
Synchronization
Continuation
bias
2.5
0.0
-2.5
-5.0
reg A
dev B
reg B
dev A

Figure A1. **Bias component results.** (A) Distribution of bias values for Exp1V grouped by task phase, trial type and sequence type (color-code: light red = global and local regular sequence A; green = global and local deviant sequence B; light green = global regular and local deviant sequence B; red = global deviant and local regular sequence A): we observed only a mild tendency to delay the tapping when the deviance is both at the local and global level and to anticipate the tapping when the deviance is only global. (B) Distribution of bias values for Exp2A grouped and color-coded as in (A): statistical analysis revealed a general local deviance effect expressed as a delay of the sequence AAAAB. (C) Distribution of scaling values for Exp3VC grouped and color-coded as in (A): also for complex sequences a statistically significant delay of the tapping was observed when performing the local deviant sequence CBAAB.

APPENDIX A – SUPPLEMENTARY MATERIAL CHAPTER 3

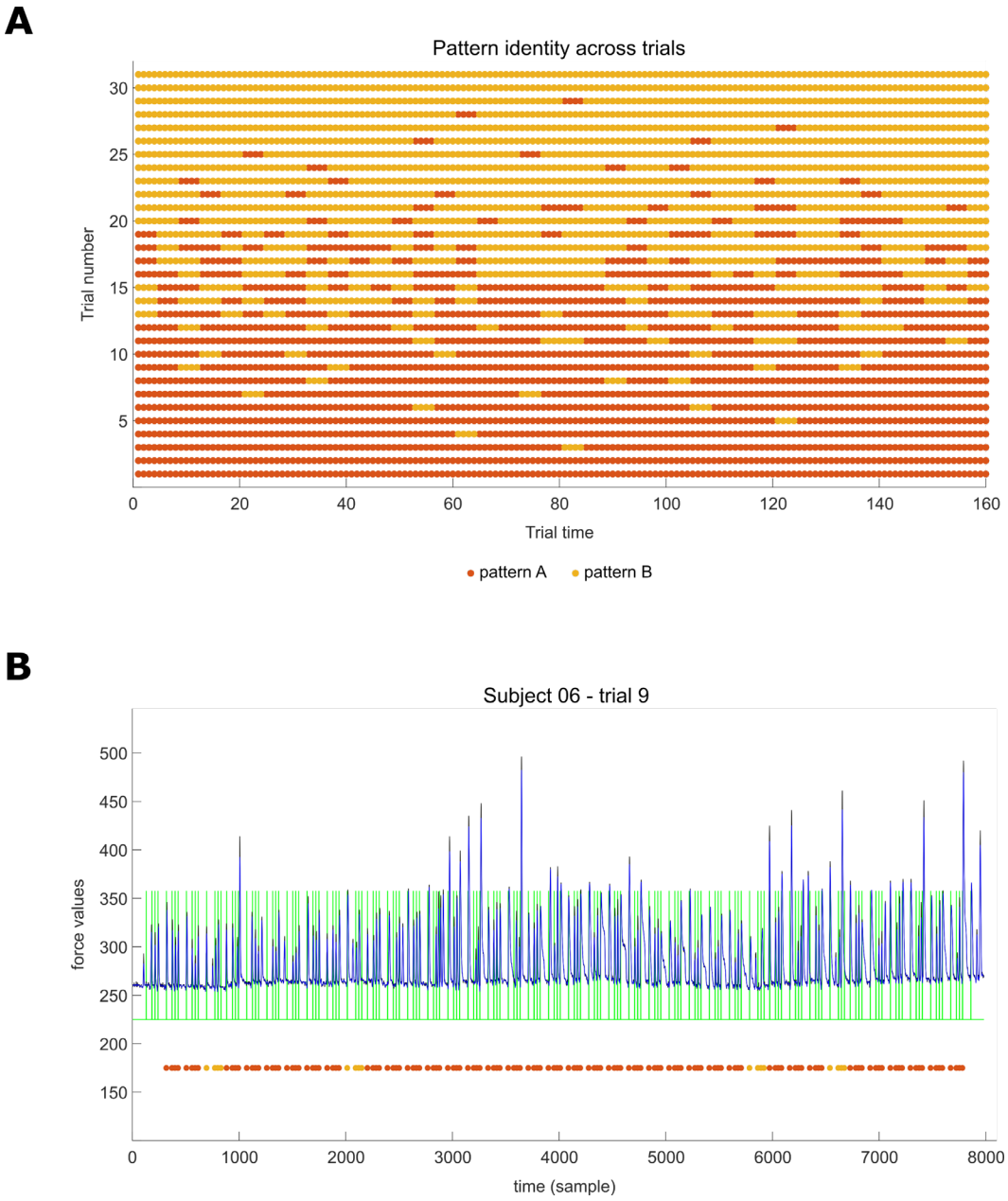


Figure B1. **Trial mirror randomization and tapping performance.** (A) Example of pattern A and B alternation across trials from one participant's mirror randomization. Each circle represents one interval of a pattern, color coded by pattern identity (pattern A = dark red, pattern B = yellow). Each trial in the second half (17th to 31st) mirrors the order of appearance of the patterns of the corresponding trial in the first half (1st-15th) that share the same probability distribution (trial 17 mirrors trial 15). (B) Example of tapping signal from one trial of participant 06. The green binary series represent stimuli onsets and offsets. The blue analogic signal is the raw tapping signal expressed in force values. The dotted line below the signals depicts the pattern series presented in this specific trial.

