An Investigation into the Perception and Production of Slow Rhythms

Rasmus Bååth

DOCTORAL DISSERTATION
by due permission of the Faculty of Humanities, Lund University, Sweden.
To be defended at Lux, room C126, Lund, October 9th at 13:15

Faculty opponent

Dirk Moelants

Lund University Cognitive Science

Department of Philosophy

LUND UNIVERSITY

Document name

Doctoral dissertation

Date of issue

Author(s): Rasmus Bååth

Sponsoring organization

Title and subtitle: An Investigation into the Perception and Production of Slow Rhythms

Abstract:

Appreciation and production of musical rhythm is a human universal and, as with other human capacities, it is imperative to understand the extent of our capacity to perceive and produce rhythm. This thesis presents my work on the cognitive and perceptual aspects of rhythm perception and production at a slow tempi.

In Paper I I establish that rhythm production, such as keeping the beat to a metronome sequence, gets subjectively more difficult at slower tempi. Difficulty increases gradually with slower tempo, however, there was a was marked increase in rated difficulty when there was more than 1800 ms between each metronome sound, supporting Repp's (2006) notion of at what tempo keeping a rhythm becomes difficult.

Paper II developed a computational cognitive model of rhythm categorization. The model used the resonance theory framework by Large (2010) to model behavioral data on how musicians categorize musical rhythm. The categorization made by the computational model and the categorizations made by the musicians agreed well, supporting the notion that resonance theory is a viable model of rhythm perception.

Paper III replicated the study by Bolton (1894) on the auditory illusion *subjective rhythmization*. Paper IV further explored aspects of this illusion and tested two theoretical explanations of why this illusion occurs. The results strongly favored the resonance theory explanation of subjective rhythmization. In connection to rhythm perception at slow tempi, the paper developed an argument for how participants' experience of subjective rhythmization relates to their slower limit of rhythm perception.

In Paper V I show that conventional methods for measuring timing performance do not work correctly when applied to data from rhythmic timing task performed at tempi slower than 30 BPM. A solution to this problem is presented in the form of a problem specific Bayesian model, which was subsequently used to calculate timing variability in Papers VI and VII.

Paper VI examine the relationship between auditory working memory, sensorimotor synchronization performance, and memory capacity for rhythms. The results showed that auditory working memory and memory capacity for rhythms are related. However, the influence of memory capacity on synchronization performance showed no interaction with sequence tempo, suggesting that auditory memory does not play an integral role in rhythm perception.

Paper VII showed that, when the tempo is sufficiently slow, performing rhythmic timing demands attentional resources and involvement of executive control. This result resonates with neural models of timing that suggest a dedicated timing mechanism for short intervals and a general, cognitive timing mechanism for longer intervals.

Key words: rhythm perception, time perception, finger tapping, slower limit of rhythm perception

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language: English

ISSN and key title: 1101-8453, Lund University Cognitive Studies, 163

ISBN: 978-91-87833-43-4

Recipient's notes

Number of pages

Price

Security classification

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature ______________________________________Date ___________

An Investigation into the Perception and Production of Slow Rhythms

Rasmus Bååth

LUND
UNIVERSITY

Back cover photo by Åsa Harvard

Cognitive Science
Department of Philosophy
Faculties of Humanities and Theology

ISBN 978-91-87833-43-4
ISSN 1101-8453
Lund University Cognitive Studies 163

Printed in Sweden by Media-Tryck, Lund University
Lund 2015

Acknowledgements

Great thanks, good karma and a big hug to all that helped me during my toils. I owe you more beers than I can buy, and you can drink. In no particular order I wish to thank:

- The research environment *Thinking in Time: Cognition, Communication and Learning* for sponsoring my PHD-studies and for giving me the opportunity to hang out with this great bunch of (now mostly former) PHD-students: Anders Rasmussen, Fredrik Johansson, Olof Sandgren, Karolina Löwgren, Jens Nirme, Emelie Grenner, Stefan Winberg, Åsa Harvard, Andreas Falck, Roger Johansson, Kristina Borgström and Simone Löhndorf.

- Christian Balkenius for supervising my PHD, for inspiring me to study cognitive science in the first place, and for always being so positive and supportive.

- My secondary supervisors Geoffrey Patching, Guy Madison, and Peter Gärdenfors for good discussions and for pushing me in the right directions. I owe an especially large beer to Geoffrey Patching for picking me up from my midterm crisis and to make me actually *publish* something. Without you there wouldn't have been a thesis.

- Erik Lagerstedt for being my MATLAB buddy and for patiently listening to my various rants on why science is broken.

- Kristin Osk Ingvarsdóttir for being a good buddy and for showing me South Korea. Gamsahabnida!

- Tomas Persson for being another good buddy and for lending me that SNES game. (I haven't forgot!)

- Birger Johansson for being my academic big brother and for keeping me company in the robot lab in Kungshuset.
- All the great people at LUCS. And just to mention a few: Agneta Gulz, Björn Sjödén, Paulina Lindström, Magnus Haake, Richard Andersson, Magnus Johnsson, Mathias Osvath, Helena Osvath, Zahra Gharaee, Can Kabadayi, Ivo Jacobs, Elainie Madsen, Jana Holsanova, Betty Tärning, Annika Wallin, Kerstin Gidlöf, Richard Andersson, Thomas Strandberg, Lars Hall, Andreas Lind, Philip Pärnamets and Petter Johansson.
- Björn Petersson, Agneta Ahlberg, Anna Cagnan Enhörning, Anna Östberg and Eva Sjöstrand for just being such solid competent people.
- The Humanities Lab and Marianne Gullberg, for giving me a space to run my experiments.
- Emmanuel Genot and Justine Jacot for being awesome, and also great row models when it comes to parenting. I might ask you questions about this in the future.
- Joel Parthemore for proofreading large portions of this thesis. If you see a spelling error, you're reading the parts he didn't proofread.
- Ullrika Sahlin for sharing my passion for Bayesian statistics. One day the world will understand...
- My parents for never being anything else than supportive, whatever I seem to be doing.
- My friends, I hope you are still there, you were last time I checked.
- Maria for being my partner in crime (except that we never seem to be doing anything criminal), for waking me up with a smile (I hate waking up, but the smile makes up for it tenfold), for accompanying me on life's journey (which takes place mostly in Lund, we don't like to travel that much), for snuggling with me on the sofa while we watch dusk turn to night (and watch another episode of Arrested Development, we really should get out more!), and for being the love of my life (a sentiment that needs no further modification).

Contents

1. List of original papers

Paper I

Bååth, R. and Madison, G. (2012). The subjective difficulty of tapping to a slow beat. *In Proceedings of the 12th International Conference on Music Perception and Cognition*. Thessaloniki, Greece.

Paper II

Bååth, R., Lagerstedt, E., and Gärdenfors, P. (2014). A prototype-based resonance model of rhythm categorization. *i-Perception* 5(6) 548–558; doi:10.1068/i0665

Paper III

Bååth, R. and Ingvarsdóttir, K. O. (2014) Subjective rhythmization: A replication and an extension. *Proceedings of the 13th International Conference on Music Perception and Cognition.* Seoul, South Korea.

Paper IV

Bååth, R. (In press) Subjective rhythmization: A replication and an assessment of two theoretical explanations. *Music Perception.*

Paper V

Bååth, R. (2015). Estimating the distribution of sensorimotor synchronization data: A Bayesian hierarchical modeling approach. *Behavior Research Methods.* doi:10.3758/s13428-015-0591-2

Paper VI

Bååth, R. (Submitted) Working memory, memory for musical rhythms, and rhythm perception.

Paper VII

Bååth, R., Tjøstheim, T., Lingonblad M. (Submitted) The role of executive control in rhythmic timing at different tempi.

2. Introduction

Musical rhythm is a human universal (Brown, 1991), and while there are cultural differences with respect to tempo, meter and instrumentation, all cultures have musical rhythm in some form (Stevens, 2012). Already newborn infants have a sense of rhythm (Honing et al., 2009) and there is only one reported case of *beat deafness* (Phillips-Silver et al., 2011), which can be compared to the many reported cases of *tone deafness* Patel et al. (2008). Appreciation and production of musical rhythm is part of what it means to be human[1] and, as with other human traits, there has been a large research effort to understand: *Why*, in the ultimate sense, do we have a sense of rhythm? *How* does it work? And *what* can it do?

Why humans appreciate musical rhythm, and music in general, is much debated but little agreed upon (Patel, 2006; Pearce and Rohrmeier, 2012). One question is: If appreciation of musical rhythm gives such an evolutionary advantage why is it not found in many more species? Or as put by Fitch (2012):

> The paradox, put simply, is this: if periodicity and entrainment are ubiquitous features of all living organisms, why can't dogs dance?

It might be the case that musical rhythm and rhythmic entrainment does not give an evolutionary advantage, but then the question becomes: Why have humans evolved the capacity for music and rhythm? One

[1]This is not to say that rhythm is not part of what it means to be a cockatoo, one other species that has shown a sense of musical rhythm (Patel et al., 2009). Many other bird species also react to musical rhythms (Schachner et al., 2009) and sea lions have been trained to move to beat of songs (Cook et al., 2013). Curiously, primates do not show a strong sense of rhythm (Merchant and Honing, 2013; Honing et al., 2012). Even though there are published reports of synchronization behavior in chimpanzee (Yu and Tomonaga, 2015), the evidence for that primates have a sense of rhythm is far from as compelling as for cockatoos.

explanation is that music and rhythm are side effects of the human capacity for language, with music being supernormal stimuli or, as Pinker (1997) puts it, "auditory cheesecake". This theory has been criticized (Levitin and Tirovolas, 2009) and other proposals are that musical ability is the result of sexual selection (Miller, 2000) or that the evolutionary benefit of musical rhythm is to help synchronizing motor activity on the group level (Merker et al., 2009).

How humans can perceive and produce musical rhythm is another area of research focused on cognitive modeling and on establishing plausible neurological mechanisms. Models can range from statistical models of specific tasks, such as keeping the beat to a metronome, to models aiming at describing mechanism general to the processing of musical rhythm (Grondin, 2010). There is also a large literature on the neural correlates of listening to and producing musical rhythms (see Grahn, 2012, for a review).

In order to understand why a capacity exists and how it works, it can be useful to investigate *what* that capacity can do and how it is limited. A well known example of a study on limitations is Miller's (1956) "The magical number seven, plus or minus two", which describes the limitations of short-term memory and, by doing this, constrained the possible answers to how short-term memory could work. With respect to musical rhythm, two possible questions are: What type of responses are made to rhythmic stimuli and what type of sounds are perceived as rhythmical? Asking these type of questions has a long history in experimental psychology where early examples of studies focusing on rhythm production are Stevens (1886), Miyake (1902), Dunlap (1910), and Woodrow (1932). These early examples all used versions of a *sensorimotor synchronization finger tapping task*, which implies that motor responses, here finger taps, are synchronized to sensory stimuli, often metronome sequences. The focus in these and similar papers are often on *performance*: how well participants can perceive and reproduce a rhythm under different conditions, exploring under what circumstances the human sense of rhythm works, and under what circumstances it deteriorates. There is also a long tradition of introspective studies that investigates how different rhythms are experienced (e.g., Bolton, 1894; MacDougall, 1903).

Here it might be suitable to introduce the term *rhythm perception* as an umbrella term for a number of capacities related to musical rhythm. Rhythm perception can be seen as a subcategory of time perception and refers to capacities such as the perception of tempo, meter, intervals and rhythmic phrases. This thesis does not touch on the *why* of rhythm perception, but mainly focuses on the *what* and, to a lesser degree, on the *how*. Rhythm perception is a vast subject and what questions I have chosen to explore has been guided by where I have found the literature wanting[2].

The main research question of this thesis is: What happens with rhythm perception when the tempo is slow? What could be considered a *slow* tempo depends, of course, on the context and is relative to what would be considered a conventional tempo. As most contemporary music has a tempo faster than 60 beats per minute (van Noorden and Moelants, 1999), a rhythm with more than a second between each beat would be considered slow in most contexts.

A secondary research question has been: What is the slower limit of rhythm perception? This question, in the same vein as Miller (1956), is concerned with a perceptual or cognitive limit. If the slower limit of rhythm perception was known, it would put a constraint on what mechanisms rhythm perception might depend on and why, in the ultimate sense, humans have acquired a sense of rhythm.

Connected to these two questions is the question: What type of model can account for a slower limit of rhythm perception and participants' responses to slow rhythms? This question led me to explore the resonance theory for rhythm perception developed by Large (2008), which is a flexible framework that can explain many phenomena related to rhythm perception.

This thesis introduction continues with an overview of basic concepts related to musical rhythm in section 2.1. Section 2.2 reviews how rhythm perception can be measured and describes the construction of the tapping board that was used to record rhythmic finger tapping data in papers I, IV, V, VI and VII. A reason for collecting rhythmic finger tapping data is often to get a measure of timing performance

[2]This thesis is also very much rooted in the literature on rhythm perception and production, and in the literature on sensorimotor synchronization (I have read and reread the comprehensive reviews of Repp, 2005 and Repp and Su, 2013 numerous times). These directions of research are often not so much focused directly on studying rhythm in music. While the the ultimate goal might be to understand the psychology of music, the focus is more on studying human performance in more simplified rhythmic tasks, where the isochronous finger tapping task is one often employed experimental paradigm. Therefore, the papers in this thesis will make relatively few direct connections to music as such, even though I hope the results presented herein will add to the understanding of the psychology of music.

and section 2.3 introduces the *asynchrony standard deviation* as such a measure. Section 3 describes the contribution of this thesis and puts it into context. Finally, section 3.5 reviews a number of methods for probing a slower limit of rhythm perception. My contribution, with respect to a slower limit of rhythm perception, is summarized in figure 3.6.

2.1 Basic concepts of musical rhythm

Four basic concepts of musical rhythm are *rhythm*, *beat*, *meter* and *tempo* (McAuley, 2010).

Rhythm

In the field of music cognition rhythm refers to the temporal organization of sounds and silences. No cyclical or hierarchical structure is implied, which is different from how the term *rhythm* is used in other fields (cf. circadian rhythm). A distinction can be made between rhythm as the temporal pattern of sound and rhythm as the *perception* of a temporal pattern of sound. An important aspect of the latter is the *grouping* of the rhythm: how the series of sounds are perceived to be clustered together. The grouping of a rhythm is a complex phenomena that is affected not only by the temporal organization but also by the intensity, timbre, duration and tempo of the sound events (Handel, 1989).

Beat

In most types of music there is a perceived regularly occurring pulse, the *beat*. The beat of a piece of music is what one would synchronize hand claps or dance steps to. A point of confusion is that both a rhythm that establishes pulses and a single pulse can be referred to as a beat, therefore it is possible to speak of “the beats of a beat”. Sound events often occur on the perceived beat and are then experienced as accented. It is, however, not necessary that sound events occur on every beat; a rhythm can evoke a strong sense of a beat while still leaving out many sounds that would occur on the beat.

Tempo

The *tempo* of a piece of music is the rate at which the beats occur. Tempo is often given as the number of *beats per minute* (BPM). In experimental psychology another common[3] measure is the *interstimulus interval* (ISI) or *interonset interval* (IOI) which define the length of the interval between adjacent beats. For example, a tempo of 120 BPM corresponds to a tempo with an ISI of $1/120 \times 60 \times 1000 = 500$ ms. The perceived speed of a song is heavily influenced by the tempo, but this relation is not linear, and the perceived speed is also influenced by other aspects of the rhythm, such as the event density (Madison and Paulin, 2010).

Meter

In the same way as tones in a song can be accented, a beat can be accented, where some beats are given more emphasis and are perceived as being more stressed. The hierarchical emphasis pattern of a beat is called the *meter*. The base level of a meter is always the beat, above the beat is the bar (or measure), where a bar often consists of 2-4 beats and where the first beat in each bar is given more emphasis. Below the beat level is a subdivision of the beat where the first part of each subdivided beat is given more emphasis. An example of a meter is visualized in figure 2.1, both by musical notation (A) and by the emphasis pattern (B). Meter not only dictates how music is performed but is also a perceptual phenomena. The hierarchical structure of a meter can be perceived even if all sound events in a sound sequence are identical. This phenomena is called *subjective rhythmization* or sometimes the *tick-tock effect* (Bolton, 1894; Bååth, In press).

2.2 Measuring rhythm perception

The different experimental tasks used to investigate rhythm perception can loosely be divided into *listening tasks* and *production tasks*. In listening tasks the participant is presented with rhythmic stimuli, for example, metronome sequences or musical excerpts and, after having listened to a stimulus, gives some kind of judgment. This type of task

[3]This thesis introduction contains a number of claims, which are given without any reference, that things are *common*. This should then be read as *common in my experience*.

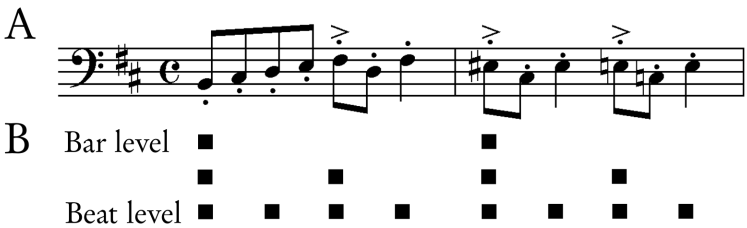


Figure 2.1: A visualization of the meter of the first two bars of Edvard Grieg's *I Dovregubbens hall.*

can often use experimental paradigms common in experimental psychology such as forced choice tasks or Likert type questionnaires (as in Geringer and Madsen, 1984 or Johnson, 1996). There are also many studies that have combined listening tasks with neuroimaging techniques like EEG and fMRI (for example, Nozaradan, 2014; Grahn and Rowe, 2013).

Production tasks instead investigate how participants produce rhythmic responses under various conditions. These type of tasks can investigate rhythm in directly musical contexts, such as the synchronization between musicians in a string quartet (Wing et al., 2014) or the amount of swing in jazz drummers as a function of tempo (Honing and De Haas, 2008). There is also a category of, what could be considered as, more artificial rhythm productions tasks which give up some ecological validity in order to focus in on specific aspects of rhythm perception[4].

One of the most common experimental paradigms when investigating rhythm production is the finger tapping task (Repp, 2005). This task was introduced more than a century ago (see Stevens, 1886, for an early example) and in its basic form a participant is asked to tap with his or her finger in synchrony with an isochronous (evenly spaced in time)

[4]That a task involves rhythm *production*, does not imply that it does not involve rhythm *perception*. Many rhythm production task are also rhythm perception tasks as what rhythm a participant produces depends on what rhythm the participant perceives. This is both the case for simple rhythm production tasks, like keeping the beat to a metronome using finger taps, and more elaborate rhythm production tasks, like synchronizing the tempo with which you play a tune to the tempo of the rest of the orchestra. In the same way an artist exercises color perception when painting, as the the artist has to relate to the colors in the emerging painting, a drummer can be said to exercises rhythm perception when playing, as the drummer has to perceive and relate to the rhythm of past sound events.

sequence of sounds. There are many variation of this basic task. For example, the amount of auditory feedback the participant is given can be varied, the participant can tap on a surface or freely flex the finger and the synchronization phase can be followed by a continuation phase where the sound sequence is muted while the participant continues tapping at the same tempo. This last modification is common when trying to dissociate *timing error* from *motor error* using, for example, the influential model of Wing and Kristofferson (1973).

Sensorimotor synchronization (SMS) is an umbrella term for rhythm production behavior that involves the synchronization of some movement to a predictable external event. Typical examples of SMS tasks are walking to the pace of a drum and tapping to the beat of a metronome, but also dancing to a piece of music or making music in an ensemble could be considered as SMS.

Reliability and latency when conducting a finger tapping study[5]

In order to conduct a finger tapping study one needs an apparatus to accurately play the sound sequence to the participant and to record the timing of the participants' taps. If one is interested in the relation between successive taps, the recorded timing of the taps need high *reliability*, that is, there should be a low amount of temporal jitter in the recorded timing. If one is interested in the relation between the taps and the sound onsets both the sound playback and the recorded timing needs high reliability and low *latency.* A sound needs to start immediately when playback is initiated and there should be no systematic discrepancy between the timing of a tap and the recorded timing. An example of the result of latency and jitter when recording timings in a sequence of taps is given in figure 2.2.

One apparatus that plays sounds and records key presses is a standard personal computer (PC). It would be convenient to use a PC, as they are readily available, but there are some issues that make it problematic to directly use a PC in a tapping study:

- There can be considerable temporal jitter and latency when playing sound through a PC. This of course depends on the brand and setup, but using a standard Windows PC with a consumer grade

[5] This section, and the following, is an abbreviated version of Bååth (2011) which describes the apparatus I constructed and used in papers I, IV, V, VI and VII.

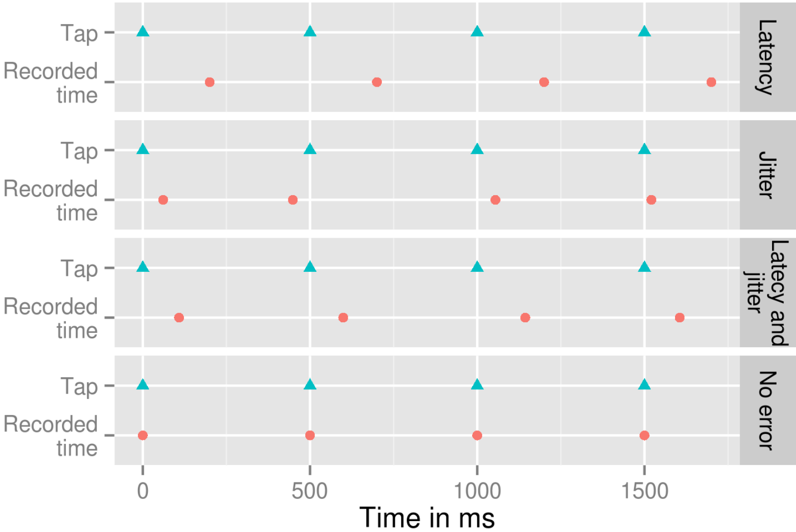


Figure 2.2: The result of latency and jitter when recording the timing of a sequence of finger taps.

sound card can result in audio delays ranging from 10 ms to 250 ms depending on the CPU load (MacMillan et al., 2001).

- There can also be temporal jitter and latency in the registration of key presses. I have not found a recent article that measures key press latency but Lane and Ashby (1987) estimate it to around 6 ms on a first generation Macintosh computer.

- It is hard to measure latency and jitter and to separate key press delay from sound delay. Wright et al. (2004) measured the key press-to-sound delay on computers running Linux and MacOS, and found delays ranging from 10 ms to 80 ms.

- Computer keyboard keys might not be ideal for tapping tasks. It is not enough to just *tap* a keyboard key, it has to be *pressed*, and most computer keyboard keys makes an audible "click" both when pressed and depressed.

One common way of getting around the problem of the tactile feel of keyboard keys, and to possibly decrease latency and jitter, is to use a MIDI interface for sound playback and registration of participants' taps. This approach is common in the literature (see, e.g., Repp and Doggett, 2007; Madison, 2001) but still suffers from the problem that it is hard to measure delays in sound playback and tap registration. One reason for why it is hard to know the delays in PCs and MIDI equipment is that these are complex, non transparent systems where there are many processes running simultaneously and where access to the hardware is hidden behind layers of abstractions.

The construction of an accurate tapping board to record finger taps

Another solution is to use a system that is simple, that is dedicated to the task of playing sounds and registering taps, and where it is possible to guarantee low upper bounds of the delays. Such a system is the *Arduino* which is a open-source electronics prototyping platform that includes, among other things, a 16 MHz processor, a USB port and several input and output pins (Mellis and Banzi, 2007). Using the Arduino remedies many of the problems with using a PC. A program implemented on the Arduino runs close to the hardware and there is no operating system that adds unpredictable delays. Because of this, when using an Arduino, it is possible to achieve millisecond accuracy when playing sounds and registering taps.

This section describes the construction of an Arduino based tapping board. The tapping board was designed to be comfortable to use and to register taps with millisecond accuracy. The on-board software was designed to support two types of tasks: A standard tapping task where a participant synchronizes his or her taps to an isochronous sequence of sounds and a spontaneous motor tempo task where there is no pacing signal and where the participant can tap at any tempo.

The apparatus consisted of an Arduino, a tapping board with an attached piezo element, a standard 3.5 mm stereo jack and a small breadboard that was used to connect the different components. The tapping board consisted of a wooden wrist rest and a 5 cm^2 tapping pad of corrugated fiberboard that rested on a piece of plastic foam of the kind commonly found in foam mattresses. This plastic foam also provided a place to rest for the fingers not involved in the tapping. Below the tapping pad was an attached piezo element that picked up vibrations from the tapping pad. Fiberboard was chosen because it was found to provide a hard surface while still having the elasticity to mediate the taps to the piezo element. For a picture of a prototype of

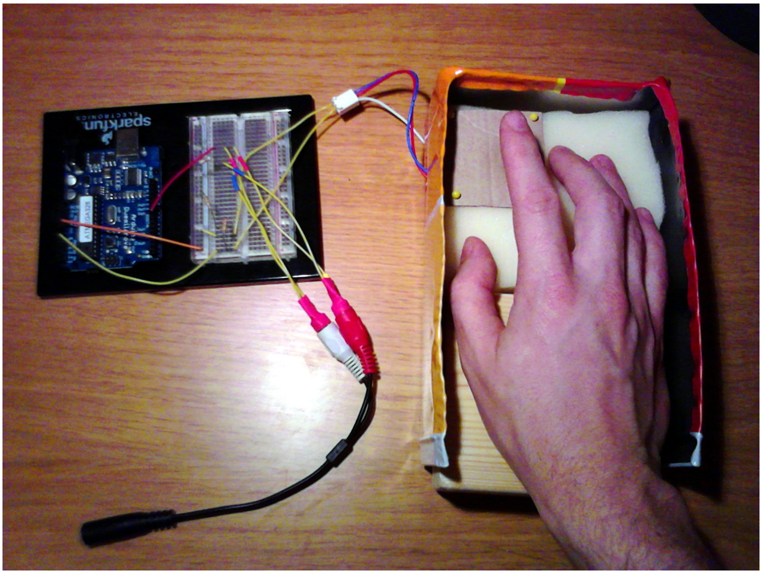


Figure 2.3: A prototype of the tapping board with all of the components exposed. The piezo element is hidden below the tapping pad.

the tapping board see figure 2.3.

The Arduino was programmed to handle two types of common rhythm production tasks: a standard SMS tapping task and a spontaneous motor tempo task (the source code can be found in the appendix of Bååth, 2011). Initiation of the tasks and handling of the resulting data is not done on the Arduino but has to be handled by a PC connected by the USB port. When the tapping task is initiated the Arduino plays a given number of square wave sounds with a given period and records the timing and amplitude of the taps made on the tapping board. The status of the piezo element is polled more than 10 times every ms. Each sound is associated with one tap and that tap is time stamped at the time of the peak amplitude reading in the time interval centered on the sound onset with a width of the ISI of the sequence. This method of defining a tap is a robust way of handling noise coming from the piezo element.

In the spontaneous motor tempo task a given number of taps is recorded without there being any pacing sequence. Here a tap is counted as every reading with an amplitude higher than a given threshold. This method of defining a tap is less robust and relies on that the threshold is carefully adjusted. If set too low, noise from the piezo element will be counted as taps and if set too high, real taps will be missed. After a tap there is a 200 ms period in which no tap will be registered; this will limit the tapping rate to five taps per second.

Even if the Arduino guarantees a millisecond resolution both when playing sounds and registering taps there is still a need to evaluate the tapping board. A high speed camera (Sanyo Xacti VPC-HD2000) with an update frequency of 600 Hz (that is, one frame every 1.7 ms) was used to test the total delay of the system. A red LED was connected to the Arduino and was programmed to light up as soon as the piezo element registered a tap. When filmed with the high speed camera, the tap onset and the lightning of the LED always occurred in the same frame, so an upper limit to the latency and jitter of the tapping board should be 1.7 ms.

2.3 Measures of sensorimotor synchronization performance

When performing a study that includes an SMS task, for example, a finger tapping task, the research objective is often to compare timing performance between different experimental conditions (for example, the difference between on-beat and off-beat tapping as in Vos and Helsper, 1992) or between different groups (for example, the difference between persons with cerebellar damage and a control group as in Ivry and Keele, 1989). A third research objective can be to investigate whether timing performance correlates with other capacities such as intelligence (Holm et al., 2011) or working memory capacity (Bååth, Submitted). These types of research objectives require the calculation of a measure of performance for each participant and condition. This section goes through some common measures, focusing on the common case where the SMS task involves keeping the beat to an isochronous metronome sequence.

Table 2.1 show data from the first ten sounds of a finger tapping trial from Bååth and Madison (2012), collected using the tapping board described in the previous section. Here the metronome sequence had an ISI of 1200 ms and the participant was asked to start tapping as soon as the sequence started. *Sound onset* and *Tap onset* are timestamps given in ms since the beginning of the trial. *Asynchrony* gives the time

Sound no.	*Sound onset*	*Tap onset*	*Asynchrony*	*ITI*
1	0	-	-	-
2	1200	-	-	-
3	2400	2384	-16	-
4	3600	3581	-19	1197
5	4800	4710	-90	1129
6	6000	5939	-61	1229
7	7200	7144	-56	1205
8	8400	8381	-19	1237
9	9600	9543	-57	1162
10	10800	10770	-30	1227

Table 2.1: An example of SMS data from the finger tapping task in Bååth and Madison (2012). See section 2.3 for a description of the variables.

difference between the target sound onset and the tap onset, where a negative asynchrony implies that the tap preceded the tone. *ITI* shows the intertap interval, the time between adjacent taps. A dash in the *Tap onset* column means that the participant did not make any response to that sound. As it may take some time before the participant gets a feeling for the rhythm of the sequence (hence the missing first two taps in table 2.1) it is common to omit the first few onsets from each trial. Figure 2.4 shows three different graphical representations of the trial data.

Given a data set, such as that in figure 2.1, one can use the ITIs or the asynchronies to calculate a measure of timing performance. In the case where the task involved self-paced tapping there are no tone onsets and only the ITIs are available. A measure of performance is then the variability of the ITIs, where a low variability means that the participant produced more consistent responses and so had better timing performance. Here there are many possible measures of variability such as the sample variance, mean absolute deviation or median absolute deviation, but the conventional choice is to use the sample standard deviation (SD). Even if the task involved externally paced tapping, it is still possible to measure timing performance using the ITIs. However, one can argue that using the asynchronies is more in line with the task – to tap along to the sound sequence – as the asynchronies measure the time difference between the sounds and the responses.

Asynchronies are often sufficiently close to normally distributed to

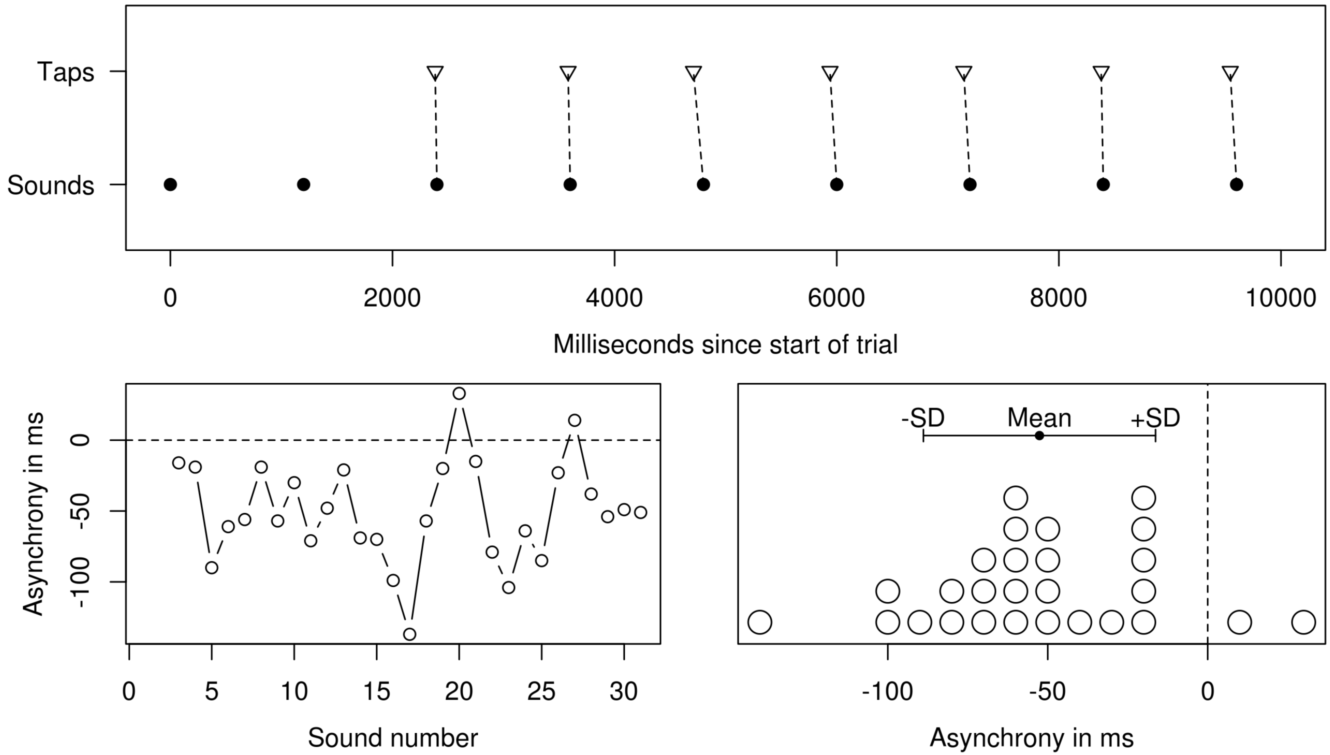


Figure 2.4: Three different graphical representations of a tapping trial from Bååth and Madison (2012). *Upper figure*: This shows a time series of the first nine sounds and the corresponding responses. *Lower left figure*: A time series showing the asynchronies. The dashed line marks where a response would fall to be perfectly in time with the stimuli onset. *Lower right figure*: A histogram showing the distribution of the asynchronies.

be well summarized by the sample mean and the sample SD (Bååth, 2015a). The mean asynchrony, sometimes called the constant error, is not a good measure of performance. As positive and negative asynchronies will cancel out, a participant with wildly inconsistent responses might still end up with a mean asynchrony that is small. However, the mean asynchrony is a good measure of the tendency of a participant to produce positive or negative asynchronies, and a well established result is that under a wide range of conditions participants have a negative mean asynchrony (Repp, 2005). This can also be seen in table 2.1 and figure 2.4 where the taps tend to precede the sounds. To avoid that the positive and the negative asynchronies cancel out, one could take the absolute value of the asynchronies and calculate the mean absolute asynchrony. This could be take as a measure of performance, however, this measure will be confounded by that mean asynchrony tends to be negative and by the considerable inter-individual difference in the magnitude of the mean asynchrony (Aschersleben, 2002). For example, a participant with perfect timing, but with a constant asynchrony of -50 ms, would by this measure perform worse than a participant with a mean asynchrony of 0 ms but with a 60 ms SD of the responses. A solution is then to take the variability of the asynchronies as a measure of performance, and again there are many possible measures of variability, but the conventional choice is to use the sample SD. The lower right sub-figure in figure 2.4 show both the mean and the SD of the asynchronies.