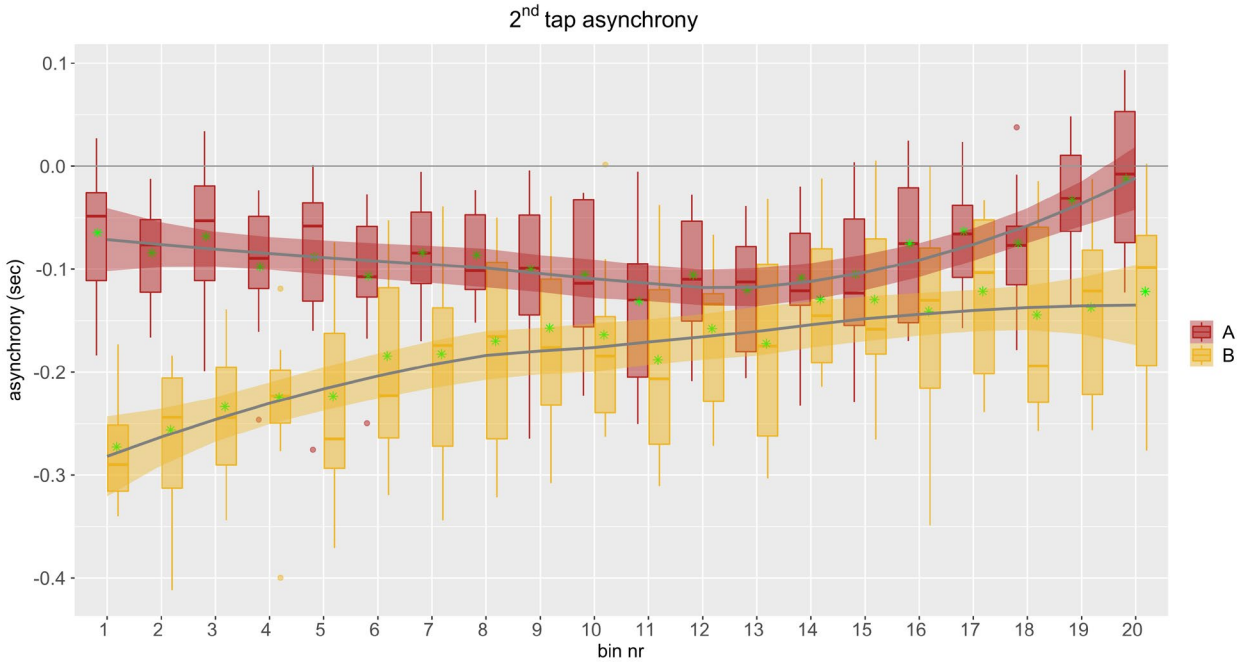


Figure B2. **Mean asynchrony of 2nd tap onset.** The plot shows the evolution across trials of the asynchrony of the 2nd tap of each pattern (color code: pattern A = dark red, pattern B = yellow). Pattern A asynchrony is constantly smaller than pattern B, whereas pattern B shows a history effect that translates into a steeper slope coefficient (here graphically represented using LOESS method).

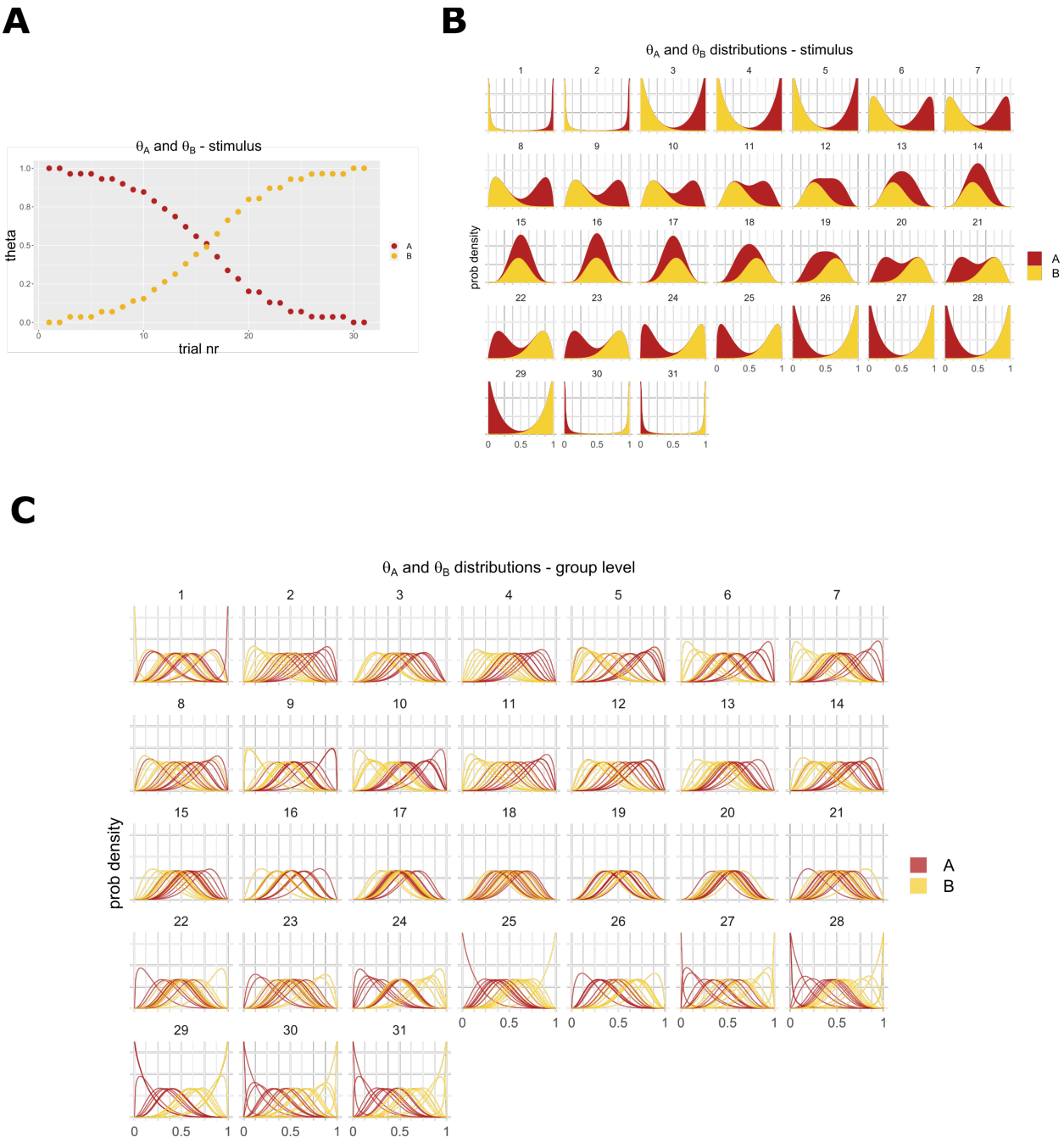


Figure B3. **Mixture model parameter θ estimation**. (A) Estimated θ_A and θ_B (color code: θ_A = dark red, θ_B = yellow) across trials for the stimuli series: this estimation was performed as sanity check to test whether the original proportions of pattern A and pattern B were accurately retrieved by the first model. (B) Beta priors on θ_A and θ_B across trials for stimuli series, as the result of the estimation of scale hyperparameters a and b for each θ, color-coded as in (A). (C) Beta priors on θ_A and θ_B across trials for all participants, as the result of the estimation of scale hyperparameters a and b for each θ, color-coded as in (A).

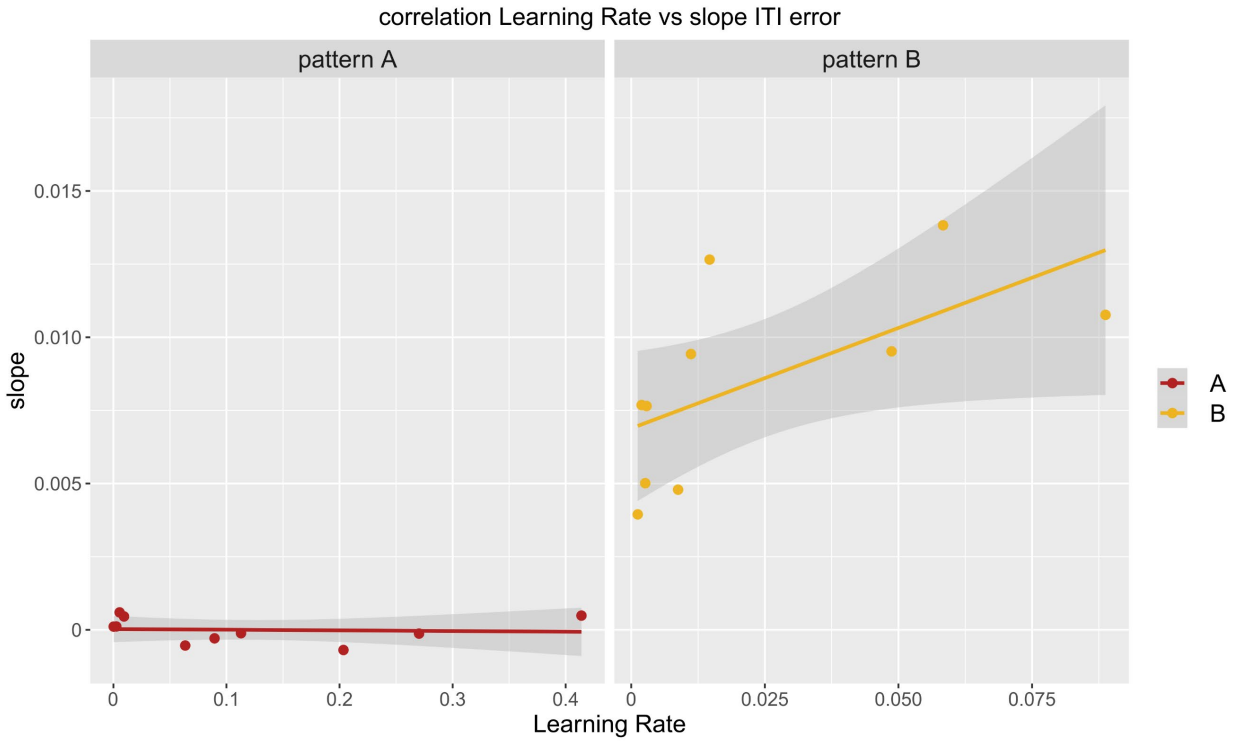


Figure B4. **Correlation between estimated learning rates and ITI error.** Each plot represents pattern A (dark red dots) and pattern B (yellow dots) estimated learning rate plotted against each pattern's slope coefficient calculated for ITI error distribution across trial history. Spearman's rank correlation revealed a significant association between the two values only for pattern B. As is evident from the plot, slope coefficient for pattern A were almost flat.

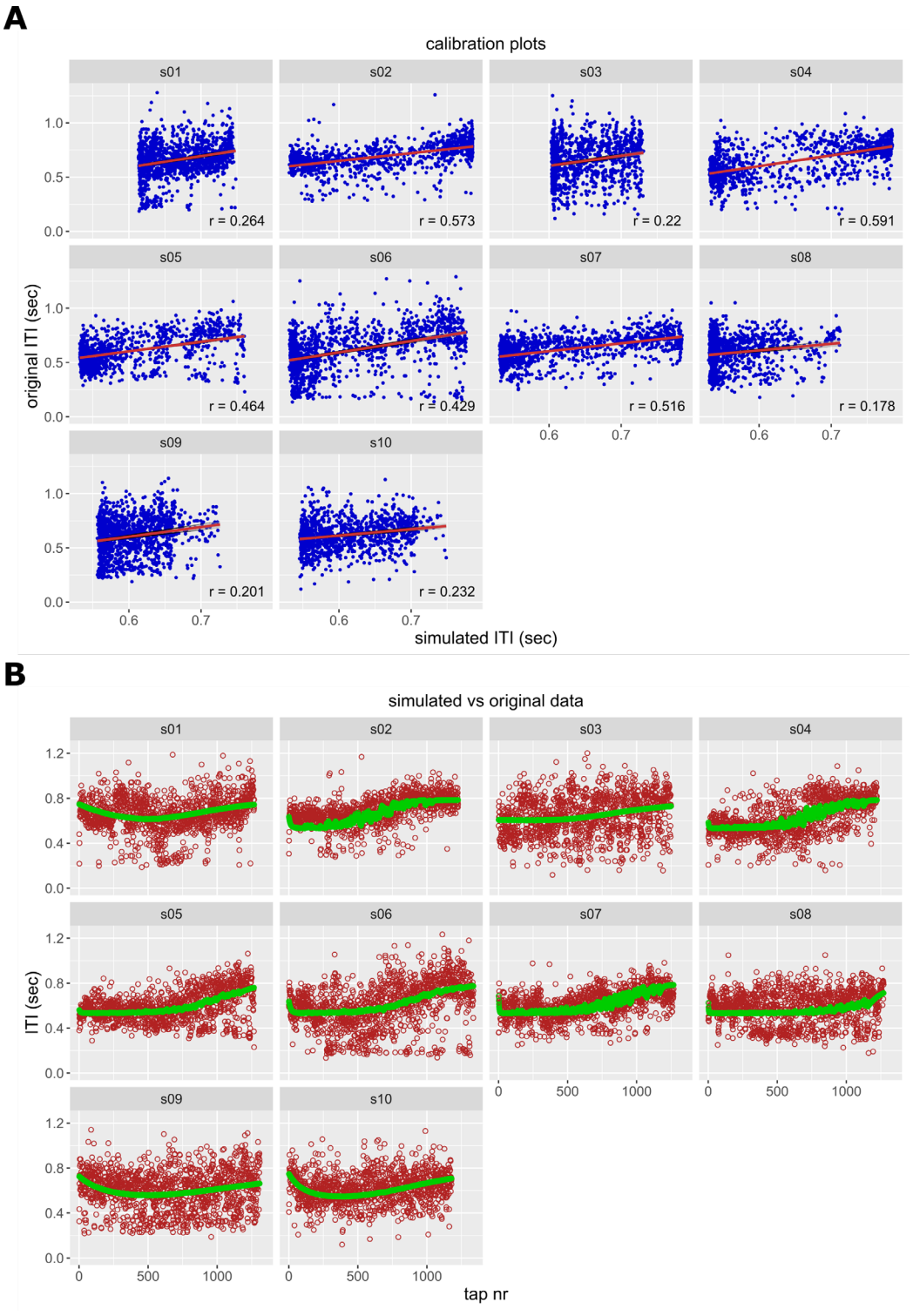
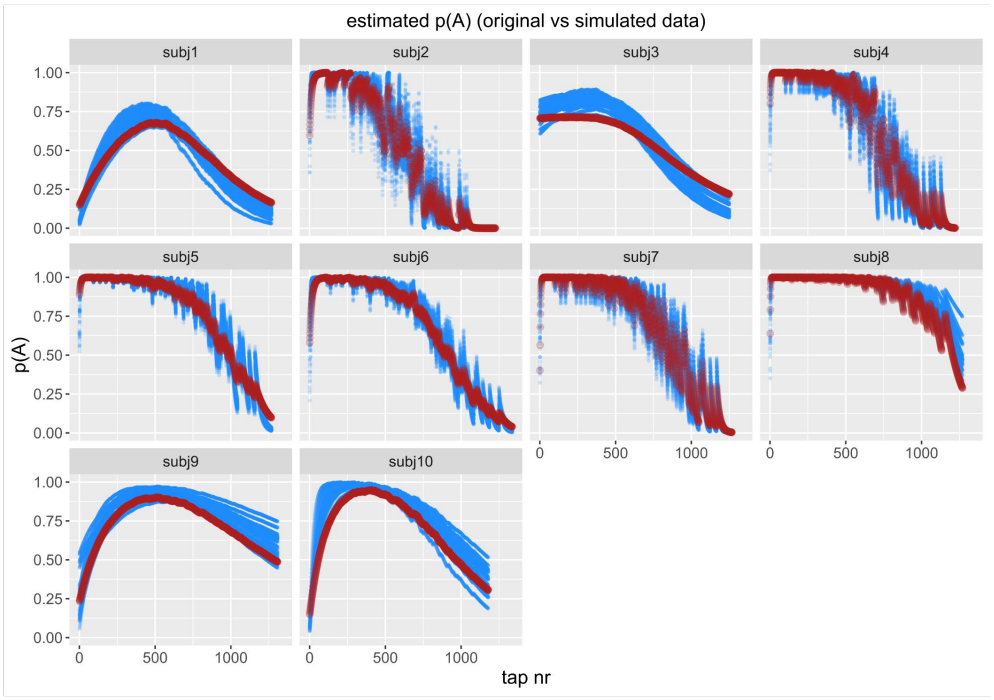
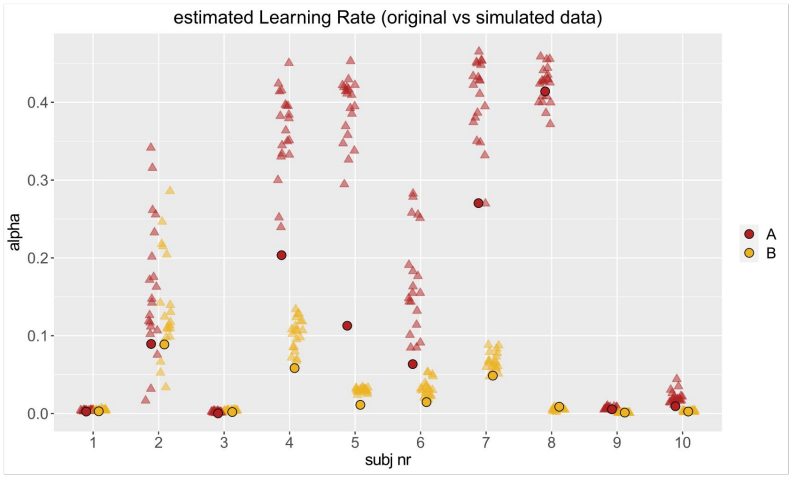