As the increase in timing variability is approximately linear within a wide range of tempi (Grondin, 2012), the coefficient of variation (CV), calculated as the SD of the asynchronies divided by the target interval, can be used to compare timing performance between different tempo levels (See figure 4 in Bååth et al. (Submitted) for an example of this use). Further, it is possible to isolate components of timing variability by using, for example, the model of Wing and Kristofferson (1973).

While the asynchrony SD is a direct measure of rhythm *production* performance it is commonly used as a proxy measure of rhythm *perception* performance, that is, how well a participant perceives aspects of a rhythmic stimulus. An example of this use is found in the paper *Tapping to Bach* where Toiviainen and Snyder (2003) were interested in measuring participants' ability to find the pulse in different musical excerpts taken from Bach's fourth organ duetto (BWV 805). Participants were asked to tap along to the perceived beat of the musical excerpts, and timing variability was used as a measure of how well the participants perceived the beat.

3. Rhythm perception and rhythm production at slow tempi

A main research question of this thesis has been: What happens with rhythm perception when the tempo is slow? Related questions are here: What is *slow*? And slow compared to what?

One approach is to first look at what ISIs could be considered a moderate or a natural tempo. A result that has been replicated several times is that when asked to freely tap out a regular rhythm at a comfortable tempo, participants tend go for a tempo around an ISI[1] of 500 ms (Moelants, 2002; Fraisse, 1982). There is, of course, a considerable difference between participants, where the resulting tempo, called the spontaneous motor tempo, can range from ISIs of 300 to 800 ms (McAuley et al., 2006). A tempo could then be considered slow when it is considerably slower than the spontaneous motor tempo.

A less common task, that directly targets *slowness*, is the slow motor tempo task. This is a variation of the spontaneous motor tempo task in where participants are asked to tap out a regular rhythm as slowly as possible while still being able to keep a regular beat. The tempi that participants produce in this task have large variability, possibly as the result of how participants interpret the task instructions, but on average the slow motor tempo is in the neighborhood of an ISI of 2.5 seconds (McAuley et al., 2006; Bååth, In press; Bååth, Submitted). The slow motor tempo does not define which tempi could be considered slow, but rather gives a point of reference for when a tempo can be considered slow.

[1] What you measure is, strictly, the intertap interval (ITI). Here interstimulus interval (ISI) is just used as a measure of tempo.

Another approach is to look at what is considered a moderate tempo, and what is considered a slow tempo, in a musical context. Van Noorden and Moelants (1999) found that in a sample of music played on the radio, the most common tempo had an ISI of 500 ms. This is very similar to the spontaneous motor tempo and corresponds to 120 BPM which would be denoted as *Moderato*[2] using Italian tempo markings (Randel, 2003). A tempo marking of *Lento* (*slow*) or *Largo* (*broad*) indicates a tempo in the range 40 to 60 BPM corresponding to ISIs in the range 1000 to 1500 ms. A tempo of 40 BPM is also the slow tempo limit of many metronomes, and has been so since the advent of the modern metronome. Here in the words of John Maelzels, the inventor of the modern metronome, from his original patent of 1815:

> **By this application of a sliding weight to the upper part of the stem or rod of the pendulum, I can render the machine very portable and convenient; for a pendulum that will vibrate 40 or 50 times per minute, will be found to suit the performance of the slowest music;**

Figure 3.1 shows a drawing of John Maelzels's metronome from this patent and the displayed tempo range, 50 to 160 BPM corresponding to ISIs in the range 375 to 1200 ms, can be seen as an indication of the range of tempi used in music[3]. This range also corresponds well with the range of tempi in the sample of radio music collected by van Noorden and Moelants (1999). There are, of course, music played at a slower tempo than 40 BMP, where *Larghissimo* can denote a tempo slower than *Largo*. An extreme example is Morton Feldman's *Last Pieces* where part III lack note durations and is to be played "very slow", making it up to the performer to choose what he or she considers to be a "very slow" tempo. In a number of performances by different pianists the average interval between the tones ranged from 1500 to 5000 ms (Moelants, 2001). Performing this piece is akin to performing a slow motor tempo task in that it is up to the performer or participant to decide what "as slow as possible" or "very slow" means. It is interesting to note that the range of tempi in performances of *Last Pieces* part III (ISIs 1500 to 5000

[2] *Moderato* would in English be *moderate* and roughly corresponds the tempo of, for example, Aretha Franklin's *Respect* or Lady Gaga's *Poker Face*.

[3] Modern metronomes, for example, those produced by Wittner GmbH, often extend this to the range 40 to 208 BPM corresponding to ISIs 289 to 1500 ms.

ms) and the range of tempi in the slow motor tempo task (ISIs 1500 to 6000 ms) (Bååth, Submitted) are very similar.

While no strong conclusions can be drawn from looking at the motor tempo tasks and conventions in a musical context, the data can be summarized as follows: Tempi in the neighborhood of an ISI of 500 ms could be considered neither fast nor slow, but moderate. A tempo at an ISI of 1000 ms could be considered slow and a tempo at an ISI of 1500 ms could be considered as very slow.

There exists a host of phenomena and effects related to rhythm perception, and there are many comprehensive reviews on the subject, for example, Patel (2006), Desain and Windsor (2000) and Honing (2012). Below I review a subset of phenomena which are more relevant to the papers of this thesis.

3.1 Subjective experience of slow rhythms

My own subjective experiences of keeping the beat to sequences of different tempi are the following: When the tempo is fast, say with an ISI of 250 ms, I need to focus as it is difficult to motorically produce such as fast rhythm by, for example, tapping with my index finger. At a moderate tempo tapping feels automatic and effortless. Starting around an ISI of 1000 synchronizing becomes gradually more demanding, requiring attention and focus. At very slow tempi, e.g., at an ISI of 3000 ms, keeping the beat is very difficult; I have no *feeling* of a rhythm, and I experience each tap as being the result of a deliberate decision.

My description of what it feels like to synchronize to a very slow rhythm is similar to the description by Repp (2006). This description is also in line with results that I presented at the 12th International Conference on Music Perception and Cognition in Thessaloniki (Bååth and Madison, 2012). I asked participants to tap along to metronome sequences at tempi ranging from ISIs of 600 ms to 3000 ms, and to after each trial rate the experienced difficulty. On average, participants rated it as easy to synchronize at ISI levels 600 and 1200 ms, but rated it as more and more difficult as the tempo approached an ISI of 3000 ms (see figure 1 in Bååth and Madison, 2012). This increase in difficulty is also mirrored in that, as the tempo goes from moderate to very slow, performing rhythmic timing increasingly demands attentional resources and involvement of cognitive control (Bååth et al., Submitted).

The increase in subjective difficulty and increasing demands on

Fig 1.

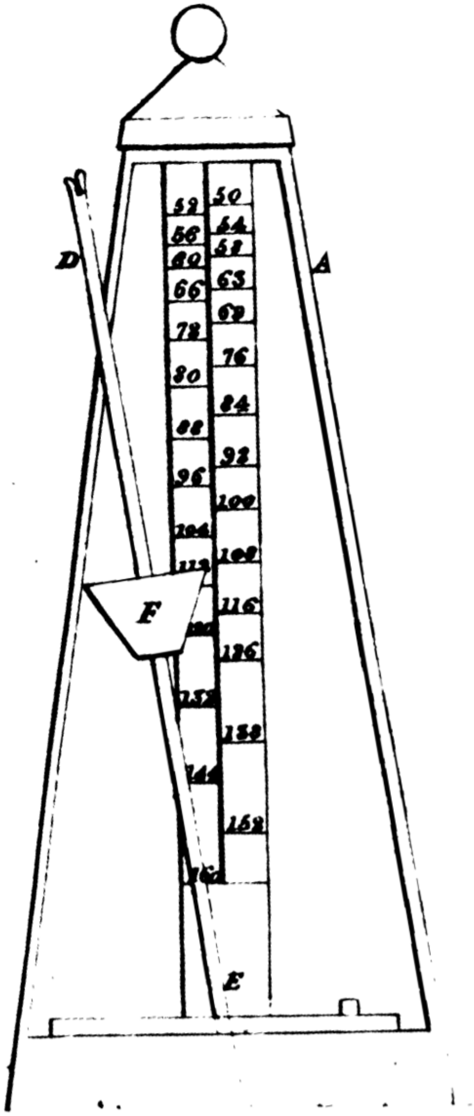


Figure 3.1: A drawing of the "Metronome or Musical Time-keeper" from the original metronome patent granted John Maelzel in 1815.

Figure 3.2: The three levels of complexity of the stimuli in the study conducted by Joel Olofsson.

attentional resources at slower tempi was seen in rhythm *production* tasks. However, the motor component of finger tapping does not get more difficult to carry out at slow tempi, if anything it should become easier to carry out as there is more time to plan the actions. It is plausible that these results instead have their basis in rhythm *perception* and that perceiving the rhythm is what is getting more difficult at slow tempi.

That slow rhythms are perceived with less intensity can also be seen in a data set collected by Joel Olofsson as part of his master thesis, which I supervised. Thirty-two participants were asked to listen to a number of rhythm sequences and rate "How much of a sense of rhythm do you experience?"[4] on a scale from 1 ("Experiences no sense of rhythm") to 7 ("Experiences a strong sense of rhythm"). The rhythm sequences were of different tempo, where the length of a bar was from 150 ms up to 3000 ms, and of different complexity, where the three complexity levels are shown in figure 3.2. The result of the study is summarized in figure 3.3 which shows the mean rating for each type of rhythm sequence. For the sequences of low complexity (1/1), the rated rhythmicity peaks when the length of a bar is 408 ms, which is close to the preferred tempo found in spontaneous motor tempo tasks (Moelants, 2002). As the tempo gets slower the rated rhythmicity drops and at a bar length of 3000 ms the mean rhythmicity is 2.0. The same pattern can be seen in the more complex sequences (1/2 and 1/4), with the difference that these sequences have higher overall mean rhythmicity and that the drop in rated rhythmicity begins at longer bar lengths.

[4]This is a rough translation of the original question, given in Swedish, "Hur mycket rytmkänsla känner du?".

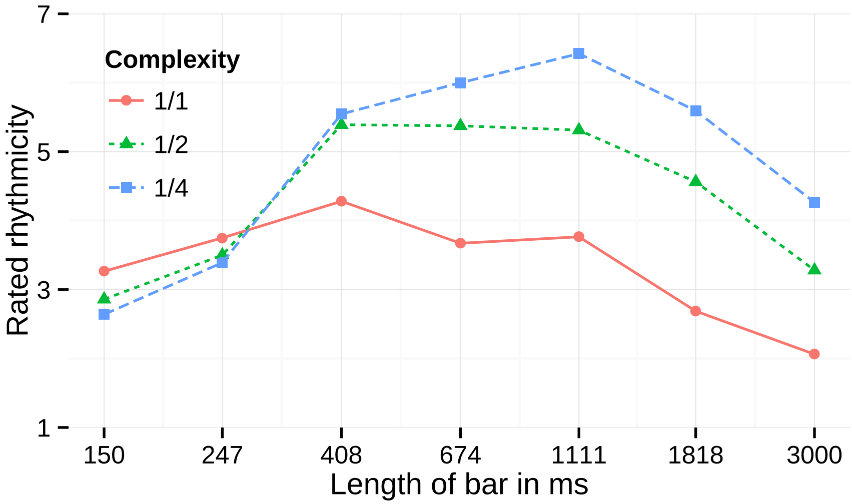


Figure 3.3: The mean ratings of rhytmicity for all combinations of complexity and tempo in the study conducted by Joel Olofsson.

3.2 Sensorimotor synchronization performance at slow tempi

Timing variability, as measured by the SD of the asynchronies, increases with slower tempi and over shorter tempo spans this increase is well described by a linear function (Semjen et al., 2000). Over larger spans the increase appears non-linear and has been better described by a quadratic function (Repp and Doggett, 2007) or by an exponentially increasing function (Bååth et al., Submitted; Bååth, Submitted).

While asynchronies recorded in SMS tasks tend to be approximately normally distributed, this is not the case when the tempo is sufficiently slow. Around an ISI of 1800 ms the distribution of the asynchronies starts to become left-skewed and bimodal, and this gets more pronounced at slower tempi (Mates et al., 1994; Miyake et al., 2004; Bååth, 2015a). An example of this is shown in figure 1 in Bååth (2015a). The reason for this increasing non-normality is that participants start to produce *reactive responses*, that is, instead of anticipating the upcoming tone onset, participants react to it producing what looks like auditory reaction time responses. This behavior has been interpreted as resulting from a deliberate shift in participants' strategy when synchronizing to slow rhythm sequences (Miyake et al., 2004). However, it has been shown that a change in strategy is not necessarily implied, as this behavior can also be explained by that participants' synchronization errors become so large so that the target tone is regularly missed and instead reacted upon (Repp and Doggett, 2007; Bååth, 2015a).

Independent of the cause of the reactive responses, it can be shown that standard estimators of the mean and SD of the asynchronies are biased towards too low estimates. That is, by using the standard estimators it will appear that participants have more negative mean asynchronies and lower tapping variability than what is warranted by the data. This is a problem if one wants to investigate rhythm production and rhythm perception at slow tempi (ISI > 1800 ms). A solution to this problem was presented in Bååth (2015a) and involved using a problem specific Bayesian model, which was subsequently used to calculate timing variability in Bååth (Submitted) and Bååth et al. (Submitted).

3.3 Rhythm perception, auditory working memory, and executive function

It has been suggested that time perception is solely dependent on memory traces in working memory (Lewis and Miall, 2006). As auditory stimuli dominate over other type of stimuli both in the temporal domain in general (Ortega et al., 2014), and with respect to rhythm perception (Barakat et al., 2015; Repp and Penel, 2002; Glenberg et al., 1989), a reasonable speculation is that *audiory working memory* and rhythm perception should be related. One would especially suspect a correlation between the temporal capacity of auditory working memory and rhythm production performance at slow tempi. A motivation for such a correlation would be that if rhythm perception depends on a memory component, and if rhythm production at a slow tempi is facilitated by a temporally extensive memory capacity, then participants with a long auditory working memory capacity should perform relatively better when synchronizing to slower tempi. To investigate this I conducted a study (Bååth, Submitted) where participants performed an auditory digit span task – a task commonly used to measure auditory working memory capacity (Baddeley, 2000; Hester et al., 2004) – and a SMS finger tapping task[5]. Despite large inter-individual variability in both auditory working memory capacity and timing performance there was a very small effect of auditory working memory capacity on timing performance. Furthermore, there was no evidence that memory capacity was related to better timing performance at slow tempi. There are two different interpretations of this result: (1) It can be taken as evidence that there is no strong connection between auditory working memory and rhythm perception and that a slower limit of rhythm perception does not directly depend on a temporal limit of working memory. (2) It can be taken as evidence that the auditory digit span task is not a good measure of auditory working memory capacity.

In line with the interpretation that there is no strong connection between auditory working memory and rhythm perception, there are models that posit a mechanism dedicated to rhythm perception (for example, Large, 2008). There are also models that posit mechanisms dedicated to timing but with a temporal limit (Pöppel, 2004; Mates et al., 1994; Miyake et al., 2004). Lewis and Miall (2006) calls timing which relies on a dedicated mechanism *automatic timing*, whereas timing of intervals beyond the temporal limit is named *cognitive timing*. This latter term reflects that timing of longer intervals require attentional and executive resources (Lewis and Miall, 2003). To investigate whether rhythmic timing requires more attentional resources at slow tempi, I, together with group of master students[6], performed a dual task study where the main task was a SMS task and where the distractor task was a covert response 2-back task (Bååth et al., Submitted). A second motivation for performing this study was that Holm et al. (2013) had not found any task interference between a rhythmic tapping and a task designed to require attentional resources, a result that I found strange. We did find that performing the 2-back task simultaneously as the SMS task resulted in worse performance in both task, and that this deterioration in performance increased with slower tempi.

[5] A second reason for why I did this study was that in many situations where I have talked with psychologists about rhythm perception, and especially the slow limit of rhythm perception, the comment I got was: "That must surely be related to working memory!"

[6] T. Tjøstheim, M. Lingonblad, F. Nelhans, D. Sivén, R. Yamazaki and H. Siljebråt.

3.4 Subjective rhythmization

There is a long tradition of investigating visual illusion in order to better understand visual perception (Eagleman, 2001). The are also cross-modal illusions, like the illusory flashing illusion, where the number of perceived flashes is modulated by the number of beeps accompanying the flashes (Shams et al., 2000). Compared to the wealth of known visual illusions there are relatively few known auditory illusions (but see Madison, 2009 and Hoopen, 2008).

One auditory illusion is *subjective rhythmization* (SR), an illusion with connections to meter perception, described in the 18th century by Kirnberger (1776) and first studied by Bolton (1894). The illusion is that sounds of a monotone metronome sequence are experienced as having different intensity, with the experienced intensity differences following a regular pattern. For example, it is common that every second or every fourth sound is perceived as accented, effectively grouping the sequence and imposing what could be described as a metric structure. Despite recent interest in the electrophysiological properties of SR (e.g., Nozaradan et al., 2011; Schaefer et al., 2011) there was, prior to my work on the subject, only one modern study, that by Vos (1973)[7], which employed Bolton's (1894) experimental paradigm. I have focused on this illusion in one short conference paper (Bååth and Ingvarsdóttir, 2014) and in one paper accepted for publication in *Music Perception* (Bååth, In press).

Subjective rhythmization is relevant to rhythm production at slow tempi, and especially the notion of a slower limit of rhythm perception, because the illusion is highly tempo dependent (as illustrated in figure 5 in Bååth, In press). When the tempo is fast, around an ISI of 200 ms, participants tend to experience every eight or every fourth sound as accented. At a moderate tempo, around an ISI of 600 ms, participants tend to experience every second sound as accented. When the tempo is very slow, around an ISI of 2000 ms, most participant respond that they do not experience the illusion anymore. That is, while a slower limit of rhythm perception can appear a somewhat elusive concept – it has not been possible to pinpoint a slow limit using SMS tasks – SR is a rare example of a task related to rhythm perception that has a slower limit. Why SR occurs can be explained within the resonance theory frame-

[7]The work of Vos (1973) is only published in Dutch, but has been reanalyzed by van Noorden and Moelants (1999).

work of rhythm perception chiefly developed by Large (2010). This framework, which is based on the notion of neural oscillation, has been used to model as diverse phenomena as SMS behavior in the presence of tempo changes (Loehr et al., 2011), the perception of polyrhytmic stimuli (Angelis et al., 2013), and categorical rhythm perception (Bååth et al., 2014a). With respect to SR, the resonance framework can be used to explain why the phenomena occurs, its tempo dependence and why participants tend to report that every second, fourth or eighth tone is accented more often than other possible divisions (Bååth, In press).

An update on Experiment 2 in Bååth and Ingvarsdóttir (2014)

In the SR paradigm employed in Bolton (1894), Vos (1973) and Bååth (In press), participants are explicitly asked whether they experience any subjective accents. This might put undue pressure on participants to report experiencing SR, and it would be desirable to demonstrate that participants experience SR without explicit instructions. A pilot study was presented in Experiment 2 in Bååth and Ingvarsdóttir (2014) where participants were asked to listen to a number of metronome sequences and indicate whether all sounds were equally intense or not. A participant got either the instruction that every second or that every fourth sound could be more intense, however, in the sequences played to the participants all sounds were identical. The only thing that was varied in the experiment was the task instructions and the tempo of the sequences. For more information regarding the method see section 3.1 in Bååth and Ingvarsdóttir (2014). Even if participants were not explicitly told about SR, they responded as if they did experience the illusion. However, the data was to weak to submit to statistical analysis.

Here I present an unpublished dataset that build upon Experiment 2 in Bååth and Ingvarsdóttir (2014). The task was identical to that in Experiment 2 except for that (1) only the ISI levels 275 and 700 ms were used, (2) each participant was given both the "every second" and "every fourth" instructions in randomized order, and (3) participants were given five trials per condition, totaling 20 trials. The ISI levels were selected as to maximize the probability of the participants experiencing that every second or fourth sound was accented at the 700 ms and 275 ms ISI level, respectively. Fourty-five participants was recruited by means of public advertising and the results are summarized in figure 3.4. In the "every second" condition participants reported more

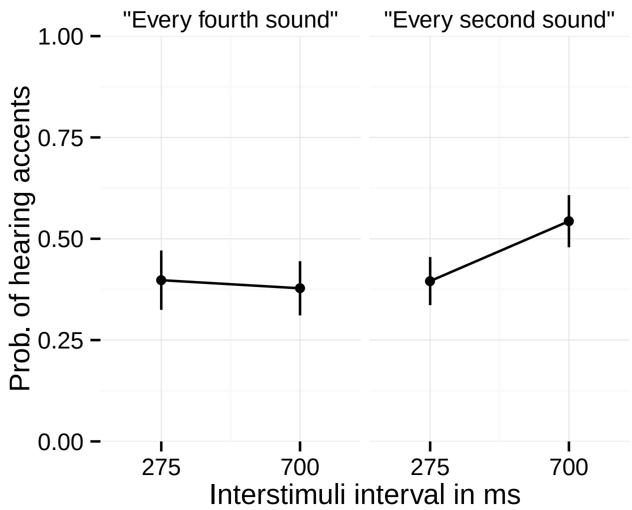


Figure 3.4: The mean proportion of trials in which the participants responded hearing an accent. The data is from the follow-up study of Experiment 2 presented in Bååth and Ingvarsdóttir (2014).

often hearing accents at the slower tempo, an mean increase of 15 percentage points, while in the “every fourth” condition there was a negligible difference between the two ISI levels, a mean difference of -2 percentage points. The difference in this mean increase of reported accents between the two conditions was statistically significant (paired t-test, M = 0.17, 95% CI: [0.044, 0.29], $p < 0.01$). This difference indicates that the participants experience some SR, despite not being explicitly instructed about the phenomena. However, the negligible difference between the two ISI levels in the “every fourth” condition is puzzling and warrants further investigation.

3.5 Estimating a slower limit of rhythm perception

A secondary research question of this thesis has been: What is the slower limit of rhythm perception? Already Woodrow (1932) noticed that “the task of synchronizing at the rates of one sound every 2 sec or every 4 sec seemed definitely different from that at rates of one sound every 1 sec or less”. He also transcribed verbal reports from his participants, here when synchronizing to an ISI of 4000 ms:

> "I never seemed to act automatically as I did with the shorter intervals. Each tap was an individual reaction. [...] I couldn't get the 'feeling' of the rate or perform with any feeling of satisfaction the task of synchronizing."
>
> "It seems almost impossible to keep the taps synchronous. There was no basis apparently for noting the time interval. There was no rhythm [...]"
>
> "It was just a matter of 'blind' guessing when to tap the key."

Compare this to verbal reports from participants synchronizing to ISIs shorter than 1000 ms:

> "I felt as though my arm became an automatic machine after I had tapped for a while, when I seemed to be doing best. Then I did not voluntarily make each tap but it went on without my will – all I had to do was to set my arm in motion at the correct rate and it went ahead tapping."
>
> "The action of the tapping finger seemed to be controlled by the stimulus-series and more or less free from conscious control. In other words it became automatic, if that expression may be used. "

I include these quotes here because the verbal reports of Woodrow (1932) well match verbal reports from participants in my own experiments, and my own experience.

Now, there does not seem to be a slower limit, within reason, where participants are completely unable to synchronize to a metronome sequence (Repp, 2006). So, in the studies presented in this thesis I have attempted to probe a slower limit of rhythm perception using other methods, however, I have not been able to produce a final answer. My attempts to estimate this slower limit are summarized in the following sections. I conclude with Figure 3.6 that compiles my findings together with other findings from the literature related to slow limits of time perception and rhythm perception. As seen in this figure, there is no clear candidate for a slower limit of rhythm perception. I do not wish to over-interpret these results, but to point out that:

- The findings stretch a range of ISIs of 1000 to 3000 ms. While this is a wide range of tempi, all findings correspond to slow tempi (cf.

Largo, 40 - 60 BPM), or slower, from a musical perspective, and are far from a moderate tempo of 120 BPM.

- One cluster of findings is seen around an ISI of 1500 ms and one cluster is seen in the range 2500 to 3000 ms.

Using the slow motor tempo task

This task, described in the beginning of section 3, is seemingly a suitable task for pinpointing a slower limit of rhythm perception. Participants are asked to tap out a rhythm that is *as slow as possible*, while still being able to maintain a regular beat. Assuming that participants need to perceive the beat in order to maintain it, this task should result in participants tapping at tempi close to their slower limit of rhythm perception. The first example of this task, that I have been able to find, is given in McAuley et al. (2006). Their result was stratified by age group, but looking at the age group 18-38 years, the mean intertap interval was 2532 ms in the slow motor tempo task. I replicated this task in Bååth (In press) and found a similar mean intertap interval of 2757 ms.

A problem with this task is that it is open to interpretation and that it is possible to approach the task in many ways. For example, participants might subdivide the beat covertly to facilitate tapping a slow beat, even when given instructions not to. To try to mitigate this problem I introduced a modified slow motor tempo task with overt counting, that is, participants were asked to count their taps aloud. The idea being that the overt counting would preclude covert counting. The resulting median intertap interval was 2719 ms (Bååth, Submitted), similar to the result in the two earlier studies.

Using subjective rhythmization

The connection between a slow limit of subjective rhythmization (SR) and a slow limit of rhythm perception is not directly apparent. However, as experiencing SR implies being able to superimpose a subjective rhythmic structure on a monotone sequence, it can be assumed that in order to experience SR at a certain tempo one must be able to perceive rhythm at that tempo. The slow limit of SR can therefore be seen as a lower bound of the slow limit of rhythm perception. Furthermore, if one accepts the resonance theory explanation of SR given in Bååth (In press), there is a direct connection between the slower limit of SR and the slower limit of rhythm perception: The slower limit of rhythm perception is approximately twice that of the slower limit of SR.

The slower limit of SR given by Bolton (1894) is at an ISI of 1600 ms. In Bååth (In press) there was no ISI level where all participants reported experiencing no SR. However, at an ISI of 1500 ms less than half of the trials resulted in that the participant experienced no SR, and this can be taken to define the slower limit. Another method to define a slower limit of SR is to calculate the *mean group period* of the reported groupings in the SR task (see the Analysis section in Bååth, In press, for more details), which in Bååth (In press) equaled 1881 ms.

Using sensorimotor synchronization data

While there is no slower limit where participants are unable to synchronize to a metronome sequence, it is possible to look at how synchronization influences other measures. For example, Miyake et al. (2004) argued for two types of anticipatory synchronization by showing that when a participant performed a distractor task and a synchronization task simultaneously, the number of reactive taps increased markedly at an ISI of 1800 ms. Marked changes, in the studies presented in this thesis, that occurred in other measures when participants synchronized at different tempi were:

- In Bååth and Madison (2012), the largest increase in rated difficulty of synchronization occurred between ISIs of 1200 and 1800 ms.
- In Bååth et al. (Submitted), the largest increase in timing performance, as measured by the log asynchrony SD, due to task interference occurred between ISIs of 897 and 1342 ms.
- In Bååth et al. (Submitted), the largest increase in the coefficient of variation, the percentage of reactive responses, and the number of errors in the n-back task, occurred between ISIs of 2006 and 3000 ms

When a change happens between two ISI levels, a reasonable summary is to take the midpoint between the two ISI levels. For the three marked changes above the midpoints are 1500, 1120 and 2503 ms, respectively. Note that these slow limits are highly tentative as a "marked change" can be defined in many ways and because the resulting midpoints are highly dependent on the ISI levels.

Using an optimality argument

One could also ask: What *should* be the slower limit of rhythm perception? There are a multitude of ways to approach this question and the exposition given below should just be viewed as an example of one such approach and as proof of concept[8].

Rhythm perception helps anticipation and prediction of events that occur in a pattern over time (Large and Jones, 1999). The ultimate purpose of all cognitive processes is to guide action. By perceiving the rhythm of a sequence, an agent *might* be able to better respond to an upcoming event by timing the response to the expected occurrence of that event, rather than just reacting to it. Here a criteria to optimize would be to minimize the distance in time between the response and the event, that is, to minimize the mean absolute error. If relying on rhythmic expectation results in a smaller mean absolute error then that is a preferred strategy over reacting to the event onsets.

However, it might not necessarily be better to rely on rhythmic expectation in all situations. One such situation is when trying to synchronize to a slow metronome rhythm. As timing variability increases as a function of the ISI, there comes an ISI that is so slow that it is a better strategy to simply react to the sound onsets. At tempi faster than this ISI, rhythmic expectation – and rhythm perception – are useful, with respect to the narrowly defined criteria to be optimized. At tempi slower than this ISI rhythm perception serves no purpose. Assume further that it is more costly to have a sense of rhythm that spans a wider tempo range. The optimal slower limit of rhythm perception is then the longest ISI where it is still advantageous to rely on rhythmic expectation over reacting to the sound onset.

It is possible to estimate this slower limit, given data on the mean absolute error at different tempo in rhythmic timing and data on auditory reaction time. Below I use the finger tapping data from the group of non-musicians in Repp and Doggett (2007) and auditory reaction time data from the group of adults in Löwgren et al. (2014)[9]. The mean absolute error was calculated for each participant and ISI level in the finger tapping data, and a line was fitted using linear regression with

[8] I thank Hans-Henning Schulze for suggesting this approach when we met at the 14th Rhythm perception and production workshop at University of Birmingham. Recently I also found that a similar calculation was presented by Woodrow (1932).

[9] Here I would like to thank Bruno Repp and Karolina Löwgren for giving me access to these datasets.

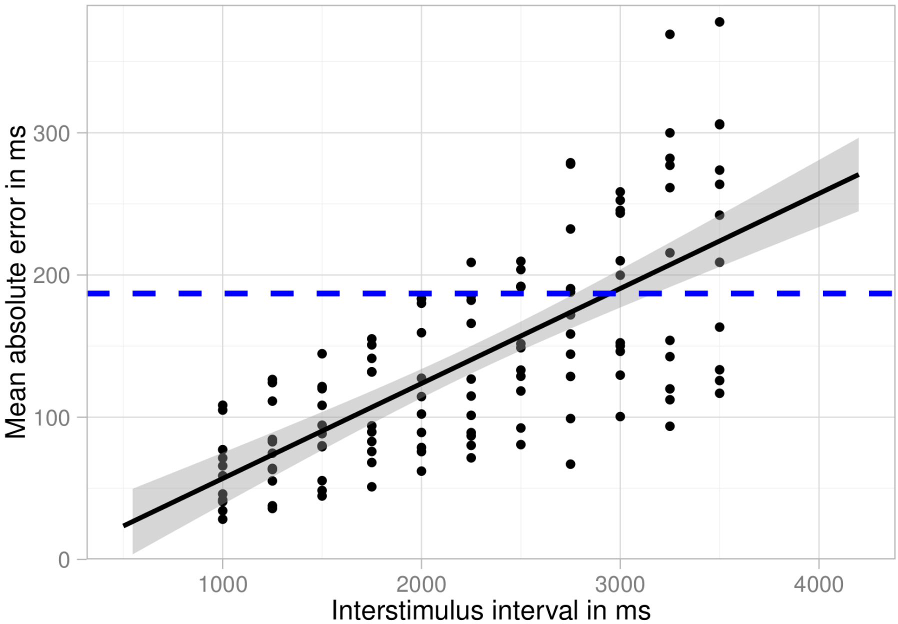


Figure 3.5: Each point show the mean abolute timing error for one of the non-musicians in Repp and Doggett (2007). The dashed line show the median auditory reaction time in Löwgren et al. (2014). The intersection between the dashed line and the solid line can be seen as an optimal slow limit of rhythm perception, according to the argument given in section 3.5.

ISI as the predictor variable and mean absolute error as the outcome variable. The median was taken of the auditory reaction time data. The resulting estimates are shown in Figure 3.5, where the optimal slow limit of rhythm perception is the intersection of the mean absolute tapping error (the solid line) and the median auditory reaction time (the dashed line). Given these data sets the optimal slow limit was at an ISI of 3000 ms.

Now, the exposition given above is, of course, a gross oversimplification. Yet, it is interesting to note that the optimal slower limit here calculated is not far removed from what would be expected, and that it coincides with the slow limit of rhythm perception as calculated using SR in Bååth (In press) and with Pöppel's (1997) window of temporal integration.

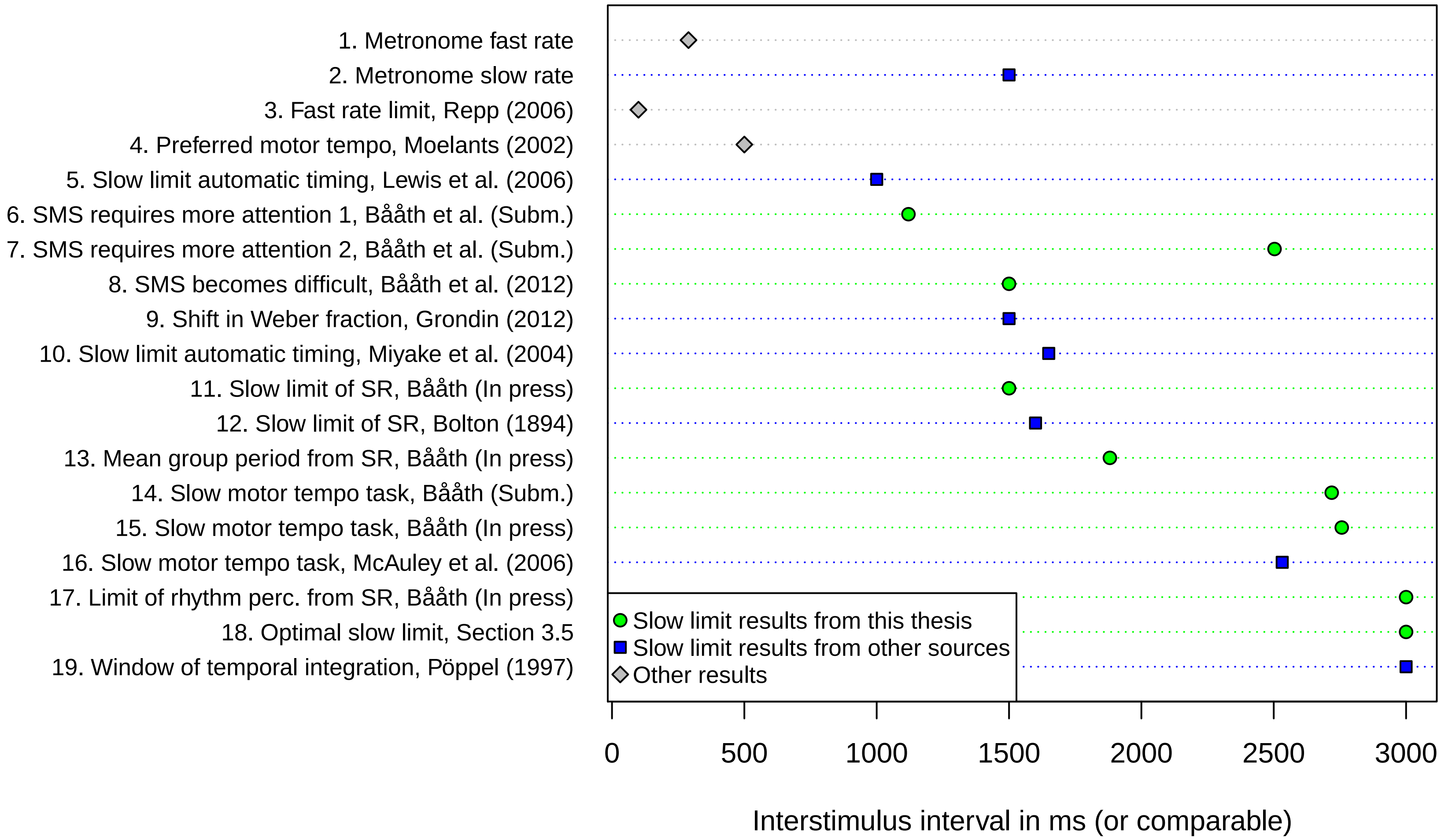

1. Metronome fast rate
2. Metronome slow rate
3. Fast rate limit, Repp (2006)
4. Preferred motor tempo, Moelants (2002)
5. Slow limit automatic timing, Lewis et al. (2006)
6. SMS requires more attention 1, Bååth et al. (Subm.)
7. SMS requires more attention 2, Bååth et al. (Subm.)
8. SMS becomes difficult, Bååth et al. (2012)
9. Shift in Weber fraction, Grondin (2012)
10. Slow limit automatic timing, Miyake et al. (2004)
11. Slow limit of SR, Bååth (In press)
12. Slow limit of SR, Bolton (1894)
13. Mean group period from SR, Bååth (In press)
14. Slow motor tempo task, Bååth (Subm.)
15. Slow motor tempo task, Bååth (In press)
16. Slow motor tempo task, McAuley et al. (2006)
17. Limit of rhythm perc. from SR, Bååth (In press)
18. Optimal slow limit, Section 3.5
19. Window of temporal integration, Pöppel (1997)
Slow limit results from this thesis
Slow limit results from other sources
Other results
0
500
1000
1500
2000
2500
3000
Interstimulus interval in ms (or comparable)

Figure 3.6: Some different pointers to the temporal location of a slow limit of rhythm perception.

All measures given in ms refer to time intervals, such as, interstimulus intervals or interresponse intervals, where the type of interval should be clear from the context.

1. For reference, the fastest rate of a Wittner Metronome No. 830 : 208 BPM equaling **289 ms**
2. The slowest rate of a Wittner Metronome No. 830 : 40 BPM equaling **1500 ms**
3. For reference, the fast rate limit of sensorimotor synchronization as described by Repp (2006) : **100 ms**
4. For reference, an estimate of the average preferred motor tempo by Moelants (2002) : **500 ms**
5. The cut-off where *automatic* timing transitions into *cognitive* timing according to Lewis et al. (2006) : **1000 ms**
6. In Bååth et al. (Submitted) the participants synchronized to a metronome while performing a distractor task designed to require attentional resources. The largest increase in timing interference, as measured by the log asynchrony SD, took place between 897 ms and 1342 ms; this is the midpoint: **1120 ms**
7. Same as in (6). But instead looking at the coefficient of variation, the percentage of reactive responses and the number of errors in the distractor task. The largest increase in interference for these three measures took place between 2006 ms and 3000 ms; this is the midpoint: **2503 ms**
8. In Bååth et al. (2012), when participants synchronized finger taps to a metronome, the largest increase in rated difficulty took place between 1200 and 1800 ms; this is the midpoint: **1500 ms**
9. Grondin (2012) found a shift in the Weber fraction between 1000 ms and 1900 ms in a number of timing tasks : **1500 ms**
10. *Automatic* finger tapping transitions into finger tapping that requires attentional resources between 1500 and 1800 ms according to Miyake et al. (2004); this is the midpoint : **1650 ms**
11. The slow limit of subjective rhythmization measured as the tempo where more than half of the trials result in the participants not experiencing the illusion in Bååth (In press) : **1500 ms**
12. The slow limit of subjective rhythmization from Bolton (1894) : **1600 ms**
13. Mean group period in the subjective rhythmization task in Bååth (In press) : **1881 ms**
14. Median slow motor tempo in Bååth (Subm.) : **2719 ms**
15. Mean slow motor tempo in Bååth (In press) : **2757 ms**
16. Mean slow motor tempo in McAuley et al. (2006) for the age group 18-38 years : **2532 ms**
17. Limit of rhythm perception measured as twice the slow limit of subjective rhythmization (11), this according to the resonance theory explanation of subjective rhythmization in Bååth (In press) : **3000 ms**
18. The optimal slow limit of rhythm perception according to the analysis in section 3.5 : **3000 ms**
19. Window of temporal integration according to Pöppel (1997) : **3000 ms**

4. Introduction to the papers

4.1 Paper I – The subjective difficulty of tapping to a slow beat

Motivation

There are verbal report, at least dating back to Woodrow (1932), of that it is *difficult* to feel and produce very slow rhythms. As Repp (2006) writes: "At [interstimulus intervals] up to 1500 ms or so, [keeping a rhythm] seems to proceed effortlessly and automatically, but the task begins to feel laborious as the [interstimulus interval] approaches 1800 ms". What is the cause of this perceived difficulty and is it generally the case, as Repp suggests, that the difficulty starts as the tempo approaches 1800 ms? These questions relate to the more general research question of this thesis: Is there a slow limit of rhythm perception, and if so, at what tempo is that limit?

Procedure

Thirty participants were recruited from the Lund community. Participants were asked to perform a standard sensorimotor synchronization tapping task where the stimuli consisted of isochronous sound sequences corresponding to tone interstimulus intervals (ISI) of 600, 1200, 1800, 2400 and 3000 ms. Each participant was presented with 20 trials, four for each ISI level. After each trial the participant rated the difficulty of tapping on a seven point scale ranging from "very easy" to "very difficult". The custom built tapping board described in section 2.2 was used both to play the sound sequences and to record the participants' responses.

Conclusion

There was a strong, statistically significant correlation between ISI and rated difficulty. This was expected, given the many verbal reports supporting such a correlation, but, to my knowledge, it is the first time it has been shown in an experiment. In general, the participants rated it as easy to synchronize at the ISI levels 600 and 1200 ms. At the 1800 ms ISI level there was marked increase in rated difficulty, supporting Repp's (2006) notion of at what tempo keeping a rhythm becomes difficult. This perceived difficulty was also shown to mostly depend on the tempo, and not on how well a participant actually manage keep the rhythm during a trial.

4.2 Paper II – A prototype-based resonance model of rhythm categorization

Motivation[1]

Every time a piece of music is transcribed the person doing the transcription has to make a choice: How to represent what is heard using a limited number of musical symbols? As the number of ways to notate the rhythm of a piece of music is finite, and as different rhythmic sequences can be notated in the same way, transcribing a rhythm implies making a categorization. Desain and Honing (2003) showed that musically trained participants reliably experienced rhythms as belonging to rhythmic categories, but that the participants' categorization depended on the metric structure of the rhythm sequences. They concluded that: "It is puzzling, however, that although meter was shown to exert a strong influence on the recognition of rhythm [...] existing computational models of meter can explain this phenomenon only to a small extent."

At this point in my PHD-studies I became interested in the prospect of using dynamical systems to model rhythm perception, where Large's (2010) resonance theory of rhythm perception was one of the most developed computational framework. This paper thus explored the

[1]This is the one paper in my thesis that is not directly related to rhythm perception at a slow tempo. However, it led me to explore the resonance framework of rhythm perception (Large, 2010), which I subsequently used in Paper IV to explain the phenomena of subjective rhythmization.

question: Can the resonance theory framework Large (2010) be used to model the type of rhythm categorization described by Desain and Honing?

Procedure

A computational model was implemented in MATLAB using the *Nonlinear time-frequency transformation workbench* developed by Large and Velasco (In preparation). The original dataset from Desain and Honing (2003) was encoded and given as input to the model which was used to make categorical decisions. That is, it took a rhythm defined in the continuous time domain and produced a categorization of this rhythm in the form of musical notation.

Conclusion

The categorization made by the computational model and the categorizations made by the participants agree to a large extent. The model also captured how changing the metric structure of the rhythm affected the participants' categorical choices. This result supports the notion that resonance theory is a viable model of rhythm perception and that by viewing rhythm perception as a dynamical system it is possible to model properties of rhythm categorization.

4.3 Paper III – Subjective rhythmization: A replication and an extension

Motivation[2]

Subjective rhythmization is an auditory illusion that shows how humans are primed to experience rhythm. The illusion occurs when one listens to a sequence of isochronous, identical sounds. A metric pattern of accents will emerge which gives the impression that there are groups of sounds. That is, even though the sounds are objectively identical

[2] In retrospect, this paper was a stepping-stone towards Paper IV. Both papers aimed at replicating the seminal work of Bolton (1894) on the phenomena of subjective rhythmization, but Paper IV improves upon Paper III in all aspects. Yet I decided to include this paper in the thesis as it constitutes an independent replication of the subjective rhythmization phenomena, and as it shows how my work on this thesis has progressed.

they sound subjectively different. Save for a small study by Vos (1973), the subjective rhythmization paradigm had not been replicated since the original study by Bolton (1894).

When I first encountered this illusion I was fascinated, both by the illusion itself, and by that the effect of subjective rhythmization is highly dependent on the sequence tempo. For example, Bolton (1894) argued for that the effect of subjective rhythmization vanished when the tempo was slower than an ISI of 1600 ms. The main research question of this study was: Can the phenomena of subjective rhythmization, and especially its tempo dependency, be replicated?

Procedure

One hundred and thirty-two participants were recruited using the online service Amazon Mechanical Turk. The participants were played monotone metronome sequences with ISIs of 200 , 300 , 400 , 600, 800 and 1500 ms and asked whether they experienced any accents.

Conclusion

The study replicated many of the findings of Bolton (1894) and Vos (1973). For example, that participants often report groupings of two, four or eight, the common meters of western music, and that what grouping participants experience is tempo dependent, with larger groupings being reported at faster tempi. When listening to the sequence with an ISI of 1500 ms, a majority of the participants (81%) reported hearing no accents which is in line with Bolton's estimate of the slow limit of subjective rhythmization. In conclusion, this study supported that subjective rhythmization is a robust phenomena that is experienced by a large proportion of the population.

4.4 Paper IV – Subjective rhythmization: A replication and an assessment of two theoretical explanations

Motivation

Paper III was a small online study and in this follow-up study I sought to study subjective rhythmization in an environment with better control over the external factors. Furthermore, I wished to test a number of prediction arising from two theoretical explanations of why subjective rhythmization occurs: the preferred tempo explanation of Temperley (1963) and the resonance theory explanation due to Large (2008). I considered the resonance theory explanation more plausible, a priori, and one exciting prediction made by this theory was that there should be a relation between how well a participant can synchronize to a slow metronome beat and how large groupings that participant will experience in the subjective rhythmization task. This relation resulting from that, within the resonance theory, both slow synchronization performance and group size can be seen as proxy variables to a slower limit of rhythm perception.

Procedure

Thirty participants were recruited from the Lund community and asked to perform a subjective rhythmization task and a number of rhythm production tasks. The custom built tapping board described in section 2.2 was used both to play the sound sequences and to record participants' responses.

Conclusion

All participants reported hearing accented tones in the subjective rhythmization task, despite being explicitly told that the sound sequences were monotone. As observed by Bolton (1894), what groupings participants experienced was highly tempo dependent. A new finding was that participants were also very consistent in their responses and tended to experience the same grouping in repeated trials of the same tempo. The results were in line with the predictions developed on the basis of resonance theory. For example, there was a correlation between how large groupings a participant experienced and how well a participant could synchronize at a slow tempo, showing a direct connection between a rhythm production task and an introspective rhythm perception task.

4.5 Paper V – Estimating the distribution of sensorimotor synchronization data: A Bayesian hierarchical modeling approach

Motivation

This is a technical paper that solves a problem that I had. I was convinced that the conventional method for calculating timing performance in sensorimotor synchronization tasks would give misleading results when used on rhythmic timing data from trials when the tempo was very slow. The reason for this is that at slow tempi participants start to react to the sound onsets, rather than anticipating the onsets, resulting in a skewed, bimodal response distribution with a long left tail and a short right tail. As I planned to study rhythm production at slow tempi I needed a robust method that allowed me to estimate timing performance at any tempo.

Procedure

I developed a statistical model that accounted for that participants make reactive responses at slow tempi. The model was developed using a Bayesian methodology, as this framework facilitated incorporating prior information regarding the distribution of reactive and anticipatory responses.

Conclusion

In a simulation study, and on a data set collected by Repp and Doggett (2007), I showed that using my method worked as well as conventional methods at moderate tempi, and worked considerable better at slow tempi. The best performing version of my method was a hierarchical Bayesian model, which was subsequently used to estimate rhythm timing performance in Papers VI and VII.

4.6 Paper VI – Working memory, memory for musical rhythms, and rhythm perception

Motivation

Here I examined the relationship between auditory working memory, sensorimotor synchronization performance, and memory capacity for rhythms. A number of predictions were made regarding the relationship between these capacities, based on the current literature on rhythm perception and working memory. One such prediction was that participants with a large memory capacity would be comparably better at synchronizing to slower sequences. This prediction was based on that synchronization to slow sequences requires longer time intervals to be retained and reproduced, and that a long auditory working memory span should be advantageous for retention of long intervals. A strong connection between memory capacity and slow synchronization would be relevant to my investigations into the slower limit of rhythm perception, as the slower limit could then be explained as resulting from the limited capacity of auditory working memory.

Procedure

Thirty-six participants were recruited through public advertising. Each participant performed a digit span task, a sensorimotor synchronization task, a slow motor tempo task, and a novel rhythm span task developed by Schaal et al. (2014). The custom built tapping board described in section 2.2 was used to record the participants' timed responses.

Conclusion

The results showed that auditory working memory – as measured by a digit span task – and memory capacity for rhythms are related. Auditory working memory and memory capacity for rhythms were also related to sensorimotor synchronization performance, albeit weakly. However, the influence of memory capacity on synchronization performance showed no interaction with sequence tempo, suggesting that auditory memory capacity does not play an integral role in rhythm production and that limits in memory capacity can not solely be used to explain the slower limit of rhythm perception.

4.7 Paper VII – The role of executive control in rhythmic timing at different tempi

Motivation

It has been proposed that different mechanisms are recruited by rhythmic timing depending on the tempo (Grondin, 2012). Relevant here is the notion of a slower limit of rhythm perception, the temporal boundary where perceiving and synchronizing to a rhythmic sequence goes from being effortless to requiring attention and executive control. This study used a dual-task setup to investigate whether rhythmic timing requires more attentional resources at slow tempi compared to a comfortable tempo. A secondary reason performing this study was that Holm et al. (2013) have argued for that rhythmic timing, even at slow tempi, does not require attentional resources, a proposition that that I found unlikely to be true.

Procedure

Twenty-four participants were recruited via public advertising. We employed a dual-task setup to investigate whether rhythmic timing required more attentional resources at slow tempi compared to comfortable tempi. The main task was a sensorimotor synchronization task in where participants synchronized their finger taps to metronome sequences ranging in tempo from an ISI of 600 to 3000 ms. On half of the trials there was a concurrent distractor task; a novel variant of the n-back task. The custom built tapping board described in section 2.2 was used to record the participants' responses.

Conclusion

The result showed that, when the tempo is sufficiently slow, performing rhythmic timing demands attentional resources and involvement of executive control. This result resonates with neural models of timing that suggest a dedicated timing mechanism for short intervals and a general, cognitive timing mechanism for longer intervals. A research question throughout my thesis project has been: What is the slower limit of rhythm perception? However, there is a more fundamental question: *Is there* a slower limit of rhythm perception? This study showed that rhythmic timing requires more cognitive resources the slower the tempo, and it can be assumed that both attentional resources and executive control are limited capacities. Therefore, independent of whether rhythmic timing depends on one or several mechanisms, this study supports the view that rhythm perception and rhythmic timing do have a slower limit.

Bibliography

Angelis, V., Holland, S., Upton, P. J., and Clayton, M. (2013). Testing a computational model of rhythm perception using polyrhythmic stimuli. *Journal of New Music Research*, 42(1):47–60.

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and cognition*, 48(1):66–79.

Bååth, R. (2010). Childfreq: An online tool to explore word frequencies in child language. *LUCS minor*, 16.

Bååth, R. (2011). Construction of a low latency tapping board. *LUCS minor*, 17.

Bååth, R. (2012). The state of naming conventions in R. *The R Journal*, 4(2):74–75.

Bååth, R. (2015a). Estimating the distribution of sensorimotor synchronization data: A bayesian hierarchical modeling approach. *Behavior Research Methods*.

Bååth, R. (2015b). Modeling match results in soccer using a hierarchical bayesian poisson model. *LUCS Minor*, 18.

Bååth, R. (In press). Subjective rhythmization: A replication and an assessment of two theoretical explanations. *Music Perception*.

Bååth, R. (Submitted). Working memory, memory for musical rhythms, and rhythm perception.

Bååth, R. and Ingvarsdóttir, K. O. (2014). Subjective rhythmization: A replication and an extension. In *Proceedings of the 13th International Conference on Music Perception and Cognition*, Seoul, South Korea.

Bååth, R., Lagerstedt, E., and Gärdenfors, P. (2014a). A prototype-based resonance model of rhythm categorization. *i-Perception*, 5(6):548–558.

Bååth, R. and Madison, G. (2012). The subjective difficulty of tapping to a slow beat. In *Proceedings of the 12th International Conference on Music Perception and Cognition*, pages 82–85, Thessaloniki, Greece.

Bååth, R., Sauerland, U., and Sikström, S. (2011a). A signal detection theory of vagueness. In *VAAG workshop*, Lund, Sweden.

Bååth, R., Seno, T., and Kitaoka, A. (2014b). Cats and illusory motion. *Psychology*, 05(09):1131–1134.

Bååth, R., Strandberg, T., and Balkenius, C. (2011b). Eye tapping: How to beat out an accurate rhythm using eye movements. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway.

Bååth, R., Tjöstheim, T., and Lingonblad, M. (Submitted). The role of executive control in rhythmic timing at different tempi.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11):417–423.

Barakat, B., Seitz, A. R., and Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current biology*, 25(2):R60–1.

Bolton, T. (1894). Rhythm. *The American Journal of Psychology*, 6(2):145–238.

Brown, D. E. (1991). *Human universals*. McGraw-Hill, New York City.

Cook, P., Rouse, A., Wilson, M., and Reichmuth, C. (2013). A california sea lion (zalophus californianus) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology*.

Dahlman, C., Sarwar, F., Bååth, R., Wahlberg, L., and Sikström, S. (2015). Prototype effect and the persuasiveness of generalizations. *Review of Philosophy and Psychology*.

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

Desain, P. and Windsor, L. (2000). *Rhythm perception and production*. Routledge.

Dunlap, K. (1910). Reaction to rhythmic stimuli with attempt to synchronize. *Psychological Review*, 17(7):399–416.

Eagleman, D. M. (2001). Visual illusions and neurobiology. *Nature reviews. Neuroscience*, 2(12):920–926.

Fitch, W. T. (2012). The biology and evolution of rhythm: unraveling a paradox. In P, R., M, R., JA, H., and I, C., editors, *Language and music as cognitive systems.*, pages 73–95. Oxford University Press.

Fraisse, P. (1982). Rhythm and tempo. In Deutsch, D., editor, *The psychology of music*, chapter 6, pages 149–180. Academic Press.

Geringer, J. and Madsen, C. (1984). Pitch and tempo discrimination in recorded orchestral music among musicians and nonmusicians. *Journal of Research in Music Education*, 32(3):195–204.

Glenberg, A. M., Mann, S., Altman, L., Forman, T., and Procise, S. (1989). Modality effects in the coding reproduction of rhythms. *Memory & Cognition*, 17(4):373–383.

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. and Rowe, J. B. (2013). Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cerebral cortex (New York, N.Y. : 1991)*, 23(4):913–21.

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.

Grondin, S. (2012). Violation of the scalar property for time perception between 1 and 2 seconds: Evidence from interval discrimination, reproduction, and categorization. *Journal of experimental psychology. Human perception and performance*, 38(4):880–890.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events.* The MIT Press, Cambridge, MA.

Hester, R. L., Kinsella, G. J., and Ong, B. (2004). Effect of age on forward and backward span tasks. *Journal of the International Neuropsychological Society : JINS*, 10(4):475–481.

Holm, L., Ullén, F., and Madison, G. (2011). Intelligence and temporal accuracy of behaviour: unique and shared associations with reaction time and motor timing. *Experimental brain research.*, 214(2):175–83.

Holm, L., Ullén, F., and Madison, G. (2013). Motor and executive control in repetitive timing of brief intervals. *Journal of experimental psychology. Human perception and performance*, 39(2):365–80.

Honing, H. (2012). Structure and interpretation of rhythm in music. In Deutsch, D., editor, *The Psychology of Music*, pages 369–390. Elsevier.

Honing, H. and De Haas, W. (2008). Swing once more: Relating timing and tempo in expert jazz drumming. *Music Perception*, 25(5):471–476.

Honing, H., Ladinig, O., Háden, G. P., and Winkler, I. (2009). Is beat induction innate or learned? probing emergent meter perception in adults and newborns using event-related brain potentials. *Annals of the New York Academy of Sciences*, 1169(Illc):93–6.

Honing, H., Merchant, H., Háden, G. P., Prado, L., and Bartolo, R. (2012). Rhesus monkeys (macaca mulatta) detect rhythmic groups in music, but not the beat. *PLoS ONE*, 7(12):e51369.

Hoopen, G. (2008). Classic illusions of auditory time perception. *Journal of the Human-Environmental System*, 11(1):27–35.

Ivry, R. B. and Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of cognitive neuroscience*, 1(2):136–52.

Johnson, C. M. (1996). Musicians' and nonmusicians' assessment of perceived rubato in musical performance. *Journal of Research in Music Education*, 44(1):84.

Kirnberger, J. P. (1776). *Die Kunst des reinen Satzes in der Musik: aus sicheren Grundsätzen hergeleitet und mit deutlichen Beyspielen erläutert*. Decker.

Lane, D. and Ashby, B. (1987). Psychlib: A library of machine language routines for controlling psychology experiments on the apple macintosh computer. *Behavior Research Methods*, 19(2):246–248.

Large, E. (2008). Resonating to musical rhythm: Theory and experiment. In Grondin, S., editor, *The psychology of time*, pages 189–231. Emerald Group Publishing.

Large, E. (2010). Neurodynamics of music. In Riess Jones, M., Fay, R. R., and Popper, A. N., editors, *Music perception*, pages 201–231. Springer.

Large, E. and Jones, M. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1):119–159.

Large, E. and Velasco, M. (In preparation). Nonlinear time-frequency transformation workbench: Matlab software for signal processing using nonlinear oscillators.

Levitin, D. J. and Tirovolas, A. K. (2009). Current advances in the cognitive neuroscience of music. *Annals of the New York Academy of Sciences*, 1156:211–31.

Lewis, P. A. and Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Current Opinion in Neurobiology*, 13(2):250–255.

Lewis, P. a. and Miall, R. C. (2006). Remembering the time: a continuous clock. *Trends in cognitive sciences*, 10(9):401–6.

Loehr, J. D., Large, E. W., and Palmer, C. (2011). Temporal coordination and adaptation to rate change in music performance. *Journal of experimental psychology. Human perception and performance*, 37(4):1292–309.

Löwgren, K., Bååth, R., Lindgren, M., Sahlén, B., and Hesslow, G. (2014). The cerebellum, timing and age: Eyeblink conditioning, isochronous interval tapping and prism adaptation in children and adults. Manuscript in preparation.

MacDougall, R. (1903). The structure of simple rhythm forms. *Psychological Monographs*.

MacMillan, K., Droettboom, M., and Fujinaga, I. (2001). Audio latency measurements of desktop operating systems. In *Proceedings of the International Computer Music Conference*, pages 259–262. sn.

Madison, G. (2001). Variability in isochronous tapping: Higher order dependencies as a function of intertap interval. *Journal of Experimental Psychology: Human Perception and Performance*, 27(2):411–422.

Madison, G. (2009). An auditory illusion of infinite tempo change based on multiple temporal levels. *PloS one*, 4(12):e8151.

Madison, G. and Paulin, J. (2010). Ratings of speed in real music as a function of both original and manipulated beat tempo. *The Journal of the Acoustical Society of America*, 128(5):3032–40.

Mates, J., Müller, U., Radil, T., and Pöppel, E. (1994). Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience*, 6(4):332–340.

McAuley, J. D. (2010). Tempo and rhythm. In Riess Jones, M., Fay, R. R., and Popper, A. N., editors, *Music Perception*, Springer Handbook of Auditory Research, pages 165–199. Springer.

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., and Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *Journal of experimental psychology. General*, 135(3):348–67.

Mellis, D. and Banzi, M. (2007). Arduino: An open electronic prototyping platform. *Conference on Human Factors in Computing Systems.*

Merchant, H. and Honing, H. (2013). Are non-human primates capable of rhythmic entrainment? evidence for the gradual audiomotor evolution hypothesis. *Frontiers in neuroscience*, 7:274.

Merker, B. H., Madison, G. S., and Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, 45(1):4–17.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63(2):81–97.

Miller, G. F. (2000). Evolution of human music through sexual selection. In Wallin, N. L., Merker, B., and Brown, S., editors, *The origins of music*, pages 329–360.

Miyake, I. (1902). Researches on rhythmic action. *Studies from Yale Psychological Laboratory*, 10:1–48.

Miyake, Y., Onishi, Y., and Pöppel, E. (2004). Two types of anticipation in synchronization tapping. *Acta neurobiologiae experimentalis*, 64(3):415–26.

Moelants, D. (2001). What is slow? timing strategies in the performance of feldman's" last pieces". In *Proceedings of the VII International Symposium on Systematic and Comparative Musicology - III International Conference on Cognitive Musicology*, pages 121–128, Jyväskylä.

Moelants, D. (2002). Preferred tempo reconsidered. In *Proceedings of the 7th international conference on music perception and cognition*, pages 580–583, Sydney.

Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1658).

Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience*, 31(28):10234–40.

Ortega, L., Guzman-Martinez, E., Grabowecky, M., and Suzuki, S. (2014). Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability. *Attention, perception & psychophysics*, 76(5):1485–502.

Osvath, M., Osvath, H., and Baath, R. (2014). An exploration of play behaviors in raven nestlings. *Animal Behavior and Cognition*, 2(2):157.

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, 24(1):99–104.

Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (2009). Avian and human movement to music. *Communicative and Integrative Biology*, 2(6):1–4.

Patel, A. D., Wong, M., Foxton, J., Lochy, A., and Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception*, 25(4):357–368.

Pearce, M. and Rohrmeier, M. (2012). Music cognition and the cognitive sciences. *Topics in cognitive science*, 4(4):468–84.

Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., and Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia*, 49(5):961–969.

Pinker, S. (1997). *How the Mind Works*. W. W. Norton & Company, New York.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in cognitive sciences*, 1(2):56–61.

Pöppel, E. (2004). Lost in time: A historical frame, elementary processing units and the 3-second window. In *Acta Neurobiologiae Experimentalis*, volume 64, pages 295–301.

Randel, D. (2003). The harvard dictionary of music.

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

Repp, B. H. (2006). Rate limits of sensorimotor synchronization. *Advances in Cognitive Psychology*, 2(2):163–181.

Repp, B. H. and Doggett, R. (2007). Tapping to a very slow beat: A comparison of musicians and nonmusicians. *Music Perception: An Interdisciplinary Journal*, 24(4):367–376.

Repp, B. H. and Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of experimental psychology. Human perception and performance*, 28(5):1085–1099.

Repp, B. H. and Su, Y.-H. (2013). Sensorimotor synchronization: a review of recent research (2006-2012). *Psychonomic bulletin & review*, 20(3):403–52.

Schaal, N. K., Banissy, M. J., and Lange, K. (2014). The rhythm span task: Comparing memory capacity for musical rhythms in musicians and non-musicians. *Journal of New Music Research*, 44(April 2015):3–10.

Schachner, A., Brady, T., Pepperberg, I., and Hauser, M. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, 19(10):831–836.

Schaefer, R. S., Vlek, R. J., and Desain, P. (2011). Decomposing rhythm processing: electroencephalography of perceived and self-imposed rhythmic patterns. *Psychological research*, 75(2):95–106.

Seddigh, A., Stenfors, C., Berntsson, E., Bååth, R., Sikström, S., and Westerlund, H. (2015). The association between office design and performance on demanding cognitive tasks. *Journal of Environmental Psychology*.