Figure B5. **Calibration plots and mean simulated data.** (A) The 20 synthetic dataset are plotted against the original data for each participant. A correlation coefficient between original and synthetic data was calculated for each participant and is indicated in the respective subplot. (B) Mean values of stimulated data (green dots) plotted together with the original dataset (red empty dots) of each participant.

A


estimated p(A) (original vs simulated data)
subj1
subj2
subj3
subj4
subj5
subj6
subj7
subj8
subj9
subj10
p(A)
1.00
0.75
0.50
0.25
0.00
0
500
1000
tap nr

B


estimated Learning Rate (original vs simulated data)
alpha
0.4
0.3
0.2
0.1
0.0
1
2
3
4
5
6
7
8
9
10
subj nr
A
B

C

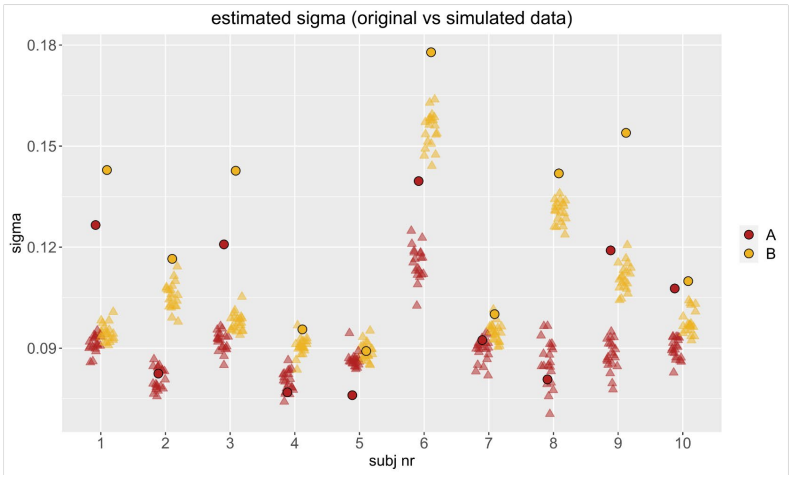

estimated sigma (original vs simulated data)
sigma
0.18
0.15
0.12
0.09
1
2
3
4
5
6
7
8
9
10
subj nr
A
B

Figure B6. **Parameter recovery results**. (A) Each plot shows the subject-wise estimation of parameter p(A) from simulated data (light blue dots) and from the original data (red dots), representing the probability of reproducing pattern A tap by tap (i.e. for pattern B {1 - p(A)}). Concerning this parameter, synthetic data led to approximately similar estimates to those obtained from original data. (B) The plot shows the estimation of learning rates α_A and α_B from simulated data (triangles) and from the original data (circles), plotted by subject (color code: α_A = dark red, α_B = yellow). (C) The plot shows the estimation of mixture component's standard deviations σ_A and σ_B from simulated data (triangles) and from the original data (circles), plotted by subject (color-coded as in (B)). Both for learning rates and sigma, parameter recovery led to very different estimates, that suggest potential limitation of the static learning rate model, namely the absence of a dynamic estimation of learning rates across trials and the great amount of variability in the data that this model is not yet able to capture.