Semjen, a., Schulze, H. H., and Vorberg, D. (2000). Timing precision in continuation and synchronization tapping. *Psychological research*, 63(2):137–147.

Shams, L., Kamitani, Y., and Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, 408:788.

Stevens, C. J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Topics in cognitive science*, 4:653–667.

Stevens, L. T. (1886). On the time-sense. *Mind*, 11(43):393 – 404.

Temperley, N. (1963). Personal tempo and subjective accentuation. *The Journal of general psychology*, 68(2):267–287.

Toiviainen, P. and Snyder, J. (2003). Tapping to bach: Resonance-based modeling of pulse. *Music Perception*, 21(1):43–80.

van Noorden, L. and Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1):43–66.

Vos, P. G. (1973). *Waarneming van metrische toonreeksen*. Nijmegen: Stichting Studentenpers.

Vos, P. G. and Helsper, E. L. (1992). Tracking simple rhythms: On-beat versus off-beat performance. In *Time, Action and Cognition*, pages 287–299. Springer Netherlands, Dordrecht.

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

Wing, A. M., Endo, S., Bradbury, A., and Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. *Journal of the Royal Society, Interface / the Royal Society*, 11(93):20131125.

Woodrow, H. (1932). The effect of rate of sequence upon the accuracy of synchronization. *Journal of Experimental Psychology*, 15(4):357–379.

Wright, M., Cassidy, R., and Zbyszynski, M. (2004). Audio and gesture latency measurements on linux and osx. Technical report, http://cnmat.berkeley.edu/system/files/attachments/latencytest.pdf.

Yu, L. and Tomonaga, M. (2015). Interactional synchrony in chimpanzees: Examination through a finger-tapping experiment. *Scientific reports*, 5:10218.

Coda

If writing a thesis is a journey, my road has not been straight. The research plan I handed in as part of my PHD application had the lofty goal of implementing "a biologically valid model of learning, generating and categorizing actions". I worked on this for a year until I realized I didn't know what I was doing and gave up. As I've was a PHD-student within the Linnaeus research environment *Thinking in Time*, I thought I would do something related to time perception instead and so I started to work on what has now become this thesis. But I still had four years to go which, in my mind, was plenty of time, and so I diverged:

- I implemented a tool for exploring word use in Child language (http://www.childfreq.sumsar.net), the accompanying technical report (Bååth, 2010) is my most cited paper by far. "Candy" overcomes "Banana" in word-use frequency around an age of 30 months.

- I wrote a paper on visual illusions in cats (Bååth et al., 2014b). To be more specific, visual illusions in *my* cat.

- I presented my work on how to use eye movements to play rhythms at a Norwegian conference on novel musical instruments (Bååth et al., 2011b).

- I have invested a lot of time in the statistical programming language R. I have presented at the last three UseR! conferences and I have published in the R journal (Bååth, 2012). Nothing of this was directly related to my thesis, but it's the part of my PHD-studies that I am most proud over.

- I developed a Bayesian model for predicting the outcome of football matches (Bååth, 2015b). This model won the Data Analysis

Contest at the UseR! 2013 conference in Albacete, and won me 60 SEK on Sevilla vs. Valencia.

- I've been to several workshops together with Sverker Sikström and presented on the topic of the semantics of *Vagueness* (Bååth et al., 2011a). To be honest, I've never really understood the precise meaning of that term.

- Thanks to Sverker I have had a hand in studies on the semantic development in children (Submitted), on misleading argumentation in legal contexts (Dahlman et al., 2015), and on distracting factors in the work environment (Seddigh et al., 2015). One thing I learned is that Sverker Sikström is a distracting factor when you are trying to write a thesis. (A distracting factor who is always fun to talk to, and who has many distractingly interesting ideas, that is!)

- I started a surprisingly popular statistics blog over at `http://sumsar.net` which had more than 80,000 visitor last year. It is a sad fact, but true, that any one of my blog posts have had more readers than everything I have published combined.

- I have given the research service of the Swedish parliament (Riksdagens utredningstjänst) a crash course in Bayesian data analysis. And for the past two years Ullrika Sahlin and I have been the main organizers of the Bayes@Lund conference (`http://www.lucs.lu.se/bayes-at-lund-2015/`).

- I developed a new "Statistics 101" course from scratch, which I taught to the cognitive science students in 2014 as part of their master program. Unfortunately it had to be a course in frequentist statistics, which it confusing at best and intellectually harmful at worst. I might have caused them irreparable damage.

- I have been privileged to be able work with the Cognitive zoology group at LUCS and had a very small hand in a paper on play in raven chicks (Osvath et al., 2014) and in a paper on the lemonade preferences of orangutans (Submitted).

- And one some point I built a robot snake, unfortunately I don't remember why[3].

[3]You can still see it crawl on YouTube: `https://youtu.be/L9n3Yo39G-c` .

In retrospect, it is actually a small miracle that there was a thesis at all...

Paper I

Bååth, R. and Madison, G. (2012). The subjective difficulty of tapping to a slow beat. *In Proceedings of the 12th International Conference on Music Perception and Cognition*. Thessaloniki, Greece.

Note: This manuscript has been reformatted to better fit the paper format of the thesis. It also includes some minor corrections compared to the paper that was presented in Thessaloníki in 2012.

The Subjective Difficulty of Tapping to a Slow Beat

Rasmus Bååth,*[1] Guy Madison,#[2]

**Lund university Cognitive Science, Lund University, Sweden*

#Department of Psychology, Umeå university, Sweden

[1]rasmus.baath@lucs.lu.se, [2]guy.madison@psy.umu.se

Abstract

The current study investigates the slower limit of rhythm perception and participants' subjective difficulty when tapping to a slow beat. Thirty participants were asked to tap to metronome beats ranging in tempo from 600 ms to 3000 ms between each beat. After each tapping trial the participants rated the difficulty of keeping the beat on a seven point scale ranging from "very easy" to "very difficult". The result strongly support the notion that subjective difficulty increased with slower tempo as this was the case for all participants. While rated difficulty increased monotonically as a function of tempo the largest increase was between the tempo of 1200 ms and 1800 ms. This is in line with earlier reports on where tapping starts to feel laborious and supports the notion that there is a qualitative difference between tapping at fast (< 1200 ms between each beat) and slow (> 2400 ms between each beat) tempi. A mixed model analysis showed that tempo, tapping error and percentage of reactive responses all affected the participants rating of difficulty. Of these, tempo was by far the most influential factor, still participants were, to some degree, sensitive to their own tapping errors which then influenced their subsequent difficulty rating.

I. Introduction

Music come at a wide range of different tempi. John Coltrane's *Giant Steps* is an example of a tune that clocks in at the faster end of the spectrum with a tempo of 285 beats per minute (bpm). An example of a piece of music at the slower end of the spectrum would be Bach's *Air* from suite No. 3 in D major which is sometimes played at a tempo below 60 bpm. There are more extreme examples, for example, John Cage's *As Slow as Possible* has months between each new note. It is, however, rare for popular music to have a tempo slower than 1500 ms and faster than 300 ms between each beat, with tempi around 500 ms being the norm (van Noorden and Moelants, 1999). This can also be seen in the tempo ranges of metronomes which generally do not go slower than 1500 ms or faster than 300 ms between each beat.

It is reasonable to believe that these limits in tempo in some way reflect the limits of rhythm perception. Both the slower and the faster limit of rhythm perception has been studied using rhythm production tasks, especially finger tapping (Repp, 2005). The faster limit of rhythm perception has been assessed using tapping tasks where participants are asked to tap to successively faster metronome sequences. In order to not be limited by motoric factors when the tempo is fast only every second tone in the metronome beat is tapped to. Using this method trained musicians are able to synchronize to sequences with an inter stimuli interval (ISI) of close to 100 ms (Repp, 2007).

The slower limit of rhythm perception has been more difficult to assess as there seems to exist no (within reason) upper limit where tapping to a beat is no longer possible. When asked to freely tap a beat as slow as possible participants tend to tap at a tempo of around 2500 ms between each tap (McAuley et al., 2006). However, participants are able to tap at a much slower rates when paced by a metronome (Miyake et al., 2004). A common observation is that, as the tempo gets slower, there is an increase both in tapping variability and in the number of reactive responses. Here a *reactive response* refers to a response where the participant reacts to the sound rather than anticipates it (Repp and Doggett, 2007; Mates et al., 1994). Even though tapping variability increases with slower tempo there is at no point a sharp change in tapping variability. Nevertheless Repp (2006) argued for a slower limit around 1800 ms as it is around this tempo that participants start having difficulties anticipating the tones and reactive responses start to occur. He also noted that tapping is a rather effortless activity up to a tempo of 1500 ms, but when the tempo approaches 1800 ms it becomes a difficult task requiring attention and effort. This observation was not supported by any experimental data, however, and the present study aims to investigate the relation between tapping error and subjective ratings of difficulty when tapping to a slow metronome sequence.

The study had three aims: (1) To establish the relation between subjective difficulty and tempo. (2) To test the hypothesis that there is a qualitative difference between tapping at fast and slow tempi and that this is reflected by a steep shift in subjective difficulty around an ISI of 1800 ms. (3) To test if subjective difficulty depends on the tempo, the trial-to-trial performance of the participants or a combination of these factors. (1) and (2) is motivated by the observation by Repp (2006) described above. A participant's experience of difficulty when tapping could be caused by many factors, both factors that made the task more difficult, for example, a slow tempo, and factors that was the result of the high difficulty, such as a large tapping error or a high percentage of reactive responses. It might be the case that participants are sensitive to their own performance. For example, a participant might notice that he or she produced many reactive responses during a trial and as a result experience that trial as more difficult. On the other hand, participants might not be influenced by their own performance but solely by the difficulty of tapping at a slow tempo. The motivation for (3) is to pinpoint what factors influences subjective difficulty when tapping at a slow tempo.

II. Method

A. Participants

Nine female and 21 male participants, ranging in age from 19 to 78 years (M=31.6 , SD=12.8) were recruited from the Lund community. All were unpaid volunteers. All participants reported being right handed. Twenty-six participants reported that they had experience playing an instrument and ten participants reported having regularly played or practiced an instrument for more than ten years. All participants gave informed consent according to the guidelines of the Swedish Research Council.

B. Material

A custom build tapping board was used to record the onsets and velocities of the participants' finger taps. For a technical report describing the tapping board see Bååth (2011). The stimuli for the tapping task consisted of isochronous sequences of 440 Hz square wave tones of 20 ms. Each sequence consisted of 31 tones and were presented at five tempi, corresponding to tone ISIs of 600, 1200, 1800, 2400 and 3000 ms. The ISI of 600 ms can not be regarded a slow tempo but was included as a baseline, as participants tend to prefer tapping at an ISI of around 600 ms when being able to choose freely (McAuley et al., 2006). Both registration of taps and generation of sound was handled by an Arduino micro-controller, this was in order to avoid the timing uncertainties resulting from using a personal computer and to guarantee millisecond accuracy. The micro-controller was connected to a Dell Vostro 3700 computer that collected the timing information through a USB interface.

C. Procedure

During a session each participant performed a number of rhythm perception and production tasks, but only the data from the tapping and rating tasks are analyzed here. In the tapping task the participants sat in front of the tapping board wearing head phones. The task consisted of four blocks where each block contained five trials, one for each tone ISI. The order of the trials was randomized within each block. First the participants were asked to adjust the volume of the head phones to a comfortable level while a tone sequence was playing. After a short test trial the participants started with the first block. They tapped using the index finger on their dominant hand which was the right hand for all participants. There was a scheduled one minute break after the second block, otherwise successive trials were started as soon as the participant indicated that he or she was ready.

A trial consisted of the participants tapping to a tone sequence on the tapping board. The instructions given were to try to tap along the given tone sequence, to try to start tapping as soon as the sequence started and to stop when the sequence stopped. The participants were especially asked not to subdivide the beat in any way, for example by covert counting or by movement of the body. After finishing each sequence the participants rated the difficulty of tapping on a seven point scale ranging from “very easy” to “very difficult”. More specific, the participants were asked to rate “How difficult did you find it to keep the beat?” (translated from the Swedish “Hur svårt tyckte du det var att hålla takten?”).

D. Analysis

The first four taps of every trial were discarded in order to use only those taps where the participants had had some time to synchronize to the sequence. For each tone in the sequence the tap-to-tone asynchrony was calculated as the difference between the tone onset and the corresponding tap so that a negative asynchrony indicated that a tap preceded the tone. Sometimes participants tapped as a reaction to the tone instead of tapping with the tone. This was especially common at the slow tempi. For each trial the percentage of reactive responses was calculated where a response was labeled as reactive if the corresponding asynchrony was larger that 100 ms. Statistical analysis was conducted using the R statistical environment Team (2010). Mixed-effects regression modeling was done using the lme4 package (http://cran.r-project.org/web/packages/lme4/) with p-values calculated using the pval.fnc function from the languageR package (http://cran.r-project.org/web/packages/languageR/) (Baayen et al., 2008).

III. Result

The participants generally used the whole rating scale and as expected there was a strong, statistically significant correlation between tempo and the mean rated difficulty for each participant (Pearson's product moment correlation, $r=0.89$, $n=150$, $p < 0.001$). Figure 1 show the distributions of difficulty ratings at the different tempi. The smallest increase in rated difficulty was between tempi of 600 ms and 1200 ms (M = 0.5) which was significantly smaller than the differences between tempi of 1200 ms and 1800 ms (paired t-test, $t(29)= -5.42$, $p < 0.001$), 1800 ms and 2400 ms ($t(29) = -5.63$, $p < 0.001$), and 2400 ms and 3000 ms ($t(29) = 2.69$, $p = 0.012$). The largest increase in rated difficulty was between tempi of 1200 ms and 1800 ms (M=1.6) which was significantly larger than the difference between tempi of 600 ms and 1200 ms ($t(29)= 5.42$, $p < 0.001$) and 2400 ms and 3000 ms ($t(29) = 3.47$, $p = 0.002$). While the difference in rating between tempi of 1200 ms and 1800 ms was larger than the difference between tempi of 1800 ms and 2400 ms it was not statistically significant ($t(29) = 1.66$, $p = 0.11$).

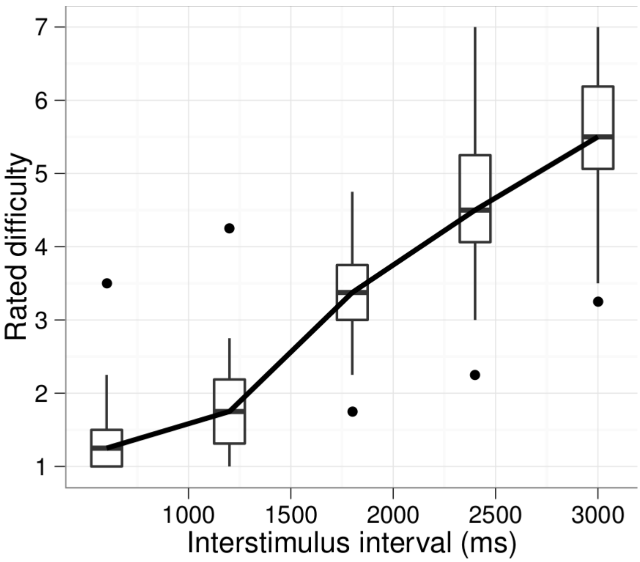


Figure 1. The distributions of difficulty ratings at the different tempi. The line connects the median ratings.

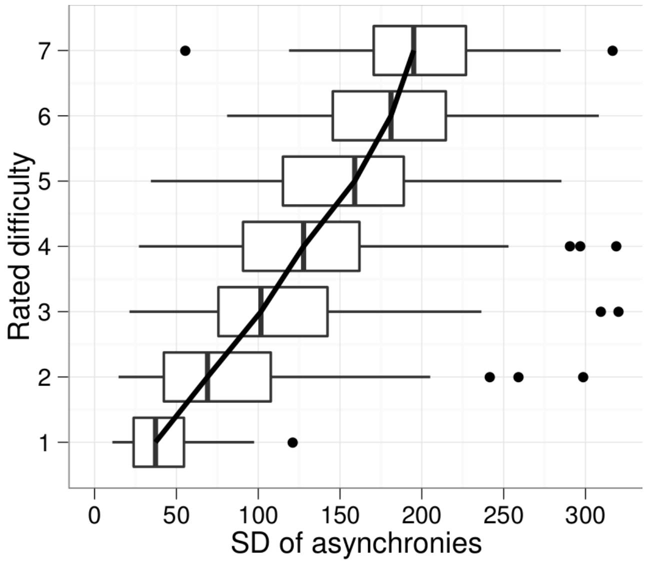


Figure 2. The distributions of tapping error as a function of rated difficulty. The line connects the median ratings.

Tapping error, as measured by the standard deviation of the tap-to-tone asynchronies, was correlated with rated difficulty ($r=0.79$, $n=150$, $p < 0.001$). The increase in rated difficulty as a function of tapping error is also visible in figure 2. This result is hard to interpret, however, as tapping error is also known to increase linearly with tempo. As expected tapping error increase with larger ISIs (as shown in figure 3) and there was a positive correlation between tapping error and tempo ($r=0.90$, $n=150$, $p < 0.001$). There was a positive correlation between rated difficulty and percentage of reactive responses ($r=0.68$, $n=150$, $p<0.001$) but the number of reactive responses also increased with slower tempo (see figure 4).

A number of linear mixed-effects models were fitted to asses the influence of tempo, tapping error and percentage of reactive responses on rated difficulty. The models were fitted on the per-trial data, not data averaged over trials, and all models included participant as a random effect. As tempo had the highest correlation with rated difficulty a first model only included tempo as predictor. A second model also included tapping error and percentage and a likelihood ratio test showed that it was justified to include those terms (chi-square = 37.8, $p < 0.001$). A third model added a term for tapping error relative to tempo, that is, the standard deviation of the asynchronies divided by ISI. This was the final model and the addition of the relative tapping error term was justified according to a likelihood ratio test (chi-square=5.99, $p=0.014$). All slopes in the final model deviated significantly from zero except for the slope for the tapping error, probably due to the inclusion of the relative tapping error term. The final model is summarized in table 1.

Predictor	B	β	p
ISI	0.0015	0.65	< 0.001
SD(asynchrony)/ISI	9.36	0.13	0.014
% reactive responses	1.76	0.11	< 0.001
SD(asynchrony)	0.0002	0.01	0.90

Table 1. A summary of the linear mixed-effects model predicting rated difficulty. Collumn B show the raw slopes of the predictors while collumn β show the standardized slopes.

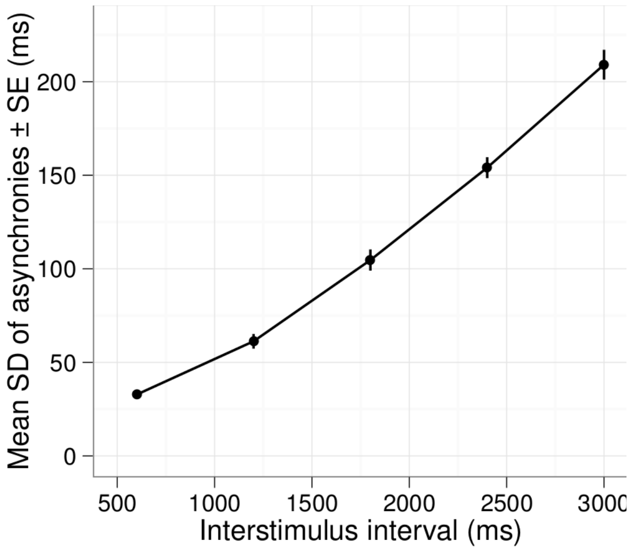


Figure 3. The tapping error as a function of tempo.

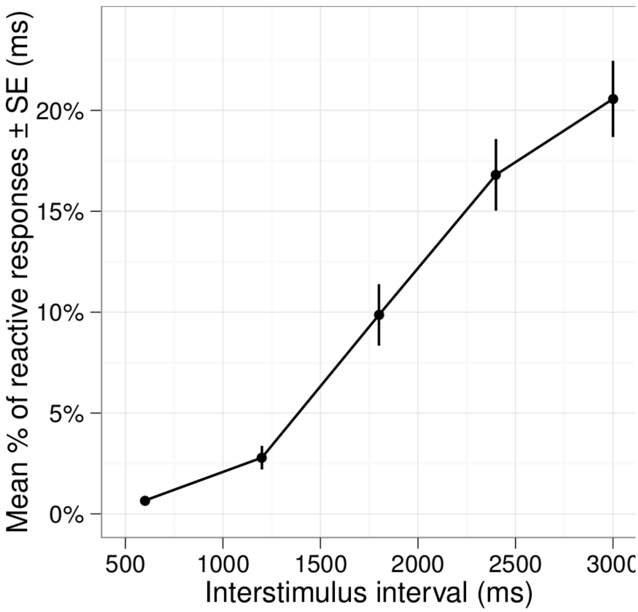


Figure 4. The percentage of reactive responses as a function of tempo.

IV. Discussion

The result strongly support the notion that subjective difficulty increases with slower tempo as this was the case for all participants. While difficulty increased monotonically as a function of tempo the largest increase was between the tempi of 1200 ms and 1800 ms. This agrees with Repp's (2006) description of a subjective slower limit were rhythm production goes from being effortless to being cognitively demanding. After having finished the session many of the participants also expressed that tapping to the slow tempi felt very taxing and that there was a great contrast between tapping at the slow tempi and at the fastest tempo. The mixed model analysis showed that tempo, tapping error and percentage of reactive responses all affected the participants rating of difficulty. Of these, tempo was by far the most influential factor as the standardized slopes in table 1 show. Still participants are, to some degree, sensitive to their own tapping errors which then influences their subsequent difficulty rating.

In this study there was relatively few ISIs levels distributed over a quite wide range. In a future study it would be interesting to narrow down the range to around 600 to 2000 ms and try to pinpoint where subjective difficulty increases the most. Another question remains: Why is there an upper limit of rhythm perception at all? This is hard to answer without addressing the larger question: What is the neural mechanism behind rhythm perception? A promising framework for explaining this mechanism is the resonance theory of rhythm perception and production which postulates that rhythm is coded as a multifrequency pattern of oscillating neural circuits (Large, 2008). The oscillating circuits can only code for rhythms that are as slow as the slowest circuit which would then explain the existence of a slower limit of rhythm perception.

Acknowledgement

The first author gratefully acknowledges support from the Linnaeus environment Thinking in Time: Cognition, Communication and Learning (CCL), Swedish Research Council grant number 349- 2007-8695.

References

Bååth, R. (2011). Construction of a Low Latency Tapping Board. *LUCS minor*, 17.

Baayen, R., Davidson, D., and Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.

Large, E. (2008). Resonating to musical rhythm: Theory and experiment. *The psychology of time*, pages 189–231.

Mates, J., Müller, U., Radil, T., and Pöppel, E. (1994). Temporal Integration in Sensorimotor Synchronization. *Journal of Cognitive Neuroscience*, 6(4):332–340.

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., and Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *Journal of experimental psychology. General*, 135(3):348–67.

Miyake, Y., Onishi, Y., and Pöppel, E. (2004). Two types of anticipation in synchronization tapping. *Acta neurobiologiae experimentalis*, 64(3):415–26.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Repp, B. (2007). Perceiving the numerosity of rapidly occurring auditory events in metrical and nonmetrical contexts. *Attention, Perception, & Psychophysics*, 69(4):529–543.

Repp, B. and Doggett, R. (2007). Tapping to a very slow beat: A comparison of musicians and nonmusicians. *Music Perception*, 24(4):367–376.

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

Repp, B. H. (2006). Rate limits of sensorimotor synchronization. *Advances in Cognitive Psychology*, 2(2):163–181.

van Noorden, L. and Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1):43–66.

Paper II

Bååth, R., Lagerstedt, E., and Gärdenfors, P. (2014). A prototype-based resonance model of rhythm categorization. *i-Perception* 5(6) 548–558; doi:10.1068/i0665

i-PERCEPTION

a Pion publication

i-Perception (2014) volume 5, pages 548–558

dx.doi.org/10.1068/i0665 ISSN 2041-6695 perceptionweb.com/i-perception

A prototype-based resonance model of rhythm categorization

Rasmus Bååth
Lund University Cognitive Science, Lund University, LUX, Lund, Sweden; e-mail: rasmus.baath@lucs.lu.se

Erik Lagerstedt
Lund University Cognitive Science, Lund University, LUX, Lund, Sweden; e-mail: lagerstedt.erik@gmail.com

Peter Gärdenfors
Lund University Cognitive Science, Lund University, LUX, Lund, Sweden; e-mail: peter.gardenfors@lucs.lu.se

Received 25 June 2014, in revised form 1 October 2014; published 4 November 2014

Abstract. Categorization of rhythmic patterns is prevalent in musical practice, an example of this being the transcription of (possibly not strictly metrical) music into musical notation. In this article we implement a dynamical systems' model of rhythm categorization based on the resonance theory of rhythm perception developed by Large (2010). This model is used to simulate the categorical choices of participants in two experiments of Desain and Honing (2003). The model accurately replicates the experimental data. Our results support resonance theory as a viable model of rhythm perception and show that by viewing rhythm perception as a dynamical system it is possible to model central properties of rhythm categorization.

Keywords: categorical perception, rhythm perception, computational modeling, music perception, dynamical systems, resonance theory.

1 Introduction

Music is a nondiscrete art form since it exists in the auditory domain where differences in rhythm, pitch, and timbre are continuous. However, when this continuous domain is described, discrete categories are often used, e.g., pitch is categorized as scale notes. Rhythm is also the subject of categorization, and an example of this is the transcription of the rhythm of a piece of music into musical notation. This process constitutes a categorization, as the number of ways to notate a rhythmic sequence is finite and different rhythmic sequences can be notated in the same way. Desain and Honing (2003) showed in two experiments that listeners reliably experienced rhythms as belonging to rhythmic categories and that categorizations were strongly influenced when the listeners were primed with a metric beat before hearing a rhythm. Furthermore, participants agreed to a large degree on which rhythms belonged to what category. Just as with categorization of other kinds of stimuli (c.f. Jäger, 2010, on color categories), the categories were roughly convex with respect to a temporal performance space as discussed in Section 3 (Gärdenfors, 2000). Honing (2013, p. 379) concludes that: "It is puzzling, however, that although meter was shown to exert a strong influence on the recognition of rhythm […] existing computational models of meter can explain this phenomenon only to a small extent."

In this article we show that an oscillation-based, *resonance theory* model of rhythm perception (Large, 1996, 2010) can replicate many of the findings of Desain and Honing (2003). Although oscillator models have previously been applied to many different aspects of music perception (e.g., Angelis, Holland, Upton, & Clayton, 2013; Large, 1996, 2010), such models have not previously, to our knowledge, been applied to categorical perception. Our results support the notion that resonance theory is a viable model of rhythm perception and show that by viewing rhythm perception as a dynamical system it is indeed possible to model the properties of categorical rhythm perception. Furthermore, these results suggest that oscillator models can be applied to other types of categorical perception, for example, pitch perception and vowel perception.

Resonance Theory and Rhythm Perception

In the field of music perception, *rhythm* refers to a temporal pattern of sound onsets (McAuley, 2010). A rhythm in this sense does not have to be periodic or recurrent. This is in contrast with how that word is used in other fields (cf. circadian rhythm or delta rhythm). A related concept that does involve periodicity is *beat*. When listening to a piece of music, a common response is to move one's body with a perceived periodic pulse (Snyder & Krumhansl, 2001), that pulse being the beat of the corresponding piece of music. It is common that not all beats in a piece of music are perceived as being equally accented (Palmer & Krumhansl, 1987) and a periodically recurring pattern of strong and weak beats is called a *meter*. For example, a duple meter would imply that every second beat is perceived as having a stronger accent while every third beat is perceived as having a stronger accent in the case of a triple meter. Rhythm perception and the ability to entrain to a musical beat was long thought to be uniquely human and, while it has recently been shown that some vocal mimicking species are, to some degree, able to move in synchrony with a beat (Schachner, Brady, Pepperberg, & Hauser, 2009), humans are still unique in their aptitude for rhythmic processing. Already infants have been shown to have a sense of rhythm (Honing, Ladinig, Háden, & Winkler, 2009) and there exists only one documented case of "beat deafness" (Phillips-Silver et al., 2011), that is, the inability to reliably synchronize to a musical beat.

Modeling of human timing and rhythm perception has a long history, one influential model being that described by Wing and Kristofferson (1973), which is based on an information theoretic perspective. Like many such models (cf. Repp, 2005), it models a participant's behavior in situations where isochronous timing responses are being elicited. An alternative to the information theoretic perspective is to take a dynamical systems perspective and model time and rhythm perception as an emergent, dynamic phenomenon. A number of models of this kind have been proposed (e.g., Large, 1996; Todd, O'Boyle, & Lee, 1999; van Noorden & Moelants, 1999). Here, the term *resonance theory* (cf. Large, 2010) will be used to refer to such models. The general idea of resonance theory is that an external auditory rhythm can be represented by the amplitude of internal oscillatory units. These oscillatory units are coupled to the external rhythm and are by definition periodic while the external rhythm does not have to be periodic. Given a rhythm sequence as input, the basic output of a resonance theory model, or resonance model for short, is the amplitude response of the oscillators over time. Resonance theory does not dictate a specific model but rather incorporates a number of related models which all can be considered dynamical system models.

Resonance theory provides a compelling framework since it is biologically plausible, has a solid base in dynamical systems theory and is able to model many aspects of meter and rhythm perception (Angelis et al., 2013; Large, 2000). A number of neuroimaging studies have shown connections between neural resonance and rhythm perception (e.g., Brochard, Abecasis, Potter, Ragot, & Drake, 2003; Fujioka, Trainor, Large, & Ross, 2012; Schaefer, Vlek, & Desain, 2011). One persuasive study that clearly showed that rhythm perception involves neural oscillatory activity is that of Nozaradan, Peretz, Missal, and Mouraux (2011). They found that playing a rhythmic beat to a participant elicited a sustained periodic neural response, as measured by EEG, that matched the frequencies of the beat.

As already noted, to our knowledge, resonance theory models have not previously been used to model categorical rhythm perception. One reason for this might be that while the amplitude response of the oscillators to a rhythm in the resonance model reflects, perhaps even represents, the rhythm sequence given as input to the system, it does not give rise to a categorization per se. That is, the state of the resonance model depends on the given rhythm sequence, but there is no single well defined set of states that can be said to constitute categories. Still, the state of the resonance model can be used as the basis of a categorical decision based on learned prototype states or a discrete partitioning of the system state space.

If the state of the resonance model is viewed as the basis for a categorical decision then two predictions regarding categorical rhythm perception can be made:

A More distinct states will facilitate categorization. Here we use *distinct state* to refer to a state of the resonance model where a small subset of oscillators has a strong amplitude response while most oscillators do not. This is in contrast to a nondistinct state where most oscillators have a similar amplitude response, that is, there are many competing signals and there is no clear single winning candidate among the categories. If some rhythmic sequence was known to result in a strongly distinctive state in a resonance model, then it could be predicted that a participant in an experimental categorization task would categorize that rhythm sequence consistently, and with more confidence than a rhythmic sequence known to result in a less distinctive state.

B Rhythm sequences resulting in similar states are categorized similarly. That is, different rhythm sequences resulting in similar states when used as the input to a resonance model should be categorized similarly by participants in an experimental task. Here similarity has to be defined using a similarity measure such as Euclidean distance or cosine similarity.

In order to test these predictions, data from the rhythm categorization task from the study by Desain and Honing (2003) were used.

The Rhythm Categorization Study of Desain and Honing (2003)

Desain and Honing (2003) employed a novel paradigm where musically educated participants were asked to categorize 66 different rhythm sequences by transcribing them into common music notation. The sequences all lasted for one second and consisted of four tone onsets and were therefore uniquely determined by the three interonset intervals (IOI) between the tones. Two such possible sequences are shown in Figure 1a and 1b where a possible categorization of 1b could be ♪♪ ♩ (or 1-1-2 when written as an integer ratio). Any possible 1 s, four-tone rhythm sequence can be thought of as a point in a three dimensional triangular *performance space* that determines the lengths of the three IOIs as shown in Figure 1. The 66 rhythm sequences in Desain and Honing's experiment were constructed so that they evenly covered the area in the performance space with the constraint that no IOI would be shorter than 153 ms. The location of these sequences in the performance space can be seen in Figure 2b where each circle marks the position of one of the 66 sequences.

In a first experiment, 29 highly trained musicians categorized the rhythm sequences and the result was that even though the rhythms occurred on a more or less continuous time scale, the participants

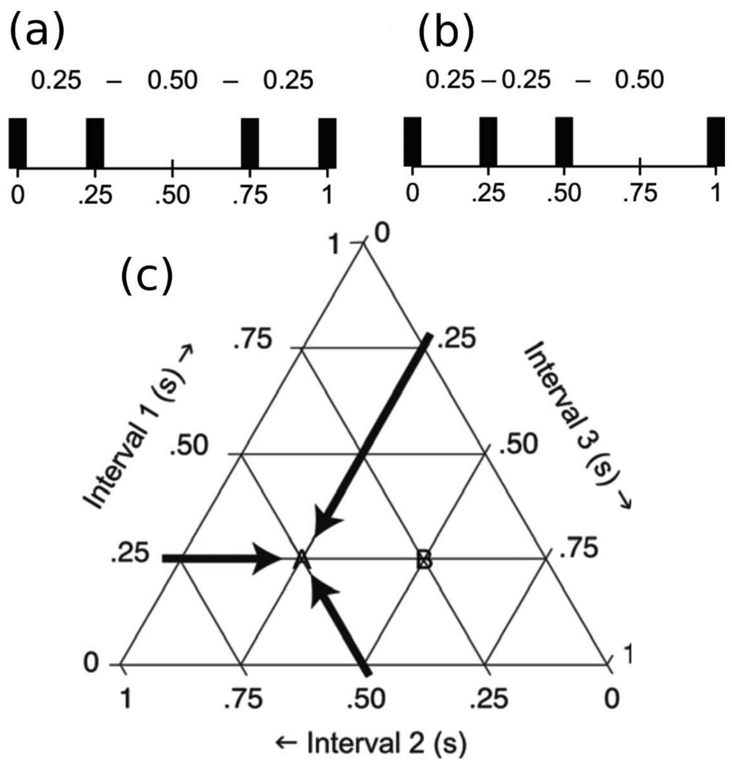


Figure 1. (a) and (b) show examples of two possible rhythms and their placement in the triangular performance space (c) defined by Desain and Honing (2003). All one second long, four sound rhythms can be represented as a point in this space. (From Honing, 2012 with permission).

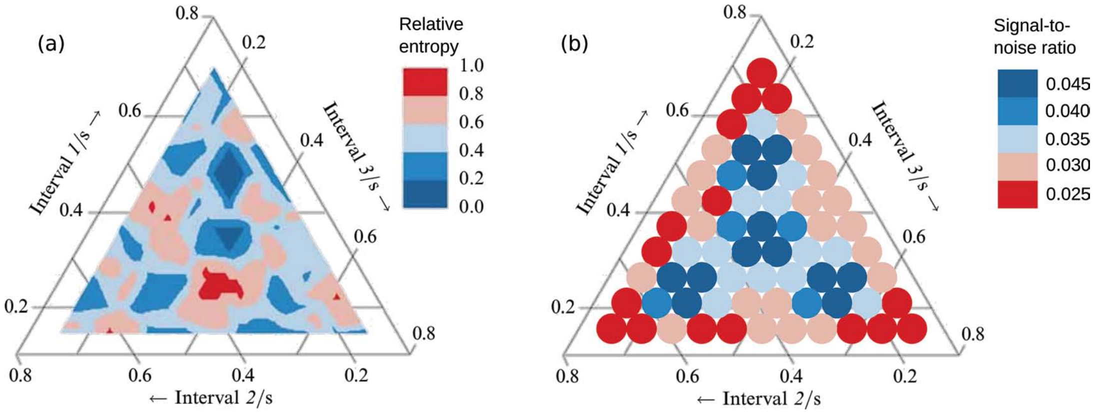


Figure 2. Maps over categorization consistency. (a) shows the relative entropy of the categorical choices for the single participant given the same rhythm sequences multiple times from Desain and Honing (2003, used with permission). The relative entropy is calculated as the Shannon entropy divided by the theoretical maximum attainable entropy. (b) shows the signal-to-noise measure calculated from the activation patterns generated by the resonance model.

tended to stick to a limited number of categories, with 1-1-1 being the most common. Twelve categories, all categories considered, stood out as being the most common and the location in performance space of these categories are shown in Figure 3a. One participant was presented with all 66 rhythm sequences at six different occasions and, as a measure of consistency, the entropy was calculated of her responses for each rhythm. These entropy values were mapped on to the performance space and the resulting entropy map is shown in Figure 2a.

In a second experiment, two metrical conditions were added. Duple meter versions of the rhythms were constructed by adding a repeated, 1-s long, two sound beats to the beginning of the rhythms, thus inducing a 2/4 meter feeling. Triple meter versions of the rhythms were similarly constructed by adding a three sound beat instead. This resulted in three different metrical conditions: The original no meter condition, a duple meter condition, and a triple meter condition. Maps over what categories the participants ascribed to the different rhythms, similar to the map shown in Figure 3a, were constructed (shown on p. 358 in Desain & Honing, 2003). A main finding was that the participants' categorization in the no meter condition was significantly more similar to the participants' categorization in the duple meter condition than in the triple meter condition.

For the purpose of the current study, data from Desain and Honing were downloaded from a web resource containing supplementary material (http://www.mcg.uva.nl/categorization). The data downloaded were the information regarding which of the 12 most common categories were most often ascribed to each of the 66 rhythm sequences for the no meter condition in experiment 1, and the duple

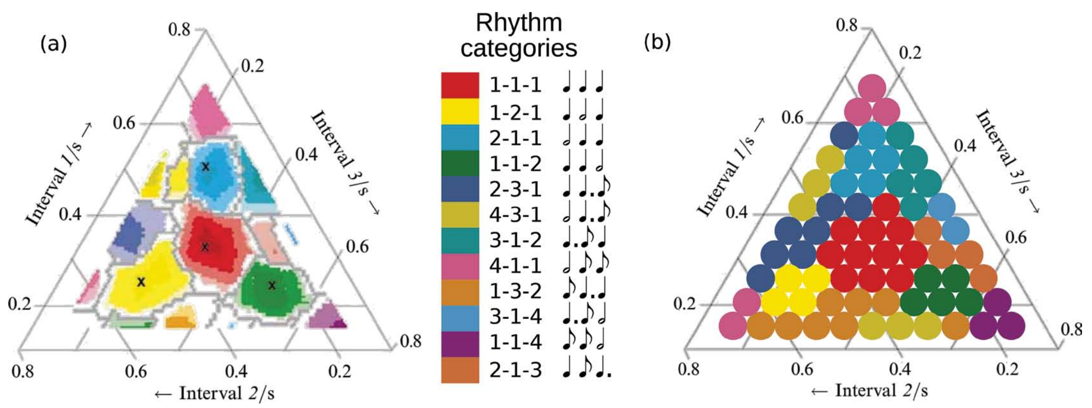


Figure 3. Categorization maps for (a) the experimental data from Desain and Honing (2003, used with permission) and (b) the resonance model. The transparent areas in (a) indicate areas where there was a large amount of disagreement between the participants.

and triple meter conditions in experiment 2. A rhythm sequence was excluded if none of the 12 most common categories was the most common for that specific rhythm. Information regarding the categorization entropy for the sole participant presented with the rhythms multiple times was calculated manually from Figure 2a.

Resonance Theory and Rhythm Categorization

It is possible to test the two predictions from resonance theory concerning how rhythms are categorized by implementing a resonance model that consists of an array of oscillators (as in Large, 2000); see Section 2 for more details. We used the rhythm stimuli from Desain and Honing (2003) as input to such a model and compared the results with the experimental data using the methods outlined below. The output of a resonance model is a multidimensional time series with the same number of dimensions as the number of oscillators in the model. This high dimensional representation might be difficult to work with directly, however, and a more convenient representation is given by creating an *activation pattern*, A, by summing the amplitude responses of each oscillator over time, as in

$$A_i = \sum_{t=t_s}^{t_e} a_{i,t}, \tag{1}$$

where $ai,_t$ is the amplitude for oscillator i at time t while t_s and t_e are the start and end time steps for the summation. Before the resonance model is given any input it is in a resting state and it takes a number of time steps before the system is activated by the stimuli. Therefore, it is not necessarily desirable to sum over the whole extent of the duration of the rhythm sequence and an activation pattern created by summing over the later time steps may represent the rhythm sequence better than an activation pattern created by summing over all time steps. By considering the activation pattern of a resonance model as a point in an n-dimensional space, n being the number of oscillatory units, this space can be partitioned into regions corresponding to rhythm categories and used to produce categorical decisions (following the general model of concepts from Gärdenfors, 2000). Given that entire equivalence classes of different spectra can give rise to the same color perception, the potential relation between the activation pattern of a resonance model and such a rhythm categorization is analogous to the relation between the hue, saturation, and lightness of a color percept and a color categorization. That is, a color percept can similarly be viewed as a point in a three-dimensional space with dimensions hue, saturation, and lightness and this space can be partitioned into regions, each representing a color category, and used to produce categorical color decisions.

Prediction (A) implies that rhythm sequences resulting in distinctive states (in the sense discussed earlier) in a resonance model should be the sequences that are categorized more consistently. In Desain and Honing's data, a measure of consistency is the categorization entropy for the participant presented with the rhythm sequences multiple times. The prediction is that this measure of consistency is correlated with a measure of distinctness of the state of a resonance model. Signal-to-noise ratio is a common measure of distinctness of a signal and a modified version of this measure can be used to quantify the distinctness of the state of a resonance model. For a resonance model that has been given a rhythm sequence as input, the activation pattern is first calculated according to Equation (1). In this activation pattern, the signal A_s is defined as being the A_i with the highest amplitude. The signal-to-noise ratio is then defined as:

$$\mathrm{SNR} = \frac{A_s}{\sum_{i=1}^{n} A_i} \qquad i \neq s, \tag{2}$$

where the sum in the denominator is over the rest of the A_i oscillator amplitudes. Notice that this measure of consistency should be negatively correlated with the entropy measure of Desain and Honing: As the signal gets weaker relative to the noise, the entropy of the participants' choices of category should go up.

Prediction (B) implies that rhythm sequences resulting in similar states when given as input to a resonance model should be categorized similarly in an experimental task. A resonance model does not directly produce a categorization, but this is not required for testing this prediction. It is possible to compare the resulting states of two rhythm sequences by calculating the respective activation patterns and comparing these. A suitable similarity measure is given by considering the activation patterns as points in an n-dimensional space, where n is the number of oscillators in the resonance model, and

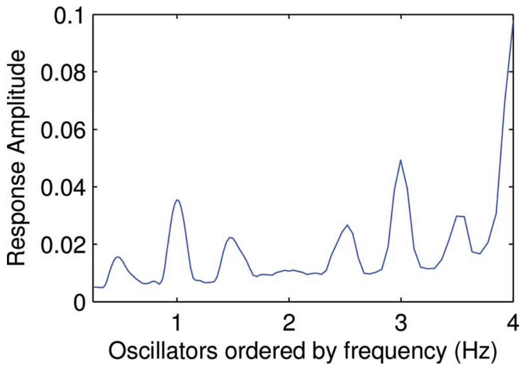


Figure 4. An example of an activation pattern generated by feeding the resonance model a rhythm with IOIs 0.25 s, 0.5 s, and 0.25 s.

then taking the Euclidean distance between these two points, where a shorter distance corresponds to more similar states. Considering the twelve most common rhythm categories chosen by the participants in Desain and Honing's study as prototype categories, it is possible to use the rhythm sequences corresponding to these categories to generate *prototype activation patterns*. For example, to generate the prototype activation pattern for the category 1-2-1 (as shown in Figure 4) the rhythm sequence with IOIs 0.25 s, 0.5 s, and 0.25 s would be used as input to the resonance model. A rhythm sequence's activation pattern can then be compared with these prototypes' activation patterns and the prototype category with the most similar activation pattern can be assigned to that rhythm sequence. In this way, all rhythm sequences can be assigned a category and these categories can be compared with the categories selected by the participants in Desain and Honing's study. Specific hypotheses are then that a resonance model categorization of the stimulus used by Desain and Honing should be similar to the categorization made by the participants in the no meter, duple meter, and triple meter conditions. Furthermore, as the participants' categorizations of the rhythm sequences in the no meter condition were more similar to the categorizations made in the duple meter condition than to the categorizations made in the triple meter condition, the same relation should be present in the categories generated by the resonance model.

2 Methods

The resonance model was implemented in MATLAB (http://www.mathworks.se/products/matlab/) using the Nonlinear Time-Frequency Transformation Workbench (Large & Velasco, in preparation). The model consisted of 145 *Hopf oscillators*, a type of oscillator that entrains to periodic input and where the amplitude of an oscillator depends on that oscillator's intrinsic frequency and the periodicities of the input. The differential equation of the Hopf oscillator used is:

$$\frac{dz}{dt} = z\left(\alpha + i\omega + \frac{\beta\varepsilon|z|^4}{1-\varepsilon|z|^2}\right) + \frac{x}{1-\sqrt{\varepsilon}x} \cdot \frac{1}{1-\sqrt{\varepsilon}\bar{z}} \tag{3}$$

$$\alpha = -0.1, \beta = -0.1, \varepsilon = 0.5$$

where α is a damping term, β is an amplitude compression factor and ε is a scale factor. The last term in Equation (3) is the resonant term, which is dependent on the stimulus x. These parameter values and this specific formulation of the Hopf oscillator were not chosen on the basis of any specific theoretical considerations (see Section 4 for a discussion of these choices); many other configurations are possible and a more general form of the Hopf oscillator is derived in Large, Almonte, and Velasco (2010). The oscillators' intrinsic frequencies were centered around 1 Hz with frequencies logarithmically distributed from 0.25 Hz to 4 Hz. Figure 5 shows an example of the activation over time for this network of oscillators given the rhythm pattern [0.5, 0.375, 0.125]. The method used for creating activation patterns was that in Equation (1) with t_s set to the time step corresponding to half the stimulus length and t_e set to the last time step. The MATLAB code for the model and both input data and the resulting output are available on request from the first author. The code for the Nonlinear Time-Frequency

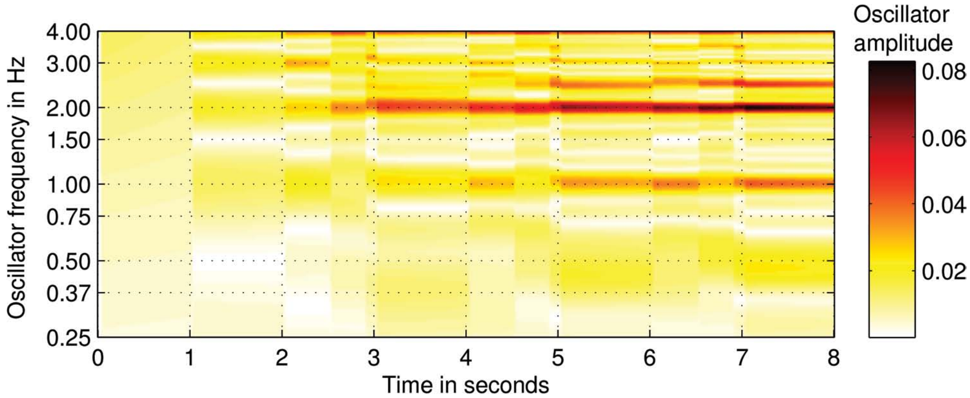


Figure 5. An example of oscillator activation over time for an oscillator network given the rhythm pattern [0.5, 0.375, 0.125].

Transformation Workbench (Large & Velasco, in preparation) has not yet been publicly released and has to be requested separately.

The 66 rhythm sequences from the no meter condition were encoded and given as input to the model yielding 66 activation patterns. In accordance with Desain and Honing (2003), the repeated rhythm pattern in each sequence was 1-s long and repeated three times, creating an eight bar long sequence of where the third, fifth, and seventh bar contained the rhythm pattern. The IOIs in the sequences range from 3/19 s to 13/19 s in steps of 1/19 s, creating the grid seen in Figures 2 and 3. This was repeated for the sequences from the duple and triple meter conditions. Additionally, the sequences of the prototype categories were encoded in the same way as in the no meter condition sequences, yielding 12 prototype activation patterns.

Analysis

The signal-to-noise ratio, defined by Equation 2, was calculated for each of the 66 rhythm sequences in the no meter condition. Here a higher value represents a stronger signal and the signal-to-noise ratio is taken as a measure of how easy it is to classify the corresponding sequence. For each of all 198 rhythm sequences, the Euclidean distance to each of the twelve prototype sequences was calculated. A sequence is categorized as the prototype it is closest to in the Euclidean space, with each dimension representing one oscillator in the oscillatory network. The resulting categories for the no meter condition can be seen in Figure 3.

Randomized permutation tests (Ernst, 2004) were used to compare the categorization of the rhythm sequences from the behavioral data with the categorization from the resonance model. Given two different categorizations of the 66 rhythms, a similarity score is calculated as the number of rhythms that are given the same category by both categorizations. In the cases where the most common categorization of a specific rhythm sequence in the behavioral data is not one of the 12 prototype categories this rhythm sequence is excluded from further analysis. Next, all category labels are randomly reassigned to the rhythm sequences and a new similarity score is calculated. This is repeated 10,000 times, yielding a randomized permutation distribution of similarity scores. This is the distribution that is expected under the null hypothesis that there is no relation between the categorization by the model and the categorization by the participants. A p-value is then calculated as the probability of achieving the actual similarity score, or a more extreme similarity score, given the distribution of randomized similarity scores. The permutation tests were two-tailed (calculated according to the method in Ernst, 2004) in all cases except where noted.

3 Results

The signal-to-noise measure was calculated for all activation patterns in the no meter condition and, as hypothesized, a negative correlation between Desain and Honing's (2003) entropy measure of consistency and the signal-to-noise ratio was found (Pearson product-moment correlation, $r = -0.32$, $p = 0.009$). These two measures of consistency are expected to have an inverse relationship, that is, low entropy in the experimental data indicates high consistency, while a low signal-to-noise ratio in the simulated data indicates low consistency. A comparison between these two measures of consistency for the experimental and the simulated data is shown in Figure 2. To facilitate comparison, the color scales have been matched so that red indicates low consistency while blue indicates high consistency. The measures of consistency are comparable, showing the same broad patterns in both the simulated (Figure 2b) and experimental data (Figure 2a).

The activation patterns for all of the three metrical conditions were compared with the prototype activation patterns using the Euclidean distance as the similarity measure and each rhythm sequence was assigned the category of the most similar prototype. A comparison with the categories assigned in the experimental task for the no meter condition is shown in Figure 3. The categorizations agree to a large extent. The 1-1-1 category is the most common in both the experimental and the simulated categorizations and both categorizations exhibit roughly convex category regions. Here convexity refers to the property that for every pair of points within a geometric object there exists a line, also within the object, that connects the points. The convexity of the category regions accords with Gärdenfors' (2000) general prediction for category representations. A randomized permutation test also showed that the categorization generated by the resonance model and the categorization from Desain and Honing's data were more similar than would be expected by chance alone for all three of the metrical conditions. In the no meter condition (shown in Figure 3) the agreement was 71% ($p < 0.001$) and in the duple and triple meter conditions 67% ($p < 0.001$) and 61% ($p < 0.001$), respectively.

Figure 6 shows the oscillator activation over time and the corresponding dynamic categorization given a rhythm sequence that was assigned low entropy in Desain and Honing's data. Compare this with Figure 7 that shows the oscillator activation and dynamic categorization for a rhythm that was assigned high entropy in Desain and Honing's data. For the low entropy rhythm the signal-to-noise ratio is high and the categorization is more stable. For the high entropy rhythm, however, the signal-to-noise ratio is low and the categorization in never stable, that is, there never emerges one clear winner.

In the experimental data, the categorization of the duple meter condition was more similar to the no meter condition than was the triple meter condition, and this was also the case for the simulated

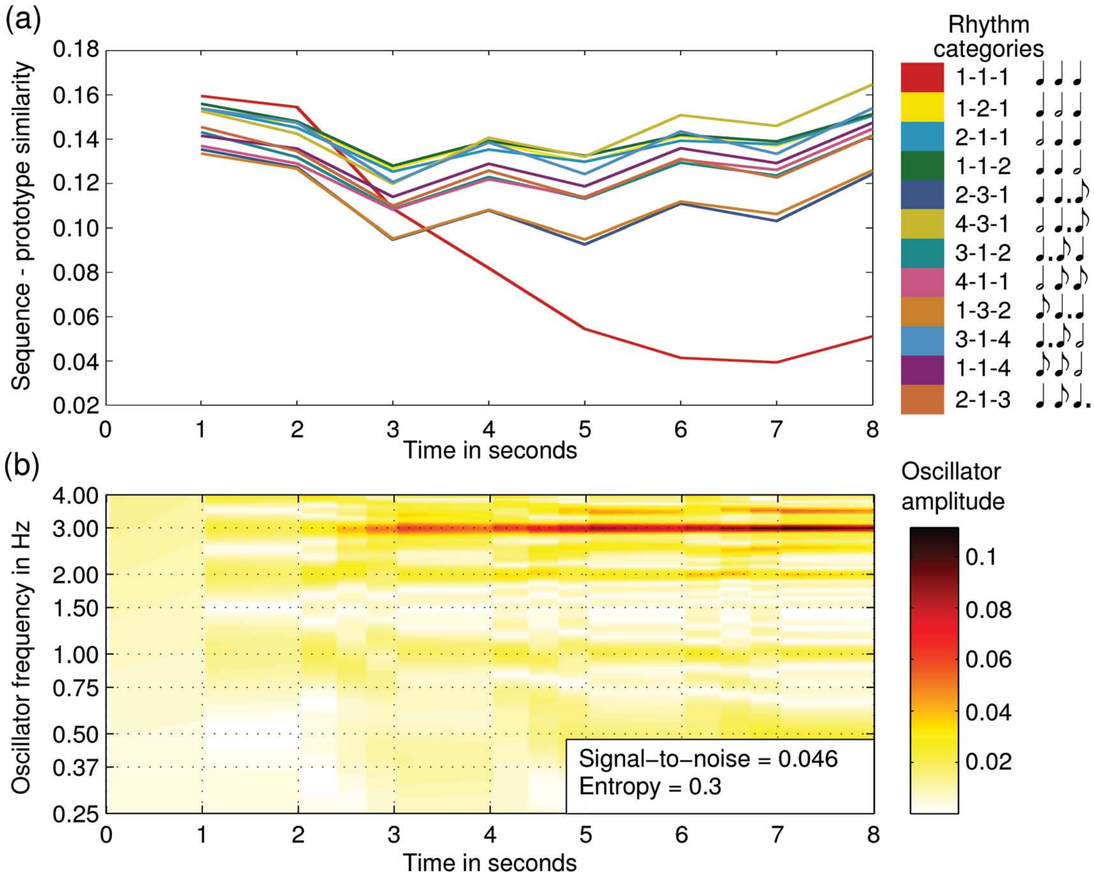


Figure 6. The oscillator activation and corresponding categorization over time for an oscillator network given a rhythm pattern that scored low entropy in Desain and Honing (2003).

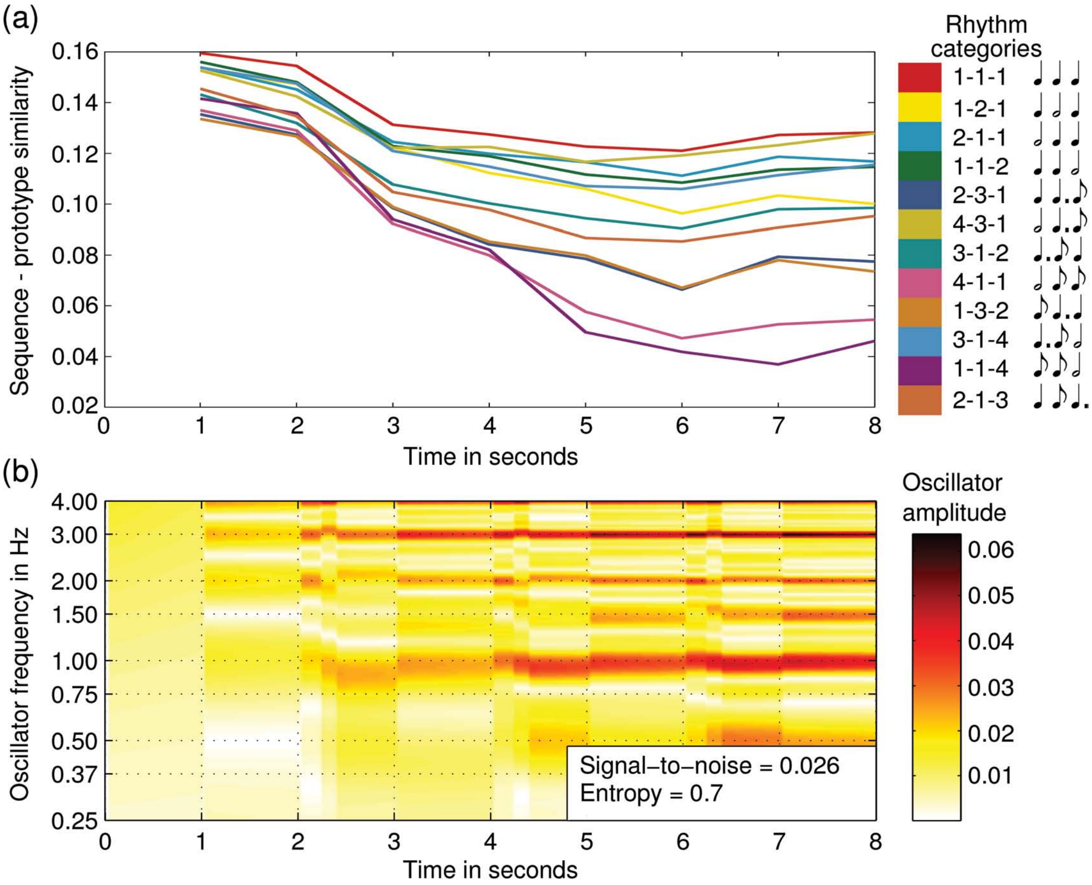


Figure 7. The oscillator activation and corresponding categorization over time for an oscillator network given a rhythm pattern that scored high entropy in Desain and Honing (2003).

categorizations. The agreement between the no meter condition and the duple and triple meter conditions for the simulated categorizations was calculated as being 77% and 71%, respectively with duple meter agreeing with the no meter categorization in 6 percentage points more of the cases ($p = 0.045$, one-tailed randomized permutation test).

4 Discussion

Many models of categorical perception have been based on neural networks and there exist several models of rhythm perception based on neural networks (Desain & Honing, 1989; Miller, Scarborough, & Jones, 1992; Mozer, 1993). We believe that using a dynamical system of resonating oscillators provides a physiologically more plausible way of modeling such phenomena. By modeling rhythm perception in such a system, we have shown that it is possible to explain empirical findings of listeners' categorical perception of rhythm. Our oscillator model has been able to accurately replicate the experimental data from Desain and Honing (2003). A possible concern is whether the model is sensitive to the choice of parameters. However, a parameter sensitivity analysis has not been performed as the purpose of the model is not to predict the experimental data as well as possible nor do we claim that the specific model configuration could not be the subject of improvements. What is claimed is that the model supports the notion that resonance theory is a viable model of rhythm perception and that by viewing rhythm perception as a dynamical system it is possible to model properties of rhythm categorization.

An advantage of oscillator models is that they can be generalized to other kinds of categorical perception. Examples from the domain of music are pitch perception and tonality perception (Large, 2010). Oscillatory models are not confined to temporal processes and can be used for other modalities. The main importance of our model is perhaps that the example of how oscillator models can be constructed for categorical rhythm perception can serve as inspiration for similar models of other cognitive phenomena. A general question is whether the convexity of rhythm categories generated from our model generalizes to other areas of categorical perception when oscillator models are used. If so, it could be interpreted as a general mechanism that can explain the convexity of categories as put forward by Gärdenfors (2000).

Acknowledgments. The authors gratefully acknowledge support from the Linnaeus environment Thinking in Time: Cognition, Communication and Learning (CCL), Swedish Research Council grant number 349-2007-8695.

References

Angelis, V., Holland, S., Upton, P. J., & Clayton, M. (2013). Testing a computational model of rhythm perception using polyrhythmic stimuli. *Journal of New Music Research*, *42*(1), 47–60. doi:10.1080/09298215.2012.718791

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

Desain, P., & Honing, H. (1989). The quantization of musical time: A connectionist approach. *Computer Music Journal, 13*(3), 56–66

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception, 32*(3), 341–365. doi:10.1068/p3370

Ernst, M. D. (2004). Permutation methods: A basis for exact inference. *Statistical Science, 19*(4), 676–685. doi:10.1214/088342304000000396

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic β oscillations. *The Journal of Neuroscience, 32*(5), 1791–1802. doi:10.1523/JNEUROSCI.4107-11.2012

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought.* Cambridge, MA: MIT Press.

Honing, H. (2013). Structure and interpretation of rhythm in music. In D. Deutsch (Ed.), *The psychology of music* (pp. 369–390). Amsterdam, The Netherlands: Elsevier.

Honing, H., Ladinig, O., Háden, G. P., & Winkler, I. (2009). Is beat induction innate or learned? Probing emergent meter perception in adults and newborns using event-related brain potentials. *Annals of the New York Academy of Sciences, 1169*(1), 93–96. doi:10.1111/j.1749-6632.2009.04761.x

Jäger, G. (2010). Natural color categories are convex sets. In *Logic, language and meaning* (pp. 11–20). New York: Springer.

Large, E. W. (1996). Modeling beat perception with a nonlinear oscillator. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 420–425).

Large, E. W. (2000). On synchronizing movements to music. *Human Movement Science 19*(4), 527–566.

Large, E. W. (2010). Neurodynamics of music. In M. Riess Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception* (pp. 201–231). New York: Springer.

Large, E. W., Almonte, F. V., & Velasco, M. J. (2010). A canonical model for gradient frequency neural networks. *Physica D: Nonlinear Phenomena, 239*(12), 905–911.

Large, E. W., & Velasco, M. (n.d.). *Nonlinear time-frequency transformation workbench: MATLAB software for signal processing using nonlinear oscillators.*

McAuley, J. D. (2010). Tempo and rhythm. In M. Riess Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception* (pp. 165–199). New York: Springer.

Miller, B., Scarborough, D., & Jones, J. (1992). On the perception of meter. *Understanding music with AI.* Cambridge, MA: MIT Press.

Mozer, M. (1993). Neural net architectures for temporal sequence processing. In A. Weigend & N. Gershenfeld (Eds.), *Predicting the future and understanding the past.* (pp. 243–264). Boston, MA: Addison-Wesley.

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience, 31*(28), 10234–10240. doi:10.1523/JNEUROSCI.0411-11.2011

Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology. Human Perception and Performance, 13*(1), 116–126. doi:10.1037/0096-1523.13.1.116

Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia, 49*(5), 961–969. doi:10.1016/j.neuropsychologia.2011.02.002

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

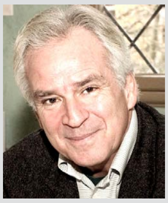
Schachner, A., Brady, T., Pepperberg, I., & Hauser, M. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology, 19*(10), 831–836. doi:10.1016/j.cub.2009.03.061

Schaefer, R. S., Vlek, R. J., & Desain, P. (2011). Decomposing rhythm processing: Electroencephalography of perceived and self-imposed rhythmic patterns. *Psychological Research, 75*(2), 95–106. doi:10.1007/s00426-010-0293-4

Snyder, J., & Krumhansl, C. L. (2001). Tapping to ragtime: Cues to pulse finding. *Music Perception, 18*(4), 455–489. doi:10.1525/mp.2001.18.4.455
Todd, N., O'Boyle, D., & Lee, C. (1999). A sensory-motor theory of rhythm, time perception and beat induction. *Journal of New Music Research, 28*(1), 5–28. doi:10.1076/jnmr.28.1.5.3124
van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research, 28*(1), 43–66. doi:10.1076/jnmr.28.1.43.3122
Wing, A., & Kristofferson, A. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics, 14*(1), 5–12. doi:10.3758/BF03198607

Rasmus Bååth is a PHD student at Lund University Cognitive Science, where he studies rhythm perception and statistical modeling. He has a background in Computer Science and writes about programming, statistics, and modeling on his blog at http://www.sumsar.net.

Erik Lagerstedt began his education with theoretical physics, and received his Master's degree in cognitive science in 2013 at Lund University, Sweden. He is about to start his PhD on socio-technical systems at the University of Skövde, Sweden.



Peter Gärdenfors has a background in mathematics and computer science and a PhD in philosophy from Lund University. He has been professor of cognitive Science at Lund University since 1988 and also has a part time appointment at University of Technology Sydney. His has developed conceptual spaces as a theory of knowledge representation and works on applying the theory to various areas from physics to event perception and semantics.

a Pion publication

Paper III

Bååth, R. and Ingvarsdóttir, K. O. (2014) Subjective rhythmization: A replication and an extension. *Proceedings of the 13th International Conference on Music Perception and Cognition.* Seoul, South Korea.

Note: This manuscript has been reformatted to better fit the paper format of the thesis. It also includes some minor corrections compared to the paper that was presented in Seoul in 2014.

Subjective Rhythmization: A Replication and an Extension

Rasmus Bååth, Kristín Ósk Ingvarsdóttir*

Lund University Cognitive Science, Sweden.

Abstract

Subjective rhythmization (SR) is the phenomena that the sounds of a monotone metronome sequence are experienced as having different intensity and that these differences follow a regular pattern. This study aimed to replicate and extend the two studies that have employed the original SR experimental paradigm (Bolton, 1984; Vos, 1973). The extensions included using a wider range of tempi and a large number of participants. The result of the study was in accordance with these two earlier studies. In addition to the original SR task, a novel task was administered where the participants were not explicitly told about the existence of the phenomena. The responses of the participants were in agreement with that subjective rhythmization was experienced. This indicates that SR is a robust phenomena that can be experience even without it having to being primed by verbal instructions.