APPENDIX C – SUPPLEMENTARY MATERIAL CHAPTER 3

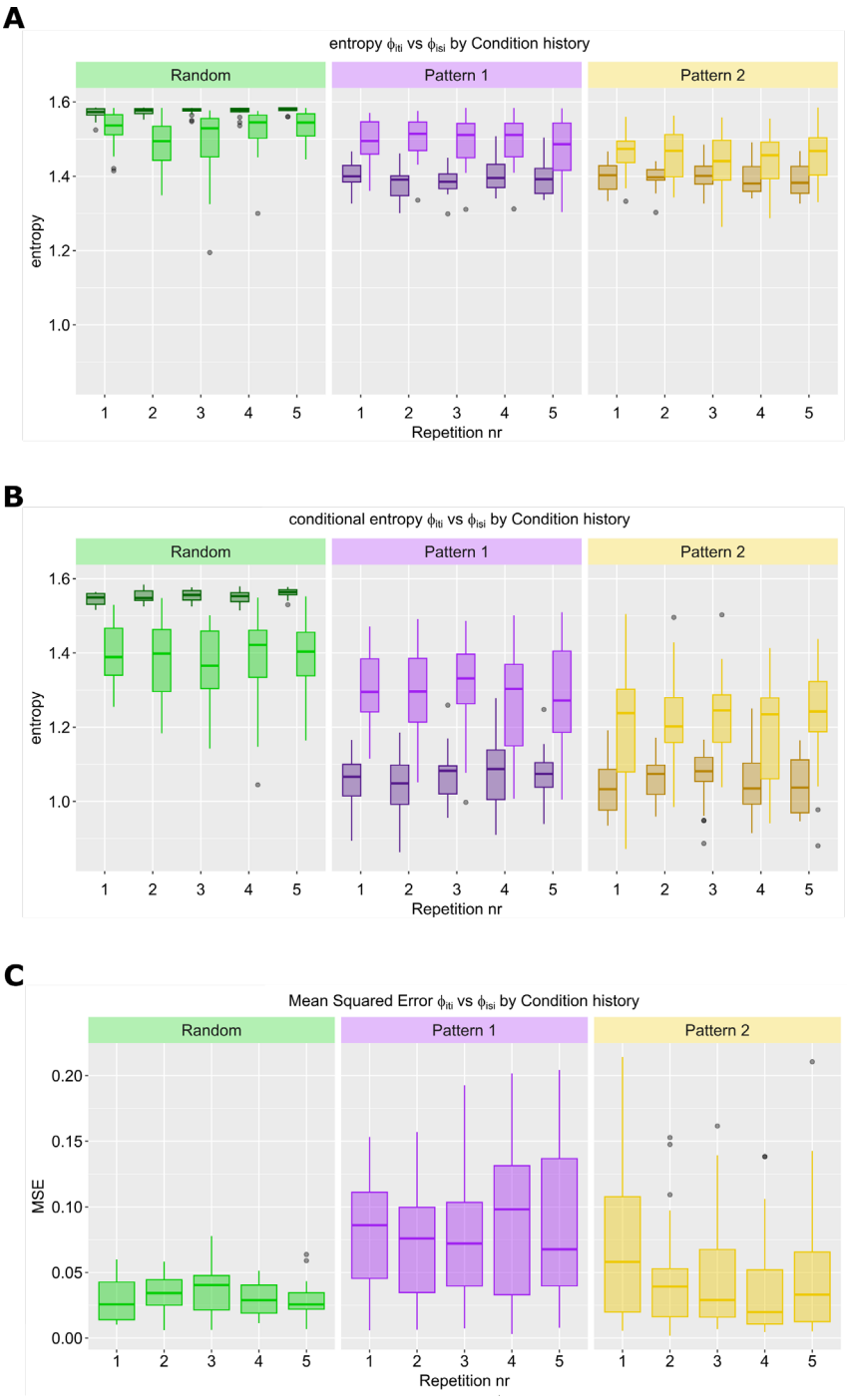


Figure C1. **Entropy rate, conditional entropy and Mean Squared Error for estimated parameter Φ_{ITI} across condition history.** Statistical comparison of entropy and MSE values showed no main effect of condition history, nor interaction with condition type. (A) Entropy rate values calculated from estimated parameters Φ_{ISI} and Φ_{ITI} across condition repetition and type (color coded as green for Random, purple for Pattern 1, yellow for Pattern 2 condition). Darker boxplot represents entropy values for Φ_{ISI}. (B) Conditional entropy values calculated from estimated parameters Φ_{ISI} and Φ_{ITI} across condition repetition and type (color coded as in A). Darker boxplot represents entropy values for Φ_{ISI}. (C) Mean Squared Error Φ_{ITI} - Φ_{ISI}, across condition repetition and type (color coded as in A). MSE = Mean Squared Error; ISI = Inter Stimulus Interval; ITI = Inter Tap Interval.