1. Introduction

When listening to a piece of music a common response is to move one's body to a perceived periodic pulse (Snyder and Krumhansl, 2001). That pulse is the *beat* of the corresponding piece of music, a series of subjectively isochronous (equally spaced in time) events that are felt as being pronounced or accented. The beat is established by the rhythm of the musical events and in a piece of music the beat and musical events tend to coincide. It is not necessary that every beat is marked by a musical event, however, and the perception of a

* Corresponding author. E-mail: rasmus.baath@lucs.lu.se

beat can be sustained even if there are conflicting musical events (Large and Palmer, 2002).

It is not common that all beats in a piece of music are perceived as being equally accented (Palmer and Krumhansl, 1987) and a periodically recurring pattern of strong and weak accents is called a meter. For example, a duple meter would imply that every second beat is perceived as having a stronger accent while every third beat is perceived as having a stronger accent in the case of a triple meter. Perceiving the beat and meter of a piece of music often comes natural and it does not require the listener to actively attend to the music. It has even been shown that some form of beat induction is functional in newborn infants (Honing et al., 2009).

One perceptual phenomena that shows our tendency to experience metrical structure is *subjective rhythmization* (SR), a phenomena that occurs when one listens to a sequence of isochronous, identical sounds. A pattern of accents will emerge that has a metrical structure and gives the impression that there are groups of sounds. Even though the sounds are objectively identical they sound subjectively different. This phenomena was described already in the 18th century (Kirnberger, 1776) but was first investigated by Bolton (1894) who systematically played monotone metronome sequences at different tempi to a number of participants and recorded their reactions. That study was later partially replicated by Vos (1973)[1], and both studies agree on some characteristics of SR. The most common groupings participants experience are two and four, the groupings of common meters of western music. Group size and tempo interacts as participants tend to perceive smaller groupings at slower tempi and larger groupings at faster tempi, though no groupings larger than eight have been reported (Bolton, 1894; Vos, 1973). There is a limit to the range of tempi where SR can be experienced. Bolton found that participants' experience of SR ceased when the interstimuli interval (ISI) between consecutive sound onsets was above 1500 ms. After reviewing the

11 Vos (1973) has not been translated into English, but the data from that thesis has been reanalyzed by van Noorden and Moelants (1999).

literature Fraisse (1982) proposed that the limit was around an ISI of 1800 ms. Here is a connection to rhythm production as this limit is in the same range as when sensorimotor synchronization (e.g. finger tapping) begins to feel laborious (Repp, 2006; Bååth and Madison, 2012).

The present study aim to replicate the result of Bolton (1894) and Vos (1973) using a wide range of tempi and a larger number of participants than in the earlier studies. Such a replication is presented in Experiment 1. A second aim was to verify the existence of the SR phenomena. In previous studies SR has been investigated by explicitly asking what groupings participants experience when listening to a metronome sequence. By explicitly asking participants about grouping it is possible that participants get primed to experience SR. In Experiment 2 we used a novel task where the response of a participant depends on whether he or she experiences SR but where SR is not mentioned in the task instructions.

2. Experiment 1

2.1 Method

Participants were recruited, and the experiment administered, using the on-line service Amazon Mechanical Turk (Buhrmester et al., 2011). Out of the 132 participants 111 reported having experience playing a musical instrument. The task instruction given to the participants were as follows:

"This task requires your full attention. Below are six sound sequences of clicks. You should listen to each sound sequence and rate if you feel any grouping or subdivision of the clicks, however weak or subtle. For example, if you hear "TICK-tick-TICK-tick-TICK-tick" that would be groups of two, while if you hear "TICK-tick-tick-tick-TICK-tick-tick-tick" that would be groups of four. This task **is not** about whether there are groups in the sequences, it is about if you **feel** any grouping. Now listen to and rate the sequences one at a time in the order they are displayed."

The participants were then given six monotone metronome sequences with ISIs of 200 ms, 300 ms, 400 ms, 600 ms, 800 ms and 1500 ms in a randomized order. Each sequence was 15 s long and consisted of 10 ms long, 440 Hz sine wave sounds. After listening to a sequence the participant indicated what grouping he or she felt by using a list with the alternatives “no group” and “groups of 2” up to “groups of 8”.

2.2 Result

The result replicated many of the findings of Bolton (1894) and Vos (1973) and Figure 1 shows the reported experienced grouping from six participants. In general, participants reported larger groupings at faster tempi, as can be seen in participant C or F. Some participants were consistent, like participants A and B, while some were less consistent, like participant C and E. There was a tendency that some participants (for example, participant D) answered “no grouping” on the faster tempi (200 ms and 300 ms).

The most reported groupings were two, four and eight with five and seven being rarely reported at all. Table 1 show how often the participants reported each possible grouping. Figure 2 show for each ISI level the proportion of participants that reported each grouping. The slower limit of SR was estimated to an ISI of 1500 by Bolton. In the current study no such sharp limit was found, however, a majority of the participants (81 %) reported experiencing no grouping at an ISI of 1500 ms.

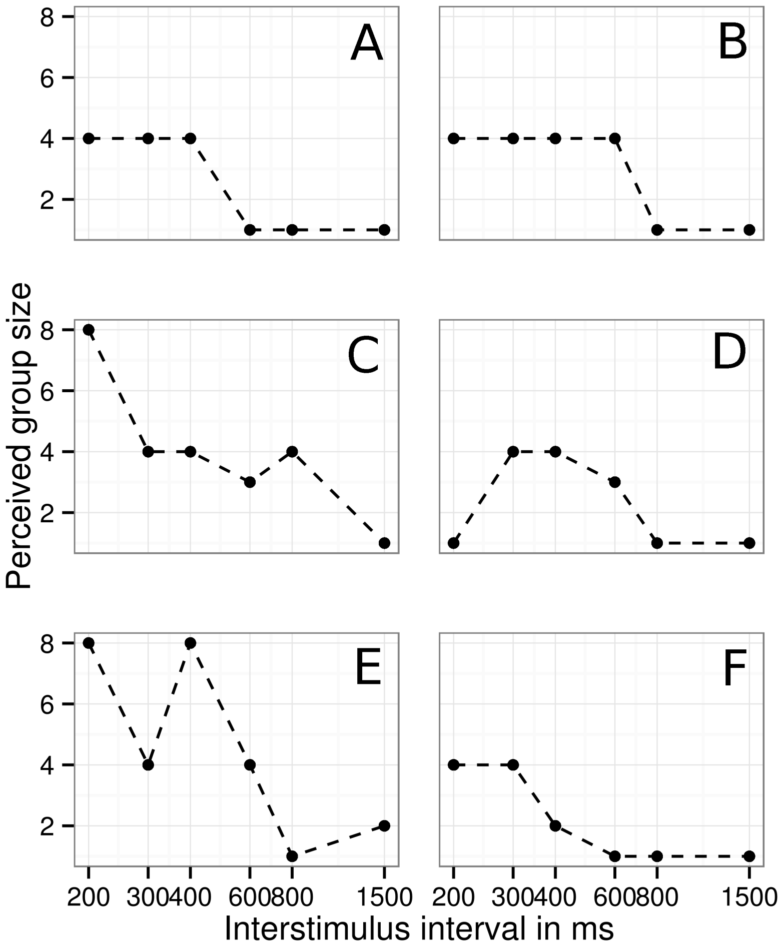


Figure 1: The reported perceived group size for six of the 132 participants in Experiment 1. Note that a group size of one corresponds to the participant having reported "no grouping".

Grouping	Peak ISI	% of responses
1 (No grouping)	1500	34 %
2	600	23 %
3	300	7 %
4	300	27 %
5	200	< 1 %
6	300	1 %
7	200	< 1 %
8	200	6 %

Table 1: Summary of the reported groupings in Experiment 1.

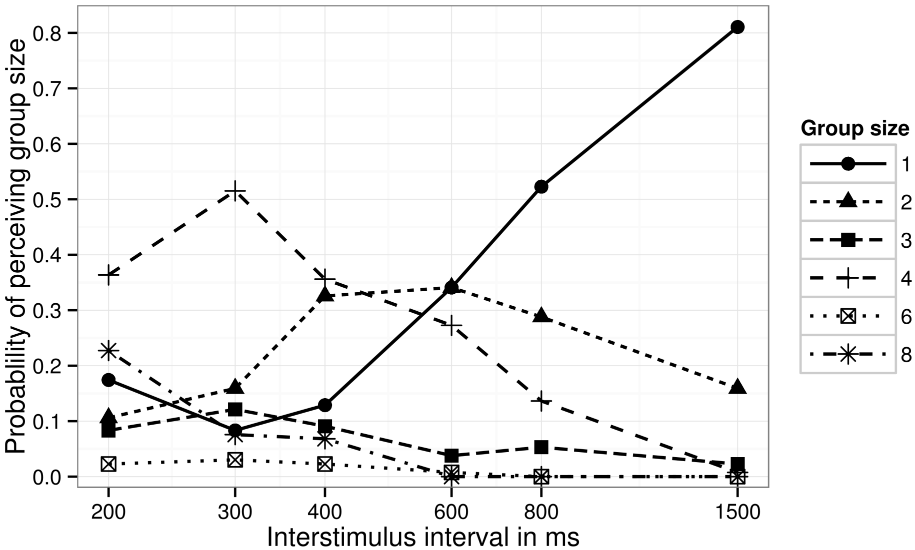


Figure 2: Percentage of reported groupings as a function of ISI. A group size of 1 corresponds to “no grouping”.

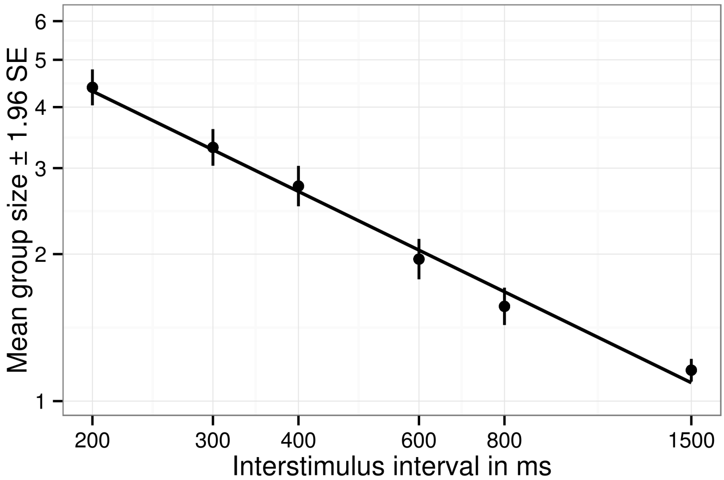


Figure 3: Log-log plot of mean group size as a function of ISI. The line show the best fitting regression line.

Figure 3 shows the mean group size reported for each ISI level with both axes being on the log scale. This relation appears highly linear, a result not previously reported in the literature. A linear regression with $\log_2$ group size as the dependent variable and and $\log_2$ ISI level as the independent variable gave an intercept of 8.1 (95% bootstrap[2] confidence interval [5.8, 10.5]), a slope of -0.77 (95% CI [-1.0, -0.53]) and an R^2 of 0.50 (95% CI [0.23, 0.77]).

3. Experiment 2

3.1 Method

The purpose of Experiment 2 was to investigate whether the phenomena of SR would influence participants responses even though SR was not mentioned in the instructions nor suggested in any way. Amazon Mechanical Turk was again used to recruit and administer the task to 120 participants, 60 in each of two conditions, where the only difference between the conditions were whether the following task instructions used the word *second* or *fourth*:

2 The bootstrap confidence intervals (CI) were calculated using 10,000 resamples.

“In this task we are interested in if it is possible to feel very small differences in loudness. Below are 14 click sequences, in some of them all clicks are equally loud and in some of them every [*second, fourth*] click is a little bit louder. The difference in loudness will be **very** small. Listen to the sequences, in the order they are given, and for each sequence try to feel if the clicks are equally loud or if every [*second, fourth*] click is louder.”

The 14 click sequences were the same as in Experiment 1, with the addition of a 2000 ms ISI sequence, each given twice in a randomized order. That is, despite the task instructions, all clicks were actually equally loud. After having listened to each sequence the participant was asked whether he or she perceived a difference in loudness or not. If the participant, without knowing it, experienced SR when listening to the sound sequence we would expect him or her to be more likely to report a difference in loudness. If a participant was given the *second*-instructions, to listen for a difference on every second click, we hypothesized that he or she would direct attention towards SR with a grouping of two and therefore be more likely to report a difference at an ISI of around 600 ms (cf. Figure 2). Similarly, if a participant was given the *fourth*-instructions and was listening for a difference on every fourth click, we would expect him or her to be most likely to report a difference around an ISI of 300 ms.

3.2 Result

There was a tendency for the participants given the *second*-instructions to report hearing a difference around an ISI of 600 ms while the participants given the *fourth*-instructions were more likely to report a grouping at the ISIs of 300 ms and 400 ms. Figure 4 show the probability of reporting hearing a difference for each ISI level, where there is a clear peak around 600 ms for the *second*-condition and around 300 ms and 400 ms for the *fourth*-condition. At the ISI levels of 1500 ms and 2000 ms, where the participants in Experiment 1 largely reported hearing no grouping, there is an increase in reporting hearing a difference for the *fourth*-condition. A reason for this could be that at such slow tempi some participants find it difficult to compare every fourth click and therefore approach the chance level of 50%.

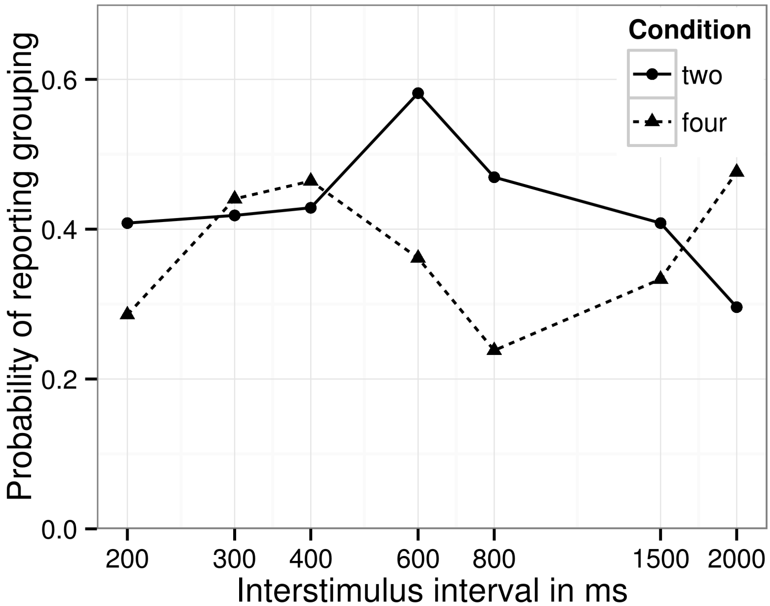


Figure 4: The probability of reporting hearing a grouping for the two instruction conditions in experiment 2.

4. Conclusion

Experiment 1 replicated the main findings of Bolton (1894) and Vos (1973):

- Subjective rhythmization is a robust phenomena that seems to be experienced by most participants.
- The reported experienced grouping is most often two, four or eight, common meters of western music.
- What grouping that is reported is highly dependent on the tempo with larger groupings being reported at faster tempi.

- Most participants do not reporting hearing a grouping when the ISI is as slow as 1500 ms.

Experiment 2 confirmed the robustness of SR and showed that SR influence participants' responses in a task that is quite different from the original SR task.

It should be noted that, as both experiments used Amazon Mechanical Turk, the experiments were administered on-line without the usual control of a perceptual experiment. This can be viewed both as a weakness and as a strength, a weakness because there was no control over what environment the participants were in when doing the experiment, a strength because despite the lack of control the result is well in agreement with the earlier studies of Bolton (1894) and Vos (1973).

5. Acknowledgement

This research was supported by the Linnaeus environment Thinking in Time: Cognition, Communication and Learning, Swedish Research Council grant number 349-2007-8695.

6. References

Bolton, T. (1894). Rhythm. *The American Journal of Psychology*. Retrieved from http://www.jstor.org/stable/1410948

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6(1), 3–5. doi:10.1177/1745691610393980

Bååth, R., & Madison, G. (2012). The subjective difficulty of tapping to a slow beat. *In Proceedings of the 12th International Conference on Music Perception and Cognition* (pp. 82–85). Retrieved from http://www.sumsar.net/papers/ICMPC_2012_rasmus_baath.pdf

Honing, H., Ladinig, O., Háden, G. P., & Winkler, I. (2009). Is beat induction innate or learned? Probing emergent meter perception in adults and newborns using event-related brain potentials. *Annals of the New York Academy of Sciences*, 1169(Illc), 93–6. doi:10.1111/j.1749-6632.2009.04761.x

Kirnberger, J. P. (1776). *Die Kunst des reinen Satzes in der Musik: Aus sicheren Grundsätzen hergeleitet und mit deutlichen Beyspielen erläutert*. Decker.

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26(1), 1–37. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S036402130100057X

Van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1), 43–66. doi:10.1076/jnmr.28.1.43.3122

Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance*, 13(1), 116–26. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2951485

Snyder, J., & Krumhansl, C. L. (2001). Tapping to Ragtime: Cues to Pulse Finding. *Music Perception*, 18(4), 455–489. doi:10.1525/mp.2001.18.4.455

Vos, P. (1973). *Waarneming van metrische toonreeksen*. Nijmegen: Stichting Studentenpers.

Paper IV

Bååth, R. (In press) Subjective rhythmization: A replication and an assessment of two theoretical explanations. *Music Perception.*

Subjective Rhythmization: A Replication and an Assessment of Two Theoretical Explanations

Rasmus Bååth

Abstract

Subjective rhythmization is that phenomenon whereby, when one is listening to a monotone metronome sequence, some sounds are experienced as accented. These subjectively accented sounds group the sequence similarly to how the metrical structure of a piece of music groups the beats. Subjective rhythmization was first investigated by Bolton (1894); the present study aims at replicating and extending that work. Consistent with Bolton's results all participants reported hearing accent patterns when listening to monotone sequences; the reported group size of an accent pattern was highly dependent on the tempo of the sequence. A power relation captured well the relation between the reported group size and the sequence interstimulus interval. Further, the mean group size reported in the subjective rhythmization task was found to correlate with the timing performance in a slow-tempo tapping task. These results are consistent with the *resonance theory* explanation of subjective rhythmization (Large, 2008).

1 Introduction

Rhythm, the temporal organization of distinct sound events, is an integral part of human speech and music (Patel, 2008). Humans have an astonishing capability both to perceive and to produce rhythms. *Subjective rhythmization* (SR) is one example of this capacity. This is the phenomenon whereby sounds of a monotone metronome sequence are experienced as having different intensity and that these intensity differences follow a regular pattern. In other words, despite the sounds having objectively equal amplitude, they are perceived as subjectively different. Bolton (1894) developed an experimental paradigm for investigating SR; only one other study exists that uses Bolton's original paradigm (Vos, 1973). The current study aims to replicate and extend Bolton's and Vos' work. Extensions include using a wider range of tempi, employing a larger number of participants and presenting those participants with a number of auxiliary tasks in addition to the SR task. The inclusion of the auxiliary tasks is motivated by three decisive predictions developed from two proposed explanations for SR: the *preferred tempo explanation* (Temperley, 1963) and the *resonance theory explanation* (Large, 2008).

A typical example of SR is when identical ticks of a clock are perceived as "tick tock" (Brochard, Abecasis, Potter, Ragot, & Drake, 2003; van Noorden & Moelants, 1999). For this reason, SR has also been called the *clock illusion* or the *tick-tock effect* (Vlek, Schaefer, Gielen, Farquhar, & Desain, 2011). An alternative way of viewing SR is as the imposition of a subjective meter onto a sequence of sounds, where no meter is enforced through physical intensity or physical pitch differences. It has been pointed out that the term subjective rhythmization is a misnomer and that a more suitable term would be subjective meter (Large, 2008) or subjective accentuation (Temperley, 1963).

Subjective rhythmization was discussed already in the 18th century (Kirnberger, 1776) but not investigated experimentally until Bolton (1894)'s seminal work. Bolton used apparatus capable of producing isochronous (temporally equally spaced) sequences of monotone clicks of equal amplitude. By systematically varying the tempi of the sequences he established the following characteristics of SR. Isochronous sequences of identical sounds produce the impression that some sounds are louder or more intense than others. The apparent increases in intensity do not appear randomly but recur every nth sound, resulting in the more intense sounds grouping the sequence. Here n can range from two up to eight but the most common reported groupings participants reports are two, three and four: the common metrical groupings of western music. Group size and tempo are related; participants reports smaller groupings at slower tempi and larger groupings at faster tempi. The range of tempi at which SR can be experienced is limited. Bolton found that SR experience ceases when the interstimuli interval (ISI) between consecutive sound onsets rises above 1600 ms, though a later review of Bolton's results suggested a slower limit of 1800 ms (Fraisse, 1982).

Only one study, that by Vos (1973), which has employed Bolton's (1894) experimental paradigm, despite recent interest in the electrophysiological properties of SR (e.g., Nozaradan, Peretz, Missal, & Mouraux, 2011; Schaefer, Vlek, & Desain, 2011). Vos' study, though limited by a relatively small number of trials and narrow tempo range of the stimuli (ISIs of 150 to 800 ms), produced results in accord with Bolton's. Subsequent analysis of Vos's data by van Noorden and Moelants (1999) emphasizes (1) the dependency between tempo and reported group size, (2) a propensity toward reporting even-numbered groups, and (3) an average interval between each group's onset longer than one second.

1.1 Explanations for Why Subjective Rhythmization Occurs

The literature offers two explanations for SR: one relating to participants' preferred tempo (Temperley, 1963) and one explaining SR using the resonance theory of rhythm perception (Large, 2008).

1.1.1 The preferred tempo explanation

When experiencing SR, one hears the sounds of a monotone sequence as grouped, with the first sound in each group being accented. This grouping of the sounds can be viewed as a modification of the period of the sequence, where the *group period* is defined as the period between group onsets. An example of such a modification would be when a participant is given a monotone tone sequence with an ISI of 250 ms and reports a grouping of two, resulting in a group period of 500 ms. The preferred tempo explanation is that participants experience a grouping that results in a group period close to their preferred tempo (Temperley, 1963)so as to facilitate entrainment to the sequence.

A regular observation is that, when participants are asked to tap an isochronous rhythm at a comfortable rate, the resulting tempi tend to cluster around a period of 500-600 ms (Fraisse, 1982). This tempo is called the *spontaneous motor tempo* (SMT) and has been shown to be strongly correlated ($r = 0.75$) with participants' verbal reports of preferred beat tempo (McAuley, Jones, Holub, Johnston, & Miller, 2006), supporting the existence of an intrinsic preferred rate for event tracking; the SMT may be seen as the tempo where rhythm perception is optimal (Moelants, 2002).

Present knowledge about SR does not favor the preferred tempo explanation, however. Especially problematic is the observation that the group period tends to be above one second (Vos, 1973, as analysed by van Noorden & Moelants, 1999) which is far from the common period of spontaneous motor tempo.

1.1.2 The resonance theory explanation

The resonance theory of rhythm perception (Large & Jones, 1999; Large & Kelso, 2002; van Noorden & Moelants, 1999) offers an alternative explanation. According to resonance theory, experiencing the beat of a piece of music or an isochronous sequence of sounds is an emergent phenomenon, caused by neural *oscillatory circuits* that resonate with incoming auditory events. An oscillatory circuit (henceforth *oscillator*) with intrinsic period T entrains to sound events with a similar period. More specifically, sound events with period T cause the amplitude of oscillators with similar periods to increase. The resulting oscillator amplitude indicates the extent to which events of period T occurred in the auditory stream.

Neural resonance is a common theme underlying resonance accounts of rhythm perception. Within this framework though, models differ in whether they model beat perception using a small number of oscillators or a large network of oscillators. The resonance theory explanation of SR assumes the latter, motivated by the observation that the brain encodes information using populations of neurons (Averbeck, Latham, & Pouget, 2006). By assuming multiple oscillators, the account allows for modeling meter perception involving the temporal organization of beats on multiple time scales (Large & Kolen, 1994).

Models using multiple oscillators (e.g., Large, 2000, Scheirer, 1998) differ in implementation, but the basic mechanism is the same. A network of oscillators, where each oscillator has an intrinsic period, is given an auditory input. The amplitude of an oscillator with period T reflects the extent to which sound events with period T occurred in the auditory stream. The sum of the amplitudes of all oscillators in the network reflects periodicities in the auditory stream. Precisely what periodicities the network is sensitive to depends on the distribution of the intrinsic periods of the oscillators.

Following Large (2008), an explanation for SR using this multiple-oscillator version of resonance theory is based on the notion that an isochronous sequence of sounds with period T will, in addition to entraining oscillators attuned to that period, entrain oscillators at subharmonics of T (i.e., $2 \cdot T$, $3 \cdot T$, $4 \cdot T$, etc.). The summed output from a network of oscillators will contain amplitude fluctuation at the subharmonic frequencies of the given sequence, even if the sequence itself has no fluctuations in amplitude. See Figure 1, where the sound sequence activates both the oscillator with matching period (Oscillator 1) and the oscillator with a period that is twice as slow (Oscillator 2), resulting in an SR with a grouping of two ("Network output"). In support of this ac-

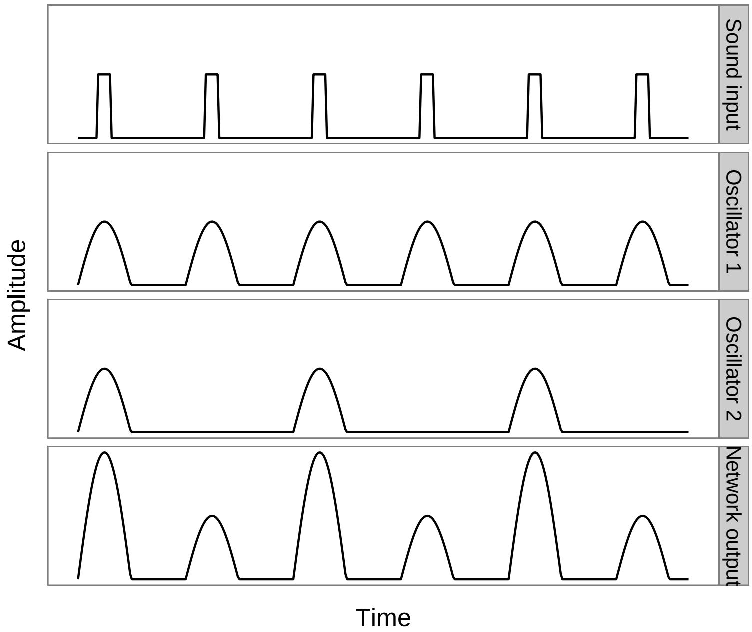


Figure 1: Schematic plot of subjective rhythmization in a resonance theory framework.

count, Nozaradan et al. (2011) found that, when participants are asked to listen to an isochronous sound sequence and subjectively impose an accent on every second beat, the resulting electroencephalogram reveals a sustained response at the period of the imposed accent.

Two other aspects of SR can be explained by a multiple oscillator resonance model: (1) why the feeling of SR disappears when the tempo is sufficiently slow and (2) why the size of the perceived groups, and consequently the number of sound onsets between subjective accents, is larger when the tempo is faster. Experiencing SR while listening to a sequence with period T requires oscillators that have at least twice the period of T, otherwise there would be no oscillators to mark every second (third, fourth, fifth, etc.) sound of the sequence. The vanishing point of SR then depends on the the *slower limit* of rhythm perception, that is, the period where the oscillator density is sufficiently low so that it is not possible to entrain reliably to a rhythm of that period. This is illustrated in Figure 2 where T_1 is the longest period to which the model is able to entrain and T_2 is the longest period at which SR is still experienced. The size of the perceived groups grow as the period of the sequence becomes shorter because there exist oscillators at higher order

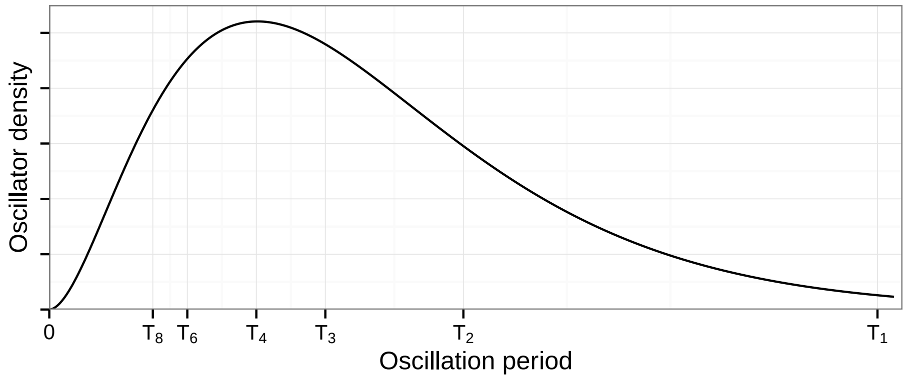


Figure 2: Schematic plot of the oscillator density as a function of the period.

subharmonics relative to the period of the sequence. As Figure 2 shows, T_3 marks the period for which T_1 is the third subharmonic of T_3, that is, T_3 marks the period for which one finds slow enough oscillators to put an accent on every third beat, resulting in an SR with a grouping of three. Further, T_4 and T_8 mark the periods for which T_1 is the fourth and eight subharmonic, respectively.

1.1.3 Predictions arising from the two explanations of subjective rhythmization

The two alternative explanations above make a number of predictions regarding participants' behavior in an SR task as well as relations between that behavior and behavior in other tasks measuring aspects of rhythm perception and production.

The first prediction regards the average group period in the SR task. Remember that the preferred tempo explanation predicts a subject's average group period to be close to her preferred tempo. According to the resonance theory explanation, on the other hand, the group period depends on the slower limit of rhythm perception (T_1 in Figure 2), and should fall somewhere between the slower limit of SR (T_2) and T_1. These two predictions are clearly distinct: preferred tempi, measured using an SMT task, tend to center on a period of 500 ms, while a slower limit of rhythm perception is believed to be above 1500 ms (Repp, 2006). This slower limit can be estimated by way of the *slow motor tempo* task in which a participant is asked to tap as slowly as possible while still maintaining a continuous, regular rhythm (McAuley et al., 2006).

The resonance theory explanation makes a second prediction, regarding the functional relation between the period of the stimulus sequence and the experienced group size in the SR task. As Figure 2 shows, the maximum possible group size g for a sequence with period T depends on the slower limit of rhythm perception T_1, so that $g \sim \frac{T_1}{T}$. This can be written more generally as the power function $g \sim \frac{k}{T^a}$, where k equals T_1 in the case where the constant exponent a equals 1. Plotted on log-log axes, power laws plot as a straight line with a slope determined by the exponent: $\log(g) \sim \log(k) - a \cdot \log(T)$.

The resonance theory explanation makes a third prediction regarding the relation between the SR task and sensorimotor synchronization performance at slow tempi. Within the resonance theory framework, both rhythm perception and rhythm production rely on the same mechanism: the entrainment of neural oscillatory circuits to regularities in the sequence of sounds. Both the slow limit of rhythm perception and rhythm production performance at slow tempi depend on the period at which there cease to be sufficient oscillators to entrain reliably to a sound sequence with corresponding period. The expectation is that participants with a relatively fast slower limit of rhythm perception should struggle to synchronize to a rhythmic stimulus at a slow tempi. As noted, the group period in an SR task is expected to be close to a participant's slower limit of rhythm perception; therefore, the *mean group period* can be seen as a proxy variable for that participant's slower limit. One can obtain a measure of synchronization performance at slow tempi by measuring variability in a finger tapping task, where participants are asked to tap in synchrony with isochronous sequences (Repp, 2005). By giving participants both sequences that are comfortably paced and ones that are in the area of the slower limit of rhythm perception, one can factor out variability due to slow tempo from variability due to motor response.

Together, these predictions motivate the inclusion of three auxiliary tasks when extending the SR task introduced by Bolton's (1894): an SMT task, a slow motor tempo task and a taping task using slow pacing sequences.

2 Method

2.1 Participants

Nine female and 21 male participants, ranging in age from 19 to 78 years (M=31.6, SD=12.8), were recruited from the Lund community. All were unpaid volunteers. All reported being right handed. Twenty-six reported experience playing a musical instrument, of which ten reported playing or practicing regularly for more than ten years.

2.2 Stimuli and Apparatus

The stimuli for the SR task were isochronous sequences of click sounds created with a click-track generator included in the sound editor Audacity[1]. Each click consisted of a 440 Hz sine wave of 10 ms. Each sequence consisted of 15 seconds of clicks repeated at a constant ISI. Sequences were presented at eight tempi, corresponding to click ISIs of 150, 200, 300, 600, 900, 1200, 1500, and 2000 ms. The sequence with an ISI of 2000 ms is slower than the proposed slower limit of SR (Fraisse, 1982); participants were expected to report no SR while listening to it. Its inclusion was for detecting any subjects who misinterpreted instructions.

For the SMT task, slow motor tempo task and the tapping task, participants used a custom-built tapping board consisting of a piezoelectric sensor mounted on a 5 cm^2 piece of corrugated fiberboard (see Bååth, 2011, for details). Participants tapped on the pad using their right index finger, with their hand resting on a plastic foam cushion. For the tapping task, the stimuli consisted of isochronous sequences of 440 Hz square wave tones of 20 ms. Each sequence consisted of 31 tones. Sequences were presented at five tempi, corresponding to tone ISIs of 600, 1200, 1800, 2400 and 3000 ms. An Arduino microcontroller controlled both generation of sounds and registration of taps. All stimuli were delivered through full-sized head phones (Philips SHP2500).

2.3 Procedure

Participants were tested individually in a quite room. The experimental tasks comprised an SR task, a tapping task, an SMT task and a slow motor tempo task, all performed during a single session which, on average, lasted one hour. The order of the SR task and the tapping task was randomized so that the SR task preceded the tapping task for 15 of the 30 participants. The SMT and the slow motor tempo tasks consisted of three trials each. The SMT trials were interleaved between the SR and the tapping task while the slow motor tempo trials were presented last. See Figure 3 for a flowchart of the experimental procedure.

[1]http://audacity.sourceforge.net/

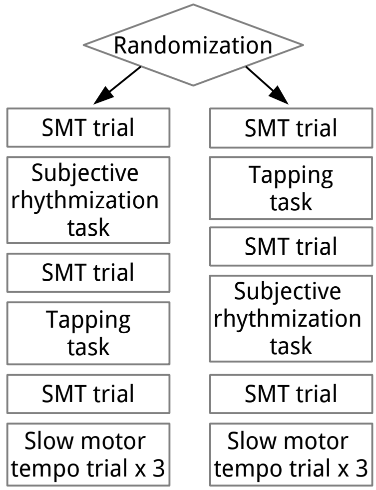


Figure 3: Flowchart of the experimental procedure.

2.3.1 The subjective rhythmization task

Each participant was placed in front of a computer with head phones. Prior to the task a 600ms ISI click sequence was played and the participants were informed that all clicks in the sequence were equally loud and equally spaced. Each participant was asked if she nevertheless experienced a grouping of the clicks or if some clicks were more dominant. The possible groupings of the sequence were explained, from none up to a grouping of eight. The 600 ms ISI click sequence was replayed. At this point, all participants reported experiencing a grouping of the clicks. These instructions conform to those described by Andrews (1905) in his discussion of Bolton's work as a *Test of Involuntary Rhythmisation with Suggestion*.

Participants then began the task proper, which consisted of four blocks of eight trials each: one for each click-sequence ISI level. The order of the trials within each block was randomized. Each participant was asked to attend to each sequence and report the first grouping that she experienced. This was done using a computer interface by selecting the appropriate alternative from a drop-down list with the alternative "No grouping/groups of one" and alternatives "Groups of two" up to "Groups of eight" (translated from Swedish) [2]. The task was self paced and no participant was interrupted while engaged in the task.

[2] A public version is available at http://sumsar.net/files/sr_task/public_sr_task.html.

2.3.2 The tapping task

Each participant sat, wearing head phones, in front of the tapping board and was asked to adjust the volume to a comfortable level while a tone sequence was played. The tapping task consisted of four blocks of five trials each, one for each ISI level. The order of the trials within each block was randomized. A trial consisted of each participant tapping along with a tone sequence, using her dominant hand. Participants were instructed to tap along to each tone sequence, to start tapping as soon as the sequence began, and to stop tapping when the sequence stopped. Participants were requested not to subdivide the beat in any way, for example, by covert counting or by moving the body.

2.3.3 The spontaneous motor tempo task

The setup was similar to the tapping task. Prior to each trial, participants were instructed to tap a regular rhythm at a tempo that felt comfortable and natural, and that felt neither too fast nor too slow. Participants were told to start tapping when ready and to continue until given notice. Thirty-one taps were recorded before participants were asked to stop.

2.3.4 The slow motor tempo task

The setup was similar to the SMT task, the only difference being that participants were asked to tap at their slowest possible rate while still able to maintain a regular beat. Again, Participants were asked to refrain from subdividing taps in any way, either overtly or covertly. These instructions conform to those described by McAuley et al. (2006). For each participant, the first fifteen taps were recorded.

2.4 Analysis

Of primary interest to the present study is participants' SR experience of monotonic tone sequences. That is, the perceptual experience is of interest, while differences in how participants approach the task are seen as a confounding variable. The slower limit of SR has been estimated to lie between an ISI of 1500 and 1800 ms (Fraisse, 1982). Any participant who repeatedly reports experiencing a grouping at an ISI well above this limit is assumed to have misinterpreted instructions. This study included four trials, with an ISI of 2000 ms added to detect such participants. Five of the thirty participants reported experiencing a grouping on all four trials at the 2000 ms ISI level. These participants were removed from further data analysis.

For each participant, the mean group period was estimated using only those trials for which a perceived grouping was reported, using the formula:

$$\frac{\sum_{i=1}^{n} T_i \cdot g_i}{n}$$

...where n is the number of trials for which the participant reported experiencing a grouping, g_i is the reported group size for the ith trial, and T_i is the corresponding ISI. As an example, consider a participant who reports hearing a grouping of four at an ISI of 300 ms, a grouping of two at an ISI of 600 ms, and a grouping of two at an ISI of 900 ms. The mean group period would then be $(4 \times 300 + 2 \times 600 + 2 \times 900)/3 = 1400$ ms.

For the tapping task, the first four taps in every trial were discarded in order to use only those taps where the participants had had some time to synchronize to the sequence. For each trial, tapping variability was calculated as the standard deviation (SD) of the tone-to-tap asynchronies. The increase in timing variability due to slowing of the tempo was estimated by fitting an ordinary least squares regression to the SD of the asynchronies as a function of ISI. The slope of such a regression line measures how much worse a participant performs as a result of slowing the tempo; a participant with a small *variability slope* is comparably better at coping with a slow tempo than a participant with a large slope.

Figure 4 shows an example of these measures. Specifically, it shows two participants' reported groupings from the SR task and timing variability from the tapping task. Participant B reported experiencing larger groupings and was better at synchronizing to a slower tempo than participant A, as reflected in the measures of mean group period and variability slope: participant B has a smaller slope and a larger mean group period.

For each participant, the mean spontaneous motor tempo and slow motor tempo were estimated by first calculating the mean intertap interval for each trial, then taking the mean of the three trials for each task. Statistical analysis was performed using the statistical computing environment R (R Core Team, 2012).

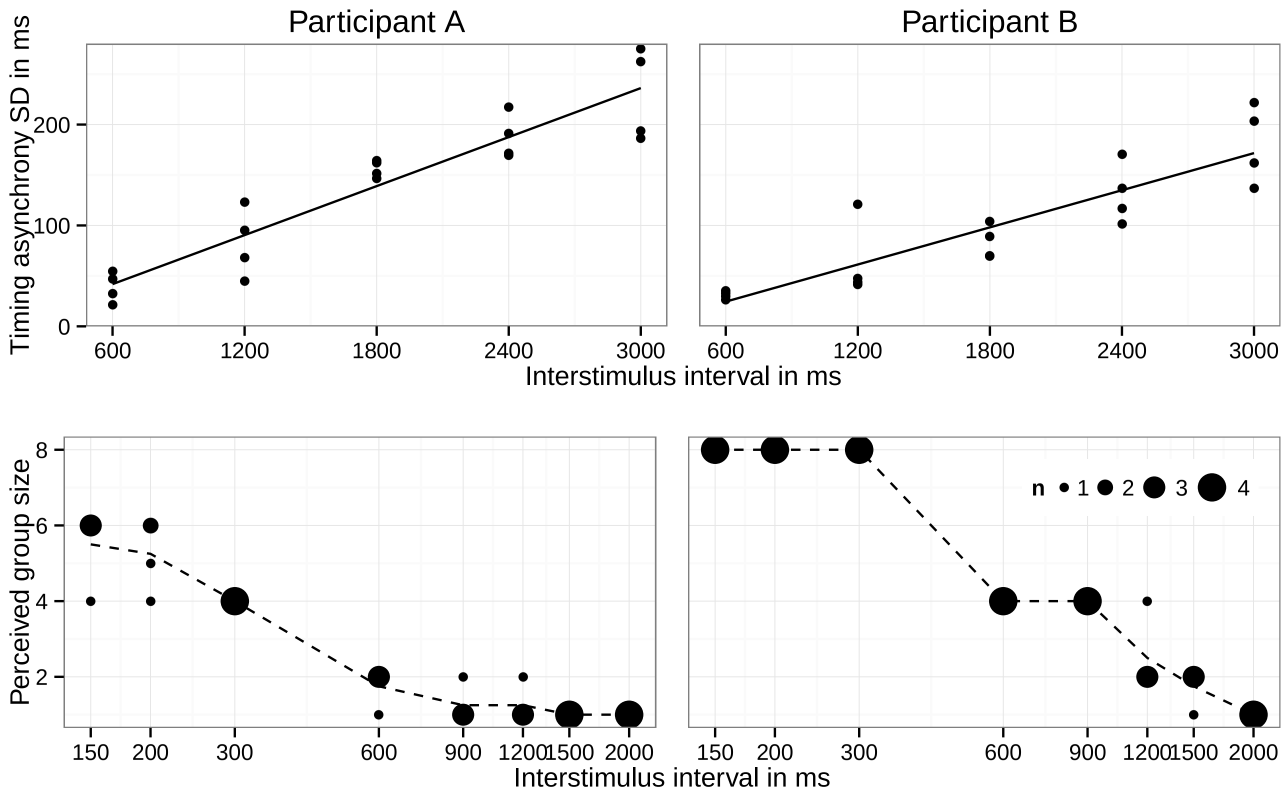


Figure 4: Timing variability from the tapping task (top row) and reported groupings from the SR task (bottom row) for two participants.

Grouping	% of Trials	Peak ISI
1	23.9	2000 ms
2	29.5	600 ms
3	3.5	300 ms
4	29.6	200 ms
5	0.8	150 ms
6	2.2	200 ms
7	0.1	150 ms
8	10.4	150 ms

Table 1: Summary of reported groupings in the SR task.

3 Results

All participants reported hearing groupings when listening to the monotone isochronous sound sequences, despite being told explicitly that the sound sequences were monotone. The most commonly reported groupings were two, four and eight; three and six were less common; five and seven were rarely reported. Table 1 shows the percentage of responses for each group size and the ISI where that group size was most commonly perceived. A group size of one indicates that the participants reported no grouping.

The percentage of reported groupings as a function of ISI is shown in Figure 5. Reported group size increases as ISI decreases, both at the group and individual level, i.e., ISI level correlated negatively to reported group size for all participants (Spearman's rank correlation with correction for tied values, mean r_ρ= -.77, SD = 0.15, $p < 0.05$ for all participants). For no ISIs did all participants cease to experience a grouping, however, more than half the trials above an ISI of 1500 ms did not result in any experienced groupings.

As a measure of consistency, the probability of reporting the same group size in two different trials with the same ISI was estimated for each participant. Using this measure, participant A in Figure 4 had a consistency of .73, participant B a consistency of .91, and the overall mean consistency was .69 (SD = 0.13). Figure 6 shows the mean consistency across participants at different ISIs. The ISI with the highest consistency was 2000 ms (M=.75); the ISI with the lowest consistency was 1200 ms (M = .65). Participants were comparably consistent at different ISIs; the standard deviation of the mean consistency across ISIs was 0.034. To put this into perspective, these consistency measures can be compared to those resulting from randomly reporting groupings,

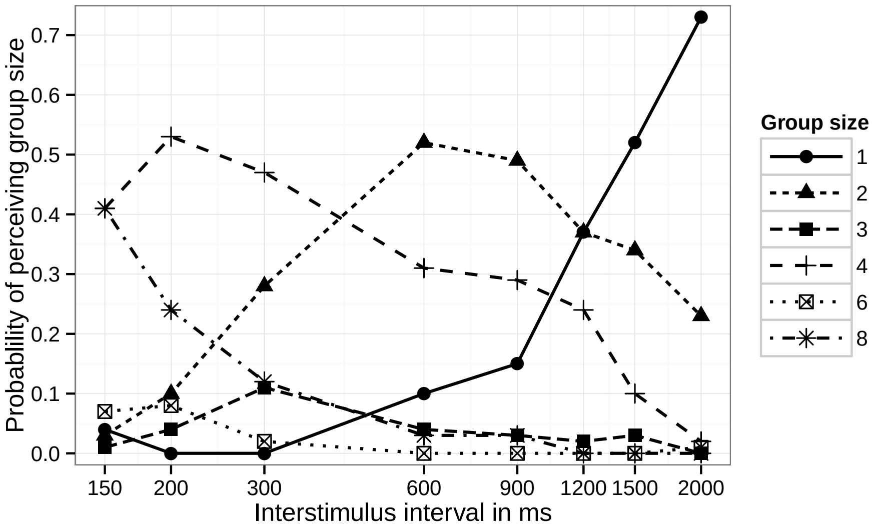


Figure 5: Percentage of reported groupings as a function of ISI.

according to the group probabilities presented in Table 1. Using this scheme, the consistency is 0.24 (marked by the dashed line in Figure 4), a much lower consistency than any of those calculated using the data.

The mean of the logarithm of reported group sizes was calculated for each participant and ISI level, as shown in Figure 7, where the grand mean is plotted against $\log_2(\text{ISI})$. The relationship between reported group size and ISI appears linear, in line with the hypothesis that this relationship would follow a power law. A linear regression between $\log_2(\text{ISI})$ and the mean of the logarithm of the reported group sizes for each participant yields the power law relation $g = \frac{k}{T^a}$, where estimates of both the factor and exponent are significantly different from zero (k = 76.7, t = 25.1, p < .001; a = 0.53, t = 19.9, p < .001 ; R^2 = 0.67, df = 198).

The grand mean of the mean group period was 1881 ms (SD = 656 ms), the grand mean spontaneous motor tempo was 622 ms (SD = 157 ms), and the grand mean slow motor tempo was 2757 ms (SD = 1100 ms). Figure 8 show the distributions of these three measures. The resonance theory explanation of SR predicted that the mean group period should fall between the slower limit of SR and the slower limit of rhythm perception. The data shown in Figure 8 are in accord with this prediction given that the slower limit of SR is estimated as the ISI where more that half of the trials result in no grouping (ISI 1500 ms) and the slower limit of rhythm perception is estimated by the average slow motor tempo (2757 ms).

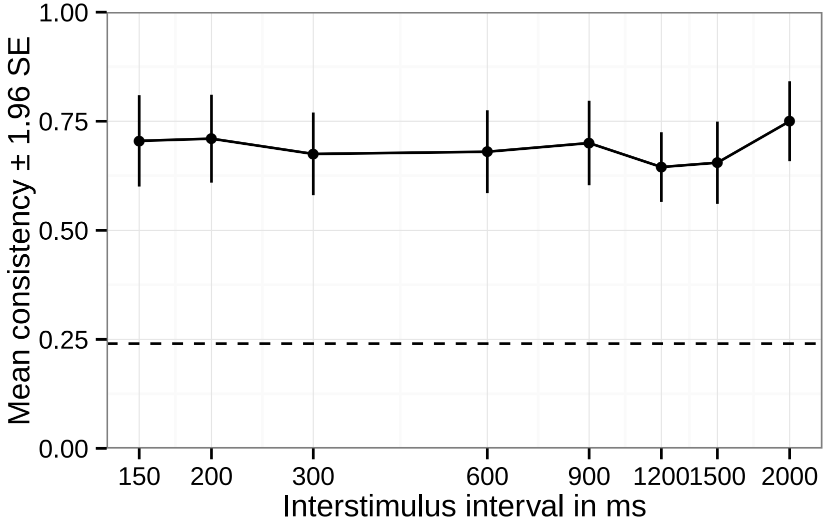


Figure 6: The mean participant consistency at different ISIs as measured by the probability of reporting the same grouping in two separate trials. The dashed line shows the expected consistency if participants would have reported groupings at random.

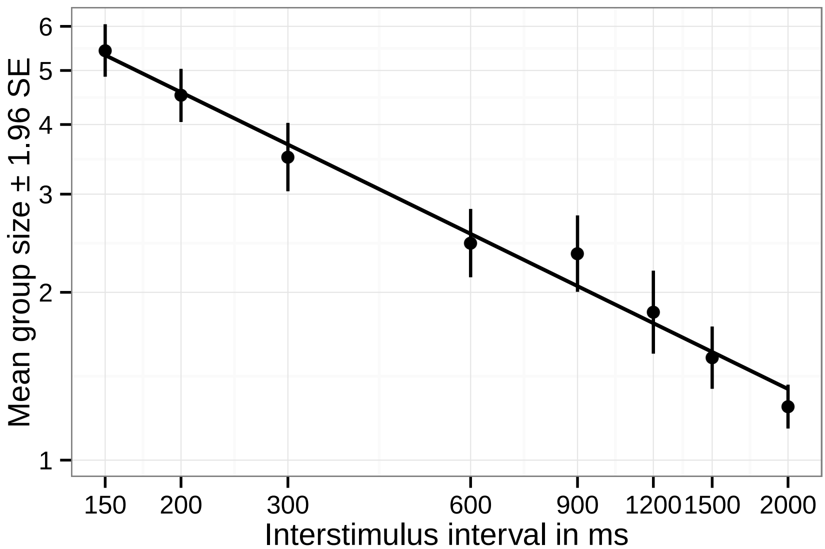


Figure 7: Log-log plot of mean group size as a function of ISI.

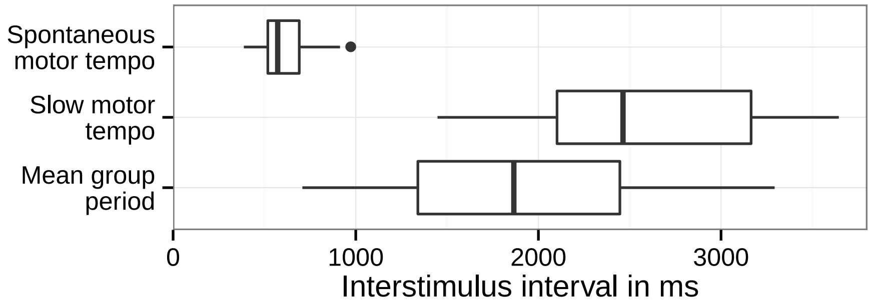


Figure 8: Distributions of participants' mean spontaneous motor tempo, mean slow motor tempo and mean group period.

Figure 9 shows the relation between a subject's mean group period and variability slope. There was a negative correlation between mean group period and variability slope across participants (Spearman's rank correlation, r_ρ = -.56, p = .0044, n = 25). From a resonance theory perspective, this implies a tendency for participants with a fast slower limit of rhythm perception to have relatively larger timing errors when tapping at slow tempi. There was no significant correlation between years practicing a musical instrument and either mean group period (Spearman's rank correlation with correction for tied values, r_ρ= -.34, p = .093, n = 25) or timing error slope (r_ρ= -.023, p = .91, n = 25).

4 Discussion

Subjective rhythmization (SR) is the phenomenon whereby the sounds of a monotone metronome sequence are experienced as having different intensity, with the experienced intensity differences following a regular pattern. The present study aimed to replicate and extend the two studies employing the original SR experimental paradigm (Bolton, 1894; Vos, 1973). The extensions were the use of a wider tempo range, the inclusion of multiple trials per tempo level, and the administration of supplemental rhythm production tasks, motivated by two theoretical explanations of SR: the preferred tempo (Temperley, 1963) and the resonance theory (Large, 2008) explanations.

The results confirmed four findings of the earlier studies. First, most participants do report that they experience SR. In the current study all participants reported experiencing SR. While this could be due to the musical training of many of the participants, it supports the position of SR as a robust phenomenon that a large part of the

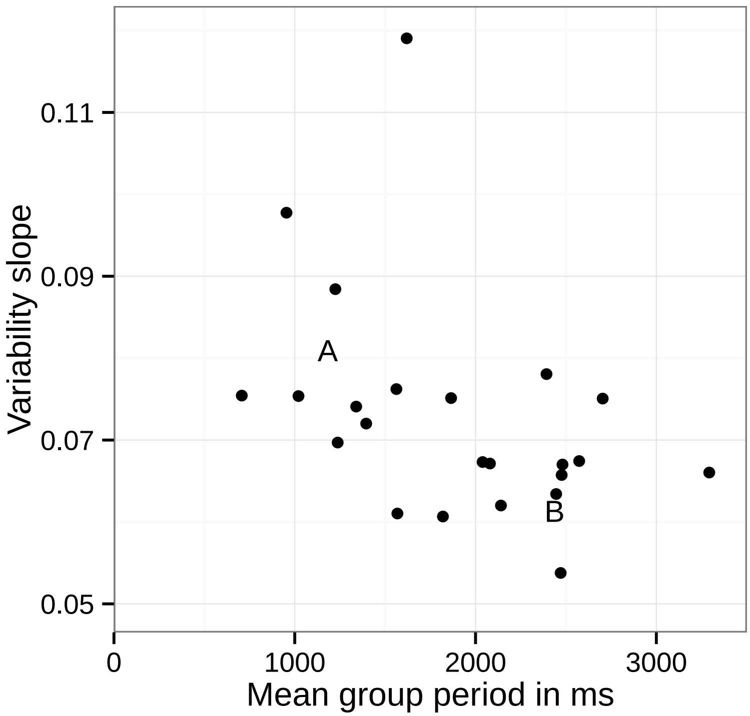


Figure 9: Variability slope plotted against mean group period for each participant. Participant A and B from Figure 4 are marked by the corresponding letter.

population experiences.

Second, the experience of SR is strongly affected by the tempo of the sound sequence, as shown by a strong negative correlation between sound sequence ISI and reported group size. Participants were highly consistent with regard to the group sizes reported at particular ISIs; the probability of reporting the same group size on any two trials with the same ISI averaged .69. Putting this into perspective, the probability of choosing the same response on two different trials would be 0.24 if choosing randomly according to the group probabilities in Table 1. Participants were also comparably consistent across ISIs, that is, although the impression of SR is strongly affected by tempo, consistency of responses is not.

Third, all group sizes are not reported with equal frequency. Groups of two, four, and eight were reported most often, followed by groups of three and six. Groups of five and seven were reported on less than 1% of the trials. This is the ordering one would expect from a Western music-theoretical perspective (van Noorden & Moelants, 1999). To date, no SR study has been conducted in a country with a non-Western musical tradition. It remains to be determined to what degree SR is affected by cultural factors. As culture has been shown to play an important role in rhythm perception (Hannon, Soley, & Ullal, 2012), a prediction is that groups of five and seven would be more commonly reported by participants accustomed to odd meters prevalent in, e.g., the traditional music of the Balkan Peninsula.

Fourth, when the tempo of the sequences is sufficiently slow, participants do not experience SR. This slower limit of SR, while not probed by Vos, was estimated to an ISI of 1500 ms by Bolton. The current study found no such sharp limit but instead found large inter-individual variability. However, at an ISI of 1500 ms more than half the trials resulted in no experienced SR, comparable to Bolton's figure.

The current study focused on how the experience of SR varies as a function of tempo but many other factors might also be influential. Time perception differs depending on the pitch of the stimulus (Hove, Marie, Bruce, & Trainor, 2014), so it is possible that pitch affects the experience of SR. Another factor that is likely to influence SR is the task instructions, even though the comparability of the results from the current study with the previous studies by Bolton (1894) and Vos (1973) shows that SR is at least somewhat robust to variation in the task instructions. That said, differences in how subjects approach the task might still heavily influence the experience of SR. The study by Nozaradan et al. (2011) is already an example of this, as whether participants were asked to actively imagine a subjective duple meter or not influenced their subsequent EEG readings. It remains an open question to what degree, and in what way, the experience of SR depends on the task instructions and on qualities of the stimulus such as amplitude, pitch and timbre.

As an aside, my experience is that perceived groupings can be changed somewhat at will, for example, listening to a monotone sequence with an ISI of 600 ms I often start out hearing an accent on every second sound. By focusing, however, I can switch the accent to every fourth sound. If SR can generally be affected by such top-down control it would not imply that SR is a purely top-down phenomena. Rather, such a finding would resonate with research regarding visual illusions, such as the Necker cube, known to be affected by both bottom-up and top-down processes (Long & Toppino, 2004).

4.1 Explanations of Subjective Rhythmization

The literature offers two explanations for SR: the preferred tempo (Temperley, 1963) and the resonance theory (Large, 2008). Resonance theory is a dynamical systems framework for modeling rhythm perception and production. The resonance theory explanation of SR is based on the notion that an isochronous sequence, in addition to entraining oscillatory units responsive to the fundamental period, entrains subharmonic oscillators, thus producing the subjective accents characteristic of SR (Large, 2008). This explanation of SR gives rise to three predictions: (1) the mean group period of the reported groupings should fall between the slower limit of SR and the slower limit of rhythm production, (2) the relation between the size of the reported grouping and ISI of the sound sequence should follow a power relation, and (3) a participant's mean group period should relate to tapping performance at slow tempi. Within the resonance theory framework, (1) follows from assuming a slower limit of rhythm perception, with the mean group period being seen as a proxy variable for this limit; (2) follows from a slower limit of rhythm perception limiting the highest possible grouping that can be perceived for any given ISI; (3) follows from the assumption that rhythm perception and rhythm production both share the same underlying mechanism.

The results of the present study are in line with the predictions developed on the basis of resonance theory. The results do not support the preferred tempo explanation, whereby the mean group period should be close to participants' spontaneous motor tempo. Instead, the mean group period was closer to the participants' slow motor tempo (see Figure 8), in line with prediction (1).

Resonance theory assumes that rhythm perception and rhythm production share a common neural substrate. Thus, there should be a relation between a participant's performance in rhythm perception tasks and rhythm production tasks. The present study did indeed find such a relation as there was a correlation between what a participant reported in the SR task and her timing performance in the tapping task. Specifically, participants that reported large groupings in the SR task tended to have smaller timing variability when tapping at a slow tempo relative to tapping at a moderate tempo. From a resonance theory perspective this is explained by that the slower limit of rhythm perception influences both timing variability at a slow tempo and what grouping is perceived in an SR task.

The relation between the reported group sizes and the ISI of the sound sequences was found to follow a power relation closely (see Figure 7). Resonance theory explains this by that of the group size perceived at a certain ISI depends on the participant's slower limit of rhythm perception. That slower limit governs the ISI at which the participant starts to experience a given group size. The relation between group size and ISI was well captured by the expression $g \sim \frac{k}{T^a}$, where g is the perceived grouping, T is the ISI of the sequence, and k and a are constants. The results are not compatible with a sharp slower limit of rhythm perception. A sharp limit would imply that a participant should experience a grouping of two at half the ISI of the slower limit, a grouping of four at a fourth of the slower limit, etc. Such behavior would result in $a = 1$, with k equal to the slower limit. The estimate of the current study was $a = 0.53$ implying that participants tend to report smaller group sizes at faster tempi compared to what a sharp limit predicts. This can be accommodated within a resonance theory framework by treating rhythm perception as an ability that, instead of having a sharp limit, deteriorates gradually as the tempo gets slower.

Overall, the current results are well explained by the resonance theory of rhythm perception. This is not to say that other models could not explain the phenomena of SR. However, the current results do suggest that any such account would need to include both a slow limit of rhythm perception and a close connection between rhythm perception and rhythm production. Subjective rhythmization is closely related to meter perception; the ability of subjects to experience widely different accent patterns while listening to the same sequences draws attention to the difference between a rhythm sequence as stimulus and as percept. Of course, it is not uncommon that different people experience the same piece of music differently. What is perhaps surprising is that, even while listening to the most simple monotone metronome sequence, what is experienced is *still* in the ear and mind of the listener.

Acknowledgement

This research was supported by the Linnaeus environment Thinking in Time: Cognition, Communication and Learning, Swedish Research Council grant number 349-2007-8695. I thank Geoffrey R. Patching and Christian Balkenius for useful comments on various aspects of this work.

References

Andrews, B. (1905). Auditory tests. *The American Journal of Psychology*, *16*(3), 302–326. Retrieved from http://www.jstor.org/stable/1412381

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature reviews. Neuroscience*, *7*(5), 358–66. doi: 10.1038/nrn1888

Bååth, R. (2011). Construction of a Low Latency Tapping Board. *LUCS minor*, *17*. Retrieved from http://www.sumsar.net/papers/ICMPC_2012_rasmus_baath.pdf

Bolton, T. (1894). Rhythm. *The American Journal of Psychology*, *6*(2), 145–238. Retrieved from http://www.jstor.org/stable/1410948

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "Ticktock" of our Internal Clock : Direct Brain Evidence of Subjective Accents in Isochronous Sequences. *Psychological Science*, *14*(4), 362–366. doi: 10.1111/1467-9280.24441

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). Academic Press.

Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *Journal of experimental psychology. Human perception and performance*, *38*(3), 543–8. doi: 10.1037/a0027225

Hove, M. J., Marie, C., Bruce, I. C., & Trainor, L. J. (2014). Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proceedings of the Na-*

tional Academy of Sciences of the United States of America, 1–6. doi: 10.1073/pnas.1402039111
Kirnberger, J. P. (1776). *Die Kunst des reinen Satzes in der Musik: aus sicheren Grundsätzen hergeleitet und mit deutlichen Beyspielen erläutert*. Decker.
Large, E. (2000). On synchronizing movements to music. *Human Movement Science*, *19*(4), 527–566. doi: 10.1016/S0167-9457(00)00026-9
Large, E. (2008). Resonating to musical rhythm: Theory and experiment. In S. Grondin (Ed.), *The psychology of time* (pp. 189–231). Emerald Group Publishing. doi: 10.1016/B978-0-08046-977-5.00006-5
Large, E., & Jones, M. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159. doi: 10.1037/0033-295X.106.1.119
Large, E., & Kelso, J. A. S. (2002). Tracking simple and complex sequences. *Psychological Research*, 3–17. doi: 10.1007/s004260100069
Large, E., & Kolen, J. (1994). Resonance and the perception of musical meter. *Connection science*, *6*(2), 177–208. doi: 10.1080/09540099408915723
Long, G. M., & Toppino, T. C. (2004, September). Enduring interest in perceptual ambiguity: alternating views of reversible figures. *Psychological bulletin*, *130*(5), 748–68. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15367079 doi: 10.1037/0033-2909.130.5.748
McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *Journal of experimental psychology. General*, *135*(3), 348–67. doi: 10.1037/0096-3445.135.3.348
Moelants, D. (2002). Preferred tempo reconsidered. In *Proceedings of the 7th international conference on music perception and cognition* (pp. 580–583). Sydney. Retrieved from http://www.ipem.ugent.be/dirkmoelants/pub/Moelants_PreferredTempoReconsidered.pdf
Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience*, *31*(28), 10234–40. doi: 10.1523/JNEUROSCI.0411-11.2011
Patel, A. (2008). *Music, language, and the brain*. Oxford University Press.
R Core Team. (2012). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org
Repp, B. H. (2005). Sensorimotor synchronization: A review of the

tapping literature. *Psychonomic Bulletin & Review*, *12*(6), 969–992. doi: 10.3758/BF03206433

Repp, B. H. (2006). Rate Limits of Sensorimotor Synchronization. *Advances in Cognitive Psychology*, *2*(2), 163–181. doi: 10.2478/v10053-008-0053-9

Schaefer, R. S., Vlek, R. J., & Desain, P. (2011). Decomposing rhythm processing: electroencephalography of perceived and self-imposed rhythmic patterns. *Psychological research*, *75*(2), 95–106. doi: 10.1007/s00426-010-0293-4

Scheirer, E. D. (1998). Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America*, *103*(1), 588–601. doi: 10.1121/1.421129

Temperley, N. (1963). Personal Tempo and Subjective Accentuation. *The Journal of general psychology*, *68*(2), 267–287. doi: 10.1080/00221309.1963.9920534

van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, *28*(1), 43–66. doi: 10.1076/jnmr.28.1.43.3122

Vlek, R. J., Schaefer, R. S., Gielen, C. C. a. M., Farquhar, J. D. R., & Desain, P. (2011). Sequenced subjective accents for brain-computer interfaces. *Journal of neural engineering*, *8*(3), 036002. doi: 10.1088/1741-2560/8/3/036002

Vos, P. (1973). *Waarneming van metrische toonreeksen*. Nijmegen: Stichting Studentenpers.