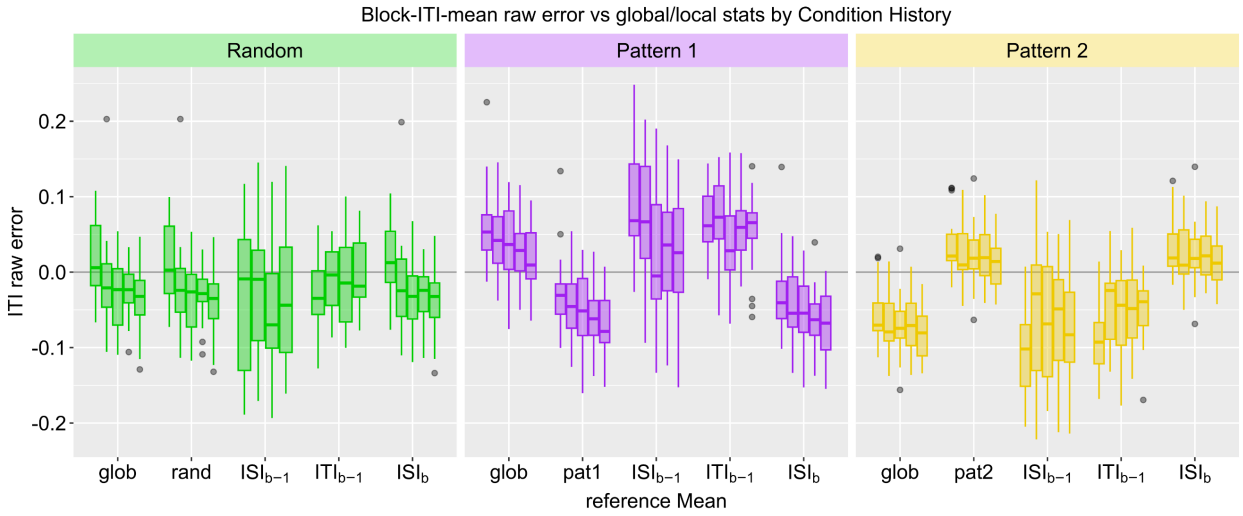


Figure C2. **Effect of condition history on block-ITI error.** Block-ITI error values plotted by error type (see Method section on ITI-error for details), condition history and grouped by experimental condition (color coded as green for Random, purple for Pattern 1, yellow for Pattern 2 condition). The plot highlights the condition history effect emerged in LME results for *glob-error* (post-hoc comparisons: rep1-rep2 = 0.01755, t.ratio(280) = 3.231, p=0.012; rep1-rep3 = 0.02604, t.ratio(280) = 4.795, p<0.0001; rep1-rep4 = 0.02937, t.ratio(280) = 5.409, p<0.0001; rep1-rep5 = 0.03836, t.ratio(280) = 7.064, p<0.0001; rep2-rep3 = 0.00849, t.ratio(280) = 1.564, p=0.5219; rep2-rep4 = 0.01183, t.ratio(280) = 2.178, p=0.1912; rep2-rep5 = 0.02081, t.ratio(280) = 3.832, p=0.0015; rep3-rep4 = 0.00333, t.ratio(280) = 0.614, p=0.9728; rep3-rep5 = 0.01232, t.ratio(280) = 2.269, p=0.1582; rep4-rep5 = 0.00899, t.ratio(280) = 1.655, p=0.4637), for *cond-error* (post-hoc comparisons: rep1-rep2 = 0.01755, t.ratio(280) = 3.235, p=0.0118; rep1-rep3 = 0.02604, t.ratio(280) = 4.800, p<0.0001; rep1-rep4 = 0.02937, t.ratio(280) = 5.415, p<0.0001; rep1-rep5 = 0.03836, t.ratio(280) = 7.071, p<0.0001; rep2-rep3 = 0.00849, t.ratio(280) = 1.566, p=0.5208; rep2-rep4 = 0.01183, t.ratio(280) = 2.180, p=0.1903; rep2-rep5 = 0.02081, t.ratio(280) = 3.837, p=0.0014; rep3-rep4 = 0.00333, t.ratio(280) = 0.614, p=0.9727; rep3-rep5 = 0.01232, t.ratio(280) = 2.271, p=0.1574; rep4-rep5 = 0.00899, t.ratio(280) = 1.657, p=0.4626) and *ISI_b-error* (post-hoc comparisons: rep1-rep2 = 0.01913, t.ratio(280) = 3.475, p=0.0053; rep1-rep3 = 0.02489, t.ratio(280) = 4.521, p<0.001; rep1-rep4 = 0.03001, t.ratio(280) = 5.451, p<0.0001; rep1-rep5 = 0.03591, t.ratio(280) = 6.524, p<0.0001; rep2-rep3 = 0.00576, t.ratio(280) = 1.046, p=0.8338; rep2-rep4 = 0.01088, t.ratio(280) = 1.976, p=0.2806; rep2-rep5 = 0.01678, t.ratio(280) = 3.049, p=0.0211; rep3-rep4 = 0.00512, t.ratio(280) = 0.930, p=0.8851; rep3-rep5 = 0.01103, t.ratio(280) = 2.003, p=0.2671; rep4-rep5 = 0.00591, t.ratio(280) = 1.073, p=0.8202), as well as the absence of history effects for and *ISI_{b-1}-error* and *ITI_{b-1}-error*.

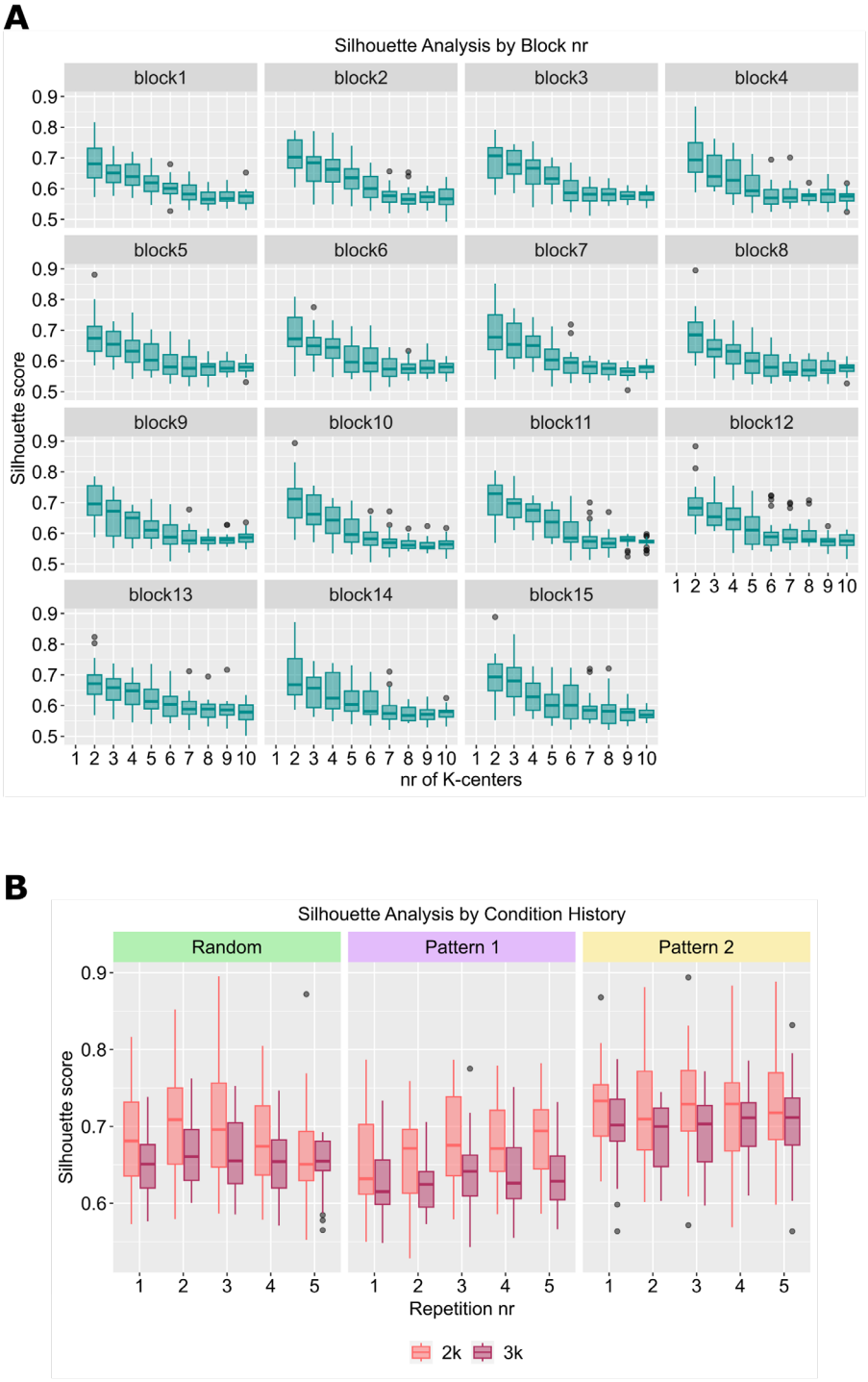


Figure C3. **K-means silhouette analysis on raw ITI**. (A) Silhouette scores are plotted for each K iteration, faceted by block number. Results of the analysis showed that the highest silhouette scores are consistently obtained for K=2 and K=3 across blocks, with a slow decrease in the score as K increases until a plateau around K=6. (B) Silhouette scores plotted for K=2 (light red bars) and K=3 (dark red bars) across experimental conditions and condition history (color coded as green for Random, purple for Pattern 1, yellow for Pattern 2 condition).