Paper V

Bååth, R. (2015). Estimating the distribution of sensorimotor synchronization data: A Bayesian hierarchical modeling approach. *Behavior Research Methods*. doi:10.3758/s13428-015-0591-2

Behav Res
DOI 10.3758/s13428-015-0591-2

Estimating the distribution of sensorimotor synchronization data: A Bayesian hierarchical modeling approach

Rasmus Bååth[1]

Abstract The sensorimotor synchronization paradigm is used when studying the coordination of rhythmic motor responses with a pacing stimulus and is an important paradigm in the study of human timing and time perception. Two measures of performance frequently calculated using sensorimotor synchronization data are the average offset and variability of the stimulus-to-response asynchronies—the offsets between the stimuli and the motor responses. Here it is shown that assuming that asynchronies are normally distributed when estimating these measures can result in considerable underestimation of both the average offset and variability. This is due to a tendency for the distribution of the asynchronies to be bimodal and left skewed when the interstimulus interval is longer than 2 s. It is argued that (1) this asymmetry is the result of the distribution of the asynchronies being a mixture of two types of responses—predictive and reactive—and (2) the main interest in a sensorimotor synchronization study is the predictive responses. A Bayesian hierarchical modeling approach is proposed in which sensorimotor synchronization data are modeled as coming from a right-censored normal distribution that effectively separates the predictive responses from the reactive responses. Evaluation using both simulated data and experimental data from a study by Repp and Doggett (2007) showed that the proposed approach produces more precise estimates of the average offset and variability, with considerably less underestimation.

Keywords Bayesian statistics · Sensorimotor synchronization · Hierarchical models · Finger tapping

Electronic supplementary material The online version of this article (doi:10.3758/s13428-015-0591-2) contains supplementary material, which is available to authorized users.

✉ Rasmus Bååth
rasmus.baath@lucs.lu.se

1 Department of Philosophy, Lund University, Box 192, 22100 Lund, Sweden

The experimental study of human timing and time perception has a long history in psychology, with the sensorimotor synchronization (SMS) task being one of the most important experimental paradigms (Roeckelein, 2008). Following Stevens (1886), this task requires a participant to produce periodic movements synchronized to a regular pacing stimulus such as a metronome (Schulze, 1992). Sensorimotor synchronization is often studied in a musical context, because the ability to engage in SMS is central to musical activities, especially in ensemble music, in which many musicians are required to follow the same rhythm and coordinate their movements together (Repp, 2006). Yet, SMS performance is also a relevant measure in many other fields. For example, SMS performance has been shown to correlate with personality traits (Forsman, Madison, & Ullén, 2009) and measures of intelligence for both children (Corriveau & Goswami, 2009) and adults (Madison et al., 2009). It is also correlated with performance in other experimental paradigms related to timing, such as simple reaction time (Holm, Ullén, & Madison, 2011) and eye blink conditioning (Green, Ivry, & Woodruff-Pak, 1999).

What is most often measured in an SMS task is the stimulus-to-response asynchronies—that is, the offset of a participant's responses from the stimulus onsets, where a negative asynchrony indicates that a participant's response preceded the stimulus (Repp, 2005). The two basic parameters estimated in an SMS task are the constant error—the average deviation from the target stimuli—and the timing variability of the asynchronies. The most straightforward and common

Published online: 01 May 2015

way to estimate these parameters is to calculate the sample mean and the sample standard deviation (*SD*) of the asynchronies. However, it is shown here that this approach results in negatively biased estimates of constant error and timing variability; that is, these parameters are generally underestimated by the sample mean and sample *SD*. The concern is with the nonnormal distribution of timing asynchronies at slow tempi and the failure of moment estimators to incorporate task-relevant information.

The present article describes a Bayesian model with which to estimate the constant error and timing variability in SMS tasks that do not suffer from the problems associated with using the sample mean and *SD*. The text is organized as follows. First the distribution of SMS data is reviewed, and it is shown that the distribution is approximately normal when the tempo of the pacing stimuli is moderate, but that the distribution becomes increasingly nonnormal when the interstimulus interval (ISI) exceeds 2 s. It is argued that this is due to a tendency for participants to react to the stimulus onset and that, because of this, SMS data are modeled better using a right-censored normal distribution. A Bayesian model is then developed that uses a censored normal distribution to model the distribution of timing asynchronies. The model is developed in two stages, first as a basic nonhierarchical model, and then as a fully hierarchical model. Finally, the model is compared with the traditional methods of using the sample mean and *SD* to estimate constant error and timing variability. This is done using both simulated data and the experimental data from a study by Repp and Doggett (2007). It is shown that using a censored normal distribution to model timing asynchronies results in considerably less bias than using traditional methods. Furthermore, using a hierarchical Bayesian approach outperforms both traditional methods and a nonhierarchical Bayesian model with regard to accuracy. This article advances the study of human timing and time perception by giving researchers a better tool to measure SMS performance. A further advance is that the Bayesian methods proposed facilitate analyzing SMS data when the ISI exceeds 2 s.

The distribution of sensorimotor synchronization data

In a typical SMS task, a participant is asked to produce responses in time with a recurring, isochronous (equally spaced in time) stimulus sequence. The commonly employed stimuli are sequences of equally spaced auditory tones that the participant synchronizes to by tapping a button using the index finger, although there are many variations of this basic experimental procedure (Repp, 2005). Stimulus-to-response asynchronies, that is, the time offsets between the stimulus onset and the participant's timed response, are of primary interest in SMS tasks (Repp & Su, 2013). Such asynchronies can be positive or negative, where a negative asynchrony indicates that the corresponding response preceded the stimulus, and a positive asynchrony indicates that the corresponding response followed the stimulus onset.

Under many circumstances the distribution of the asynchronies is approximately normal (Chen, Ding, & Kelso, 1997; Mates, Müller, Radil, & Pöppel, 1994; Moore & Chen, 2010), but it is rarely normal when the ISI of the pacing stimuli is longer than 2 s (cf. Mates et al., 1994, and Miyake, Onishi, & Pöppel, 2004). An example of timing asynchrony distributions at different ISIs is shown in Fig. 1 using data from a study by Bååth and Madison (2012), in which 30 participants were asked to tap with their index finger in synchrony with a pacing tone sequence. In Fig. 1A, the asynchrony distributions for a representative participant are shown; for ISIs of 600 ms and 1,200 ms, the distributions can be seen to be heap shaped and symmetric. The central tendencies of the distributions are not centered around zero (i.e., at the onset of the pacing tone) but are slightly negative, a well-known phenomenon termed the *mean negative asynchrony* (Aschersleben, 2002). At ISIs of 1,800 ms and above, a visible peak from 100 to 200 ms makes the distribution left skewed, or even bimodal. This peak coincides with where auditory reaction time responses to the stimulus onset would be likely to occur (Gottsdanker, 1982). These deviations from normality not only can be seen by visual inspection, but a Shapiro–Wilk normality test is also rejected, with $p < .01$ as the

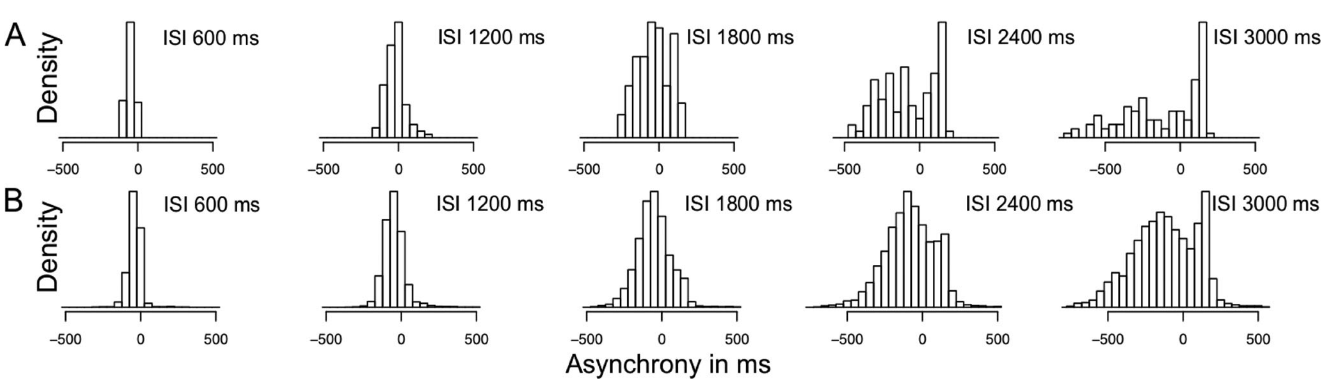


Fig. 1 Tone-to-response asynchrony distributions (**A**) for a single participant and (**B**) for all participants in Bååth and Madison (2012)

rejection criterion, for ISIs of 1,800 ms ($p = .004$), 2,400 ms ($p < .001$), and 3,000 ms ($p < .001$), but not for ISIs of 600 ms ($p = .62$) and 1,200 ms ($p = .90$). This was even after using Tukey's (1977) boxplot method to label and remove outliers. A similar pattern can be seen when looking at the distributions of the asynchronies from all participants. The percentages of the 30 participants who produced asynchronies at different ISI levels that resulted in rejection by a Shapiro–Wilk test were 7% (600 ms), 3% (1,200 ms), 7% (1,800 ms), 27% (2,400 ms), and 47% (3,000 ms). This resonates with the finding of Mates et al. (1994) that asynchronies produced at ISIs above 1,800 ms tend to be increasingly nonnormal.

According to Repp and Doggett (2007), the distribution of timing asynchronies departs from normality at ISIs longer than 2s because participants occasionally overshoot the target stimuli and instead react to them. At moderately fast ISIs, around 600 ms (see Fig. 1), this rarely happens, because the asynchronies tend to be within 100 ms of the stimulus onset. At longer ISIs the variability of the asynchronies increases, and at an ISI of 2 s many asynchronies are smaller than –300 ms but rarely exceed 300 ms. This asymmetry results in a skewed, bimodal asynchrony distribution with a long left tail and a short right tail at long ISIs.

If Repp and Doggett's (2007) interpretation is correct, the observed distribution of timing asynchronies at ISIs greater than 2 s is then a mixture of predictive asynchronies, which generally are the responses of interest, and contaminating reactive asynchronies. The mean and *SD* of the whole asynchrony distribution is then uninformative, and methods should be used to separate out these distributions so as to estimate the mean and *SD* of the predictive distribution adequately. If this is not done, different phenomena are measured at long and short ISIs: At short ISIs, a participant's ability to synchronize to a pacing sequence is being measured, whereas at longer ISIs, a mixture of reaction time and timing ability may be measured. The data from an SMS task can then be thought of as being sampled from two different distributions. At short ISIs, the predictive timing asynchronies can be seen as samples from a normally distributed random variable X_P. At longer ISIs, at which the timing variability is large enough that participants sometimes make reactive responses, the timing asynchronies can be seen as samples from a random variable $X_A = \min(X_P, X_R)$, where the random variable X_R is distributed as a reaction time distribution. There are many proposed models for the distribution of reaction time responses, all of which are right skewed and, for practical purposes, left bounded (Ulrich & Miller, 1994; van Zandt, 2000). For the present purpose of modeling the predictive responses, the distribution of the reactive responses could be assumed to be any of those—for example, the ex-Gaussian distribution.

An illustration of the distribution of X_A is shown in Fig. 2. The rationale for taking the *minimum* of X_P and X_R is that these two random variables can be thought of as representing two

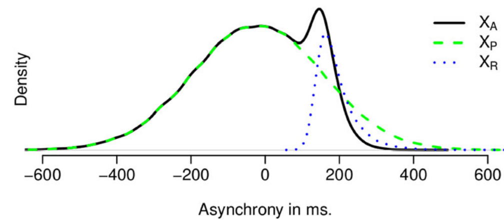


Fig. 2 Theoretical distribution of timing asynchronies (X_A, solid line), modeled as a combination of a distribution of predictive responses (X_P, dashed lines) and reactive responses (X_R, dotted line)

independent processes that can either trigger a response (e.g., a buttonpress or drum stroke), with the response being initiated by whichever process triggers first. For example, if 300 ms is an outcome of X_P and 200 ms is an outcome of X_R, then the outcome of X_A, which represents a participant's response, would be min(300 ms, 200 ms) = 200 ms. Because X_A is the minimum of X_P and another random variable, X_A will always have a shorter right tail than X_P. Note that there are other combinations of X_P and X_R that do not result in this behavior—for example, the average of X_P and X_R, or a mixture distribution constructed from X_P and X_R. The average of X_P and X_R would instead result in a distribution with a positively shifted mean, as compared to X_P. This would not be in agreement with the well-established finding that the mean asynchrony from timing tasks tends to be slightly negative (Aschersleben, 2002). The mixture distribution defined by $X_{mix} = wX_P + (1 - w)X_R$ would not resemble X_A in that, depending on the location of X_R and the mixture weight w, the distribution of X_{mix} could have a longer, rather than a shorter, right tail than X_P.

In an SMS task, the interest is in the distribution of X_P, and whereas X_P is possible to measure at short ISIs, at long ISIs X_P can be considered a latent variable. Using the sample mean and *SD* to estimate the distribution of X_P is then problematic because reactive responses may confound estimates of constant error and timing variability, resulting in considerable negative bias. This happens because the distribution of X_A has a shorter right tail than the distribution of X_P. Due to the asymmetry of X_A, a sample mean estimate will be biased toward a negative mean asynchrony, and a sample *SD* estimate will be smaller than the actual *SD* of X_P, due to the right tail of the distribution of X_A being less spread out than the distribution of X_P. In other words, using moment estimators will make it appear that participants are responding earlier and more accurately than they actually are.

At first, it might seem that the distributions of X_P and X_R could be separated using a standard mixture-model approach, by modeling an outcome as being generated by first selecting one of the underlying distributions and then using that distribution to generate the outcome. This approach does not consider that the distribution of X_A was the result of taking the *minimum* of two random variables, and so will not result in a consistent estimator of the parameters underlying X_A. Another approach would be to implement the fully generative model shown in Fig. 2. This would, however, require specifying not only the distribution for X_P, but also that for X_R. Although many possible distributions could be assumed for reactive responses (van Zandt, 2000), it is not known whether these are applicable for the reactive responses in an SMS task, and misspecification of this distribution would impact the parameter estimates of both X_P and X_R.

One way of estimating the distribution of X_P is by noticing that the distribution of X_R does not depend on the ISI and is left-bounded by participants' ability to react to the pacing tones. Therefore, the distribution of X_P can be retrieved by modeling the asynchronies as coming from a right-censored normal distribution (see Fig. 3) with the censoring threshold being selected to exclude reactive taps—for example, set to 100 ms. That is, all asynchronies below 100 ms would be assumed to be direct observations of X_P, whereas asynchronies above 100 ms would also be assumed to be observations of X_P, but the actual asynchrony would be disregarded and the only information retained would be that the observation fell in the range (100, ∞) ms. The rest of this article will focus on a model implementing these assumptions and Bayesian estimation of the parameters of this model.

Bayesian modeling of sensorimotor synchronization data

Bayesian methods of data analysis are becoming increasingly popular in psychology and other fields (Andrews & Baguley, 2013; Kruschke, 2011b). The rationale for using Bayesian modeling is the ease with which nonnormal data can be modeled, hierarchical dependencies in the data can be specified, and prior knowledge, such as task-specific constraints, can be included (Kruschke, 2011a). The model described here is Bayesian and is implemented using Markov chain Monte Carlo (MCMC) methods. Both the terminology and the philosophy of Bayesian statistics are different from those of classical frequentist statistics. The following sections assume some acquaintance with Bayesian statistics and MCMC methods, and many good text cover these topics—for example, the books by Kruschke (2011a), Lunn, Jackson, Best, Thomas, and Spiegelhalter (2012), and Lee and

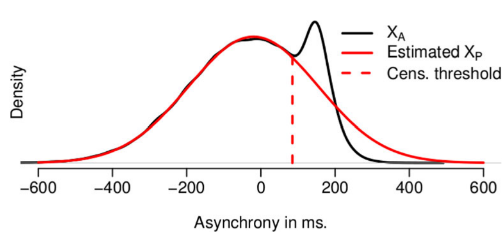


Fig. 3 Schematic diagram of the fit of a right-censored normal distribution to the theoretical distribution of timing asynchronies shown in Fig. 2

Wagenmakers (2014). Because the data from an SMS experiment often are hierarchically organized—that is, they consist of many participants performing multiple trials—the model is presented in a hierarchical as well as in a nonhierarchical version. The advantages of hierarchical modeling are described well by Gelman and Hill (2006).

A number of Bayesian models are designed to analyze reaction time data (Craigmile, Peruggia, & van Zandt, 2010; Farrell & Ludwig, 2008; Rouder, Sun, Speckman, Lu, & Zhou, 2003), and these models and the model presented in this article have much in common, in that they are hierarchical, deal with timing data, and model data as coming from nonnormal distributions. Differences include the types of distributions used and what type of task-related information is incorporated in the model. The present model makes two important assumptions. First, the asynchronies are treated as independent observations. This is strictly not true, because asynchronies tend to be autocorrelated when they are considered as forming a time series (Chen et al., 1997). Estimates of constant error and timing variability do not, then, constitute an exhaustive description of the underlying data. Still, these measures are two of the most common in the literature (Mates et al., 1994), and together they form a useful summary of timing performance. Second, the asynchronies are assumed to be from trials with the same ISI. In an experimental setup in which many ISI levels are used, the constant error and timing variability have to be estimated separately for each ISI level.

Below, the nonhierarchical version of the model is first described, which is then extended into a fully hierarchical model. The model uses priors that are vague and noninformative, except with regard to certain task-specific constraints that are used to inform the priors. How to extend the model in order to use more informative priors and how to model interresponse intervals instead of asynchronies is described in the supplementary text, and software implementing the model is freely available at https://github.com/rasmusab/bayes_timing. The software is implemented using the R statistical environment (R Development Core Team, 2012) and the JAGS framework (Plummer, 2003). The JAGS framework takes a model definition and automatically generates a sampling scheme using Gibbs sampling. Because the technical details regarding sampling schemes are handled by JAGS, it is relatively straightforward to extend and modify the models given below—for example, by changing the priors or the distributional assumptions.

The nonhierarchical model

The timing asynchronies (Y) are modeled as coming from a normal distribution in which asynchronies exceeding a threshold c are assumed to be right censored. Strictly, it is the values that are censored, and not the actual distribution, but the procedure of first censoring values above a threshold and then modeling them as being normally distributed could be seen as defining a right-censored normal distribution with three parameters: μ, σ, and c. In the present model, the threshold c is fixed at a value that is assumed to censor asynchronies that could be the results of reactive responses. A conservative value for c would be, for example, 100 ms, which is a reasonable tradeoff between censoring reactive responses and maintaining predictive responses.

The prior distributions for both μ and σ use noninformative Jeffreys (1946) priors, but with additional constraints imposed by the SMS task. In an SMS task, a single asynchrony cannot be farther away from a stimulus onset time than half the ISI. For example, if the ISI is 600 ms and a response is registered 250 ms before a stimulus onset, this will be encoded as a –250-ms asynchrony, but if that response was registered 375 ms before the same stimulus onset, it would instead be encoded as a 225-ms asynchrony with respect to the preceding stimulus onset. If asynchronies are bounded by the interval [–ISI/2, ISI/2] ms, then μ is also bounded by this interval. The Jeffreys prior on μ is an unbounded uniform distribution (Lunn et al., 2012), but when using the task-specific information, the prior becomes a uniform distribution bounded on the interval [–ISI/2, ISI/2] ms. Given that the difference between any two asynchronies can be at most the length of the ISI, σ can be at most ISI/2 ms.[1] The Jeffreys prior on σ is an unbounded uniform distribution on log(σ) (Lunn et al., 2012), but using the present task-specific constraints bounds the distribution on the interval [log(1), log(ISI/2)] ms. The lower bound on log(σ) is set to log(1) because the probability that the *SD* of a participant's asynchronies would be close to 1 ms is negligible. The full specification of the nonhierarchical model is then:

$$Y_i \sim \text{Right-Censored-Normal}(\mu, \sigma, c),$$
$$\mu \sim \text{Uniform}(-\text{ISI}/2, \text{ISI}/2),$$
$$\log(\sigma) \sim \text{Uniform}[\log(1), \log(\text{ISI}/2)],$$

where Y_i is the ith asynchrony.

Point estimates of the parameters μ and σ can be calculated by taking the mean or the median of their respective posterior distributions (Robert, 2007). These estimates can then be used as a "drop-in" replacement for the sample mean and *SD* estimates of constant error and timing variability. This approach is useful in the case in which a researcher wants to avoid the bias associated with using the sample mean and *SD* but prefers a classical analysis of the point estimates rather than a fully Bayesian analysis.

It should be noted that a point estimate generated using the Bayesian model above and a maximum likelihood approach would be very similar. This is because the maximum a posteriori estimate from a Bayesian model with flat priors and a maximum likelihood estimate are identical (Hastie, Tibshirani, & Friedman, 2009). When only a point estimate is required, it can therefore be convenient to use a maximum likelihood method, which is computationally more efficient and, perhaps, a better-known method than Bayesian estimation. Ulrich and Miller (1994) have described a maximum likelihood approach for fitting a right-truncated normal distribution that is applicable in this case. This method has also been implemented and is available at https://github.com/rasmusab/bayes_timing.

[1] This maximum *SD* would occur in the unlikely event that half of the asynchronies were –ISI/2 ms and half were ISI/2 ms.

The hierarchical model

Hierarchical modeling is an elegant solution to the problem of analyzing a data set with repeated measurements (Kruschke, 2011a). It is an increasingly used technique in psychological research and is variously known as *hierarchical modeling*, *multilevel modeling*, or *mixed modeling* (Baayen, Davidson, & Bates, 2008). One of the reasons to use a hierarchical model is to better describe data that have a multilevel structure (Gelman & Hill, 2006), and the typical SMS experiment is inherently multileveled. Timing responses can be assumed to be related within a trial, between trials, within a participant, and between participants. Furthermore, it is reasonable to assume that relations exist between the timing responses at different tempi; for example, a participant who produces highly variable asynchronies at an ISI of 500 ms will probably produce highly variable asynchronies at longer ISIs. The hierarchical model described below does not take all of the possible multilevel relations into account. Asynchronies from different trials produced by the same participant are assumed to have the same distribution; that is, no training or exhaustion effect is assumed to exist. Although it is well known that timing variability increases as the tempo gets slower, this relation does not seem to be linear (Grondin, 2012; Repp & Su, 2013), and it is not known whether it follows any simple function. Therefore, the relationship between the asynchronies produced at different tempi is not part of the model.

What this model adds over the nonhierarchical version is that the relation between participants' timing performance is modeled, allowing measurements made on all participants to inform the parameter estimates of single participants. The hierarchical formulation also facilitates investigating individual differences as constant error and timing variability are estimated at both the individual and group levels. The mean of the censored normal distribution of the jth participant, μ_j, is modeled as coming from a normal distribution with mean μ_μ and *SD* σ_μ. The subscript μ is used to indicate that μ_μ and σ_μ are hyperparameters of the prior distribution on μ_j. The *SD* of the censored normal distribution, σ_j, is modeled as coming from a log-normal distribution with the parameters μ_σ and σ_σ (as was proposed by Lunn et al., 2012). Note that these parameters are not the mean and *SD* of the log-normal distribution. Through reparameterization, this distribution can be described by its mean, m_σ, and *SD*, s_σ (Limpert, Stahel, & Abbt, 2001). The rationale for using the log-normal distribution is that it allows for modeling participants' timing variability on the same scale as σ_j, which makes the hyperparameters m_σ and s_σ easy to interpret; the posterior distributions of m_σ and s_σ can be directly used to estimate the population's mean timing variability and the *SD* of the population's mean timing variability. A popular choice for the prior distribution on a variability parameter is otherwise $1/\sigma^2 \sim$ Gamma(., .) (Gelman, 2006), but the problem with using this prior is that it is not defined on the scale of the *SD* of the asynchronies, which makes it hard to interpret the posterior distributions of its parameters. This is fine in the case in which the constant error is the main interest and the *SD* of the asynchronies is considered a nuisance parameter. When analyzing SMS data, however, it is often the case that timing variability is the main interest (Repp, 2005).

The prior distributions on μ_μ and m_σ are the same as those used for the parameters μ and σ in the nonhierarchical model. The prior distributions for the *SD* parameters σ_μ and s_σ are uniform distributions (as was proposed by Gelman, 2006) with their left boundaries at 0, and with task-related constraints defining the right boundaries as for μ and σ in the nonhierarchical model. The left boundary for σ_μ is set to ISI/2 because this is the largest possible between-participants mean asynchrony *SD*, since all μ_js are bounded within the interval [–ISI/2, ISI/2]. The largest possible within-participants asynchrony *SD* is ISI/2, which implies that the largest possible value of σ_σ, the between-participants *SD* of the asynchrony *SD*, is ISI/4. The full specification of the hierarchical model is then

$$
\begin{aligned}
&Y_{ij} \sim \text{Right-Censored-Normal}(\mu_j, \sigma_j, c_j),\\
&\mu_j \sim \text{Normal}(\mu_\mu, \sigma_\mu),\\
&\sigma_j \sim \text{Log-Normal}(\mu_\sigma, \sigma_\sigma),\\
&\mu_\mu \sim \text{Uniform}(–\text{ISI}/2, \text{ISI}/2),\\
&\sigma_\mu \sim \text{Uniform}(0, \text{ISI}/2),\\
&\mu_\sigma \sim \log(m_\sigma) - \sigma_\sigma^2/2,\\
&\sigma_\sigma \sim \sqrt{\log(s_\sigma^2/m_\sigma^2 + 1)},\\
&\log(m_\sigma) \sim \text{Uniform}[\log(1), \log(\text{ISI}/2)],\\
&s_\sigma \sim \text{Uniform}(0, \text{ISI}/4),
\end{aligned}
$$

where Y_{ij} is the *i*th asynchrony of the *j*th participant. After Kruschke (2011a), a graphical model diagram of this model is shown in Fig. 4.

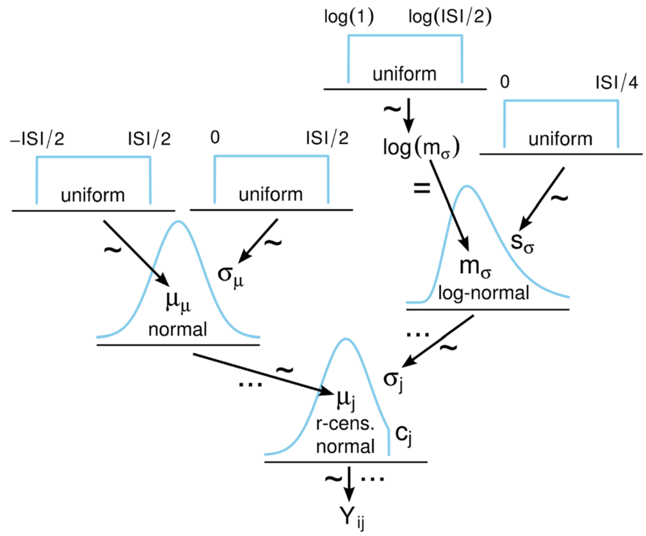


Fig. 4 Model diagram showing the specification of the fully hierarchical model. Here, Y_{ij} is the *i*th asynchrony of the *j*th participant

An example using the hierarchical model

The data from Bååth and Madison (2012) were analyzed using the hierarchical model. The study included responses from 30 participants who synchronized finger taps to isochronous tone sequences with five different ISIs: 600, 1,200, 1,800, 2,400, and 3,000 ms. Because the data include repeated measurements, with each participant producing many asynchronies at each ISI level, the hierarchical model was fitted to the data separately for each ISI level.

After a Bayesian model has been fitted, the estimated parameters can be investigated in many ways. Depending on where the interest lies, the parameters can be examined at the participant level (σ_j and μ_j) or the group level (m_σ and μ_μ). If individual differences are of interest, the variance components s_σ and σ_μ can be inspected, because they index the degree to which the participant-level parameters differ. Because the model is fully Bayesian, estimates for all of the parameters are readily available after the model has been fitted, including reliability measures in the form of credible intervals.

Because tapping variability is often of interest in SMS studies (Repp, 2005), the group-level timing variability and the variability on the participant level were investigated further. Figure 5 shows the point estimates and 95% credible intervals, extracted from the model, of the group mean asynchrony *SD* (the m_σ parameter in the model) and the participant-level asynchrony *SD*s for three participants (σ_1, σ_2, and σ_3) at the five ISI levels. The group mean *SD* increases almost linearly as a function of ISI, but we also see that there are large differences between the three participants. Using the posterior distribution of the fitted model, it was now possible to investigate any relation between the parameters. For example, there was an 89% probability that Participant 1 had a lower timing variability than Participant 2 at an ISI of 1,800 ms, and there was a 99% probability that the increase in timing variability between the ISIs of 1,800 and 2,400 ms was larger for Participant 1 that for Participant 2. A script that fully replicates these calculations and the fitting of the model can be found at https://github.com/rasmusab/bayes_timing, together with the full data set from Bååth and Madison (2012).

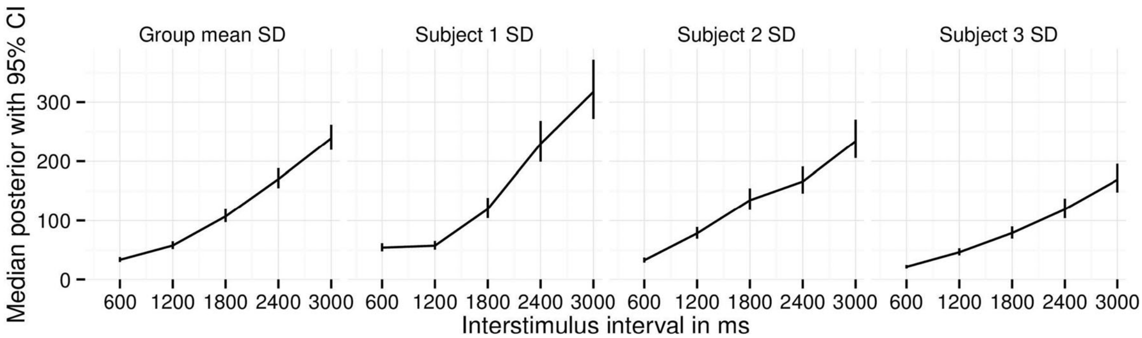


Fig. 5 Point estimates with 95% credible intervals of the group timing variability (measured as the asynchrony *SD*s) and the participant timing variability for three participants from the hierarchical model, fitted to the data from Bååth and Madison (2012)

Evaluation of the model

In order to evaluate the model, it was applied to both simulated data and the experimental data from a study by Repp and Doggett (2007). The estimates generated by the model were compared with the common estimates of constant error and timing variability, the sample mean, and the sample *SD*. These estimators will be referred to as the *moment estimators*. The Bayesian model can be used to generate point estimates of the constant error and timing variability for further analysis, or the hyperparameters of the model can be used directly to draw inferences. In order to be able to compare the Bayesian model with the moment estimators, evaluations were made using the first approach of analyzing estimated point values. There are different methods for calculating point values from the marginal posterior distributions of a Bayesian model (see Robert, 2007). Here the median of the marginal posterior distribution of the parameter of interest was taken as the point estimator. When fitting the Bayesian models, the censoring limit c was fixed to 100 ms.

Evaluation using a simulated data set

The point with a simulation study is to simulate data using a distribution in which the parameters are known, since comparisons can then be made as to how well different estimators retrieve the true parameters using the simulated data. This makes it possible to compare different estimators and to gauge the magnitude of estimation error when the models are applied to real data.

Simulation of timing asynchronies In order to simulate timing asynchronies, a number of assumptions have to be made. As was argued earlier, the timing responses will be assumed to come from two sources: Either a response is predictive, resulting from a prediction of the timing of the target stimulus, or a response is reactive, resulting from a reaction to the target stimulus. Furthermore, as is shown in Fig. 2, the distribution of predictive responses is assumed to be a normal distribution with a mean and *SD* that are dependent on the ISI. The reactive responses are assumed to be distributed as an exponentially modified Gaussian (ex-Gaussian) distribution—a right-skewed distribution that has been used to describe the distribution of reaction time responses (Hohle, 1965; Palmer, Horowitz, Torralba, & Wolfe, 2011). The distribution of the reactive responses is assumed to be independent of the ISI. If the predictive and reactive responses are represented by the random variables X_P and X_R, the actual timing responses are distributed as $\min(X_P, X_R)$. This assumes that a timing response is initiated by whichever of the reactive and predictive responses is triggered first, and also that participants tend not to respond twice to target stimuli.

The distribution of the timing responses has five parameters; μ_P and σ_P of the normal distribution for the predictive responses, and μ_R, σ_R, and λ_R of the ex-Gaussian distribution for the reactive responses. In order to simulate the timing responses at different ISIs, these parameters needed to be assigned reasonable values. For μ_P and σ_P, such values were generated by taking the sample mean and *SD* of the asynchronies at different ISIs using the finger-tapping data from the group of musicians in the study by Repp and Doggett (2007) (see Table 1). Although it has been argued in the present article that there are better ways to estimate these parameters than using the sample mean and *SD*, the performance of

Table 1 Values of μ_P and σ_P that were used when simulating the timing asynchronies at different interstimulus intervals (ISIs)

ISI	μ_P	σ_P
1,000 ms	–19 ms	28 ms
1,250 ms	–23 ms	37 ms