References

Andreou, L. V., Griffiths, T. D., & Chait, M. (2015). Sensitivity to the temporal structure of rapid sound sequences - an MEG study. *Neuroimage*, 110, 194-204.

Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, 113(5), E616-E625.

Bartoń, K. (2020). MuMIn: Multi-Model Inference. R package version 1.43.17. https://CRAN.R-project.org/package=MuMIn

Bartolo, R., Prado, L., & Merchant, H. (2014). Information processing in the primate basal ganglia during sensory-guided and internally driven rhythmic tapping. *Journal of Neuroscience*, 34(11), 3910-3923.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1-48.

Baum, L. E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, 41, 1554–1563.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences*, 106(5), 1672-1677.

Bouwer, F. L., Honing, H., & Slagter, H. A. (2020). Beat-based and memory-based temporal expectations in rhythm: similar perceptual effects, different underlying mechanisms. *Journal of Cognitive Neuroscience*, 32(7), 1221-1241.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10:433-436.

Breska, A., & Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS biology*, 15(2), e2001665.

Breska, A., & Ivry, R. B. (2018). Double dissociation of single-interval and rhythmic temporal prediction in cerebellar degeneration and Parkinson's disease. *Proceedings of the National Academy of Sciences*, 115(48), 12283-12288.

Cadena-Valencia, J., Garcia-Garibay, O., Merchant, H., Jazayeri, M., & de Lafuente, V. (2018). Entrainment and maintenance of an internal metronome in supplementary motor area. *Elife*, 7, e38983.

Cannon, J. (2021). Expectancy-based rhythmic entrainment as continuous Bayesian inference. *PLoS Computational Biology*, 17(6), e1009025.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1).

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of cognitive neuroscience*, 20(2), 226-239.

Chen, Y., Repp, B. H., & Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human movement science*, 21(4), 515-532.

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience*, 32(3), 1056-1060.

Coull, J. T., & Nobre, A. C. (2008). Dissociating explicit timing from temporal expectation with fMRI. *Current opinion in neurobiology*, 18(2), 137-144.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, memory, and cognition*, 30(5), 1119.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*, 88(1), 2-19.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, 32(3), 341-365.

El Karoui, I., King, J. R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., Adam, C., Navarro, V., Baulac, M., Dehaene, S., Cohen, L., & Naccache, L. (2015). Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: an intracranial study in humans. *Cerebral cortex*, 25(11), 4203-4212.

Elliott, M. T., Wing, A. M., & Welchman, A. E. (2014). Moving in time: Bayesian causal inference explains movement coordination to auditory beats. *Proceedings of the Royal Society B: Biological Sciences*, 281(1786), 20140751.

Essens, P. J., & Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & psychophysics*, *37*(1), 1-7.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in cognitive sciences*, 19(3), 117-125.

Gámez, J., Yc, K., Ayala, Y. A., Dotov, D., Prado, L., & Merchant, H. (2018). Predictive rhythmic tapping to isochronous and tempo changing metronomes in the nonhuman primate. *Annals of the New York Academy of Sciences*, 1423(1), 396-414.

Gámez, J., Mendoza, G., Prado, L., Betancourt, A., & Merchant, H. (2019). The amplitude in periodic neural state trajectories underlies the tempo of rhythmic tapping. *PLoS biology*, 17(4), e3000054.

García-Garibay, O., Cadena-Valencia, J., Merchant, H., & De Lafuente, V. (2016). Monkeys share the human ability to internally maintain a temporal rhythm. *Frontiers in psychology*, 7, 1971.

Gavornik, J. P., & Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature neuroscience*, 17(5), 732-737.

Geiser, E., Ziegler, E., Jancke, L., & Meyer, M. (2009). Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex*, 45(1), 93-102.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457-472.

Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception & psychophysics*, 18(1), 1-8.

Gower, J. C., & Dijksterhuis, G. B. (2004). Procrustes problems (Vol. 30). OUP Oxford.

Grahn, J. A. (2012). Neural mechanisms of rhythm perception: current findings and future perspectives. *Topics in cognitive science*, 4(4), 585-606.

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of cognitive neuroscience*, *19*(5), 893-906.

Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex*, 45(1), 54-61.

Grahn, J. A., & Rowe, J. B. (2013). Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cerebral cortex*, 23(4), 913-921.

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, 72(3), 561-582.

Grube, M., Cooper, F. E., Chinnery, P. F., & Griffiths, T. D. (2010a). Dissociation of duration-based and beat-based auditory timing in cerebellar degeneration. *Proceedings of the National Academy of Sciences*, 107(25), 11597-11601.

Grube, M., Lee, K. H., Griffiths, T. D., Barker, A. T., & Woodruff, P. W. (2010b). Transcranial magnetic theta-burst stimulation of the human cerebellum distinguishes absolute, duration-based from relative, beat-based perception of subsecond time intervals. *Frontiers in psychology*, 1, 171.

Hardy, N. F., & Buonomano, D. V. (2016). Neurocomputational models of interval and pattern timing. *Current Opinion in Behavioral Sciences*, 8, 250-257.

Hardy, N. F., Goudar, V., Romero-Sosa, J. L., & Buonomano, D. V. (2018). A model of temporal scaling correctly predicts that motor timing improves with speed. *Nature Communications*, 9(1), 1-14.

Heideman, S. G., van Ede, F., & Nobre, A. C. (2018). Temporal alignment of anticipatory motor cortical beta lateralisation in hidden visual-motor sequences. *European Journal of Neuroscience*, 48(8), 2684-2695.

Heilbron, M., & Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex?. *Neuroscience*, 389, 54-73.

Henry, M. J., Cook, P. F., de Reus, K., Nityananda, V., Rouse, A. A., & Kotz, S. A. (2021). An ecological approach to measuring synchronization abilities across the animal kingdom. *Philosophical Transactions of the Royal Society B*, 376(1835), 20200336.

Honing, H. (2013). Structure and interpretation of rhythm in music. In D. Deutsch (Ed.), *The psychology of music* (pp. 369-404). Elsevier Academic Press.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50(3), 346-363.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130(4), 658–680.

Iversen, J. R., & Balasubramaniam, R. (2016). Synchronization and temporal processing. *Current Opinion in Behavioral Sciences*, 8, 175-180.

Ivry, R. B., & Hazeltine, R. E. (1995). Perception and production of temporal intervals across a range of durations: evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1), 3.

Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, 27(3), 359-370.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature neuroscience*, 13(8), 1020-1026.

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: encoding time in neural network states. *Neuron*, 53(3), 427-438.

Kay, S. M. (1988). Modern Spectral Estimation. Englewood Cliffs, NJ: Prentice-Hall, 1988.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83(2), B35-B42.

Kleiner, M.H., Brainard, D., Pelli, D. (2007). "What's new in Psychtoolbox-3?" *Perception*, 36 ECVP Abstract Supplement.

Koelsch, S., Busch, T., Jentschke, S., & Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Scientific reports*, 6(1), 1-11.

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in cognitive sciences*, 23(1), 63-77.

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of rhythm processing. *Trends in cognitive sciences*, 22(10), 896-910.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, 82, 1-26.

Laje, R., Cheng, K., & Buonomano, D. V. (2011). Learning of temporal motor patterns: an analysis of continuous versus reset timing. *Frontiers in Integrative Neuroscience*, 5, 61.

Large, E. W., Fink, P., & Kelso, S. J. (2002). Tracking simple and complex sequences. *Psychological research*, 66(1), 3-17.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological review*, 106(1), 119.

Lenth, R. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.0. https://CRAN.R-project.org/package=emmeans

Leow, L. A., & Grahn, J. A. (2014). Neural mechanisms of rhythm perception: present findings and future directions. In: Merchant, H., de Lafuente, V. (Eds.) *Neurobiology of Interval Timing*. Advances in Experimental Medicine and Biology, vol 829 (pp. 325-338). Springer, New York, NY.

Lewis, P. A., Wing, A. M., Pope, P. A., Praamstra, P., & Miall, R. C. (2004). Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. *Neuropsychologia*, 42(10), 1301-1312.

London, J. (2004). Hearing in time: Psychological aspects of musical meter. Oxford University Press.

Lumaca, M., Trusbak Haumann, N., Brattico, E., Grube, M., & Vuust, P. (2019). Weighting of neural prediction error by rhythmic complexity: A predictive coding account using mismatch negativity. *European Journal of Neuroscience*, 49(12), 1597-1609.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2019). cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0.

Maheu, M., Dehaene, S., & Meyniel, F. (2019). Brain signatures of a multiscale process of sequence learning in humans. *elife*, 8, e41541.

Maheu, M., Meyniel, F., & Dehaene, S. (2022). Rational arbitration between statistics and rules in human sequence processing. *Nature Human Behaviour*, 1-17.

MATLAB (2018) 9.7.0.1471314 (R2019b). Natick, Massachusetts: The MathWorks Inc.

McAuley, J. D. (2010). Tempo and rhythm. In *Music perception* (pp. 165-199). Springer, New York, NY.

McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: a comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception and Performance*, 29(6), 1102.

Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., & Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1664), 20140093.

Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in neuroscience*, 274.

Merchant, H., Zarco, W., Pérez, O., Prado, L., & Bartolo, R. (2011). Measuring time with different neural chronometers during a synchronization-continuation task. *Proceedings of the National Academy of Sciences*, 108(49), 19784-19789.

Meyniel, F., Maheu, M., & Dehaene, S. (2016). Human inferences about sequences: A minimal transition probability model. *PLoS computational biology*, 12(12), e1005260.

Morillon, B., Schroeder, C. E., Wyart, V., & Arnal, L. H. (2016). Temporal prediction in lieu of periodic stimulation. *Journal of Neuroscience*, 36(8), 2342-2347.

Nobre, A. C., & Van Ede, F. (2018). Anticipated moments: temporal structure in attention. *Nature Reviews Neuroscience*, 19(1), 34-48.

Notter, M. P., Hanke, M., Murray, M. M., & Geiser, E. (2019). Encoding of auditory temporal gestalt in the human brain. *Cerebral Cortex*, 29(2), 475-484.

Paton, J. J., & Buonomano, D. V. (2018). The neural basis of timing: distributed mechanisms for diverse functions. *Neuron*, 98(4), 687-705.

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage*, 50(1), 302-313.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10:437-442.

Pereda, E., Quiroga, R. Q., & Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Progress in neurobiology*, 77(1-2), 1-37.

Pollok, B., Krause, V., Butz, M., & Schnitzler, A. (2009). Modality specific functional interaction in sensorimotor synchronization. *Human brain mapping*, 30(6), 1783-1790.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music perception*, *2*(4), 411-440.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ravignani, A., Delgado, T., & Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour*, 1(1), 1-7.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic bulletin & review*, 12(6), 969-992.

Repp, B. H., & Keller, P. E. (2004). Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *The Quarterly Journal of Experimental Psychology Section A*, 57(3), 499-521.

Repp, B. H., & Keller, P. E. (2008). Sensorimotor synchronization with adaptively timed sequences. *Human movement science*, 27(3), 423-456.

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.

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic bulletin & review*, 20(3), 403-452.

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in cognitive sciences*, 22(10), 870-882.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53-65.

Sadakata, M., Desain, P., & Honing, H. (2006). The Bayesian way to relate rhythm perception and production. *Music Perception*, 23(3), 269-288.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, *26*(1), 43-49.

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, 112(29), 8987-8992.

Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8), 1627-1639.

Schulze, H. H., Cordes, A., & Vorberg, D. (2005). Keeping synchrony while tempo changes: Accelerando and ritardando. *Music Perception*, 22(3), 461-477.

Stan Development Team (2020). RStan: the R interface to Stan. R package version 2.21.2. http://mc-stan.org/

Stan Development Team (2021). Stan Modeling Language Users Guide and Reference Manual, version 2.26.1. https://mc-stan.org

Strauss, M., Sitt, J. D., King, J. R., Elbaz, M., Azizi, L., Buiatti, M., Naccache, L., Van Wassenhove, V., & Dehaene, S. (2015). Disruption of hierarchical predictive coding during sleep. *Proceedings of the National Academy of Sciences*, 112(11), E1353-E1362.

Teki, S., Grube, M., & Griffiths, T. D. (2012). A unified model of time perception accounts for duration-based and beat-based timing mechanisms. *Frontiers in integrative neuroscience*, 5, 90.

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *Journal of Neuroscience*, 31(10), 3805-3812.

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music–brain responses to rhythmic incongruity. *cortex*, 45(1), 80-92.

Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Østergaard, L. (2006). It don't mean a thing…: Keeping the rhythm during polyrhythmic tension, activates language areas (BA47). *Neuroimage*, 31(2), 832-841.

Vuust, P., & Witek, M. A. (2014). Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Frontiers in psychology*, 5, 1111.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, 108(51), 20754-20759.

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, 14(1), 5-12.

Zarco, W., Merchant, H., Prado, L., & Mendez, J. C. (2009). Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *Journal of neurophysiology*, 102(6), 3191-3202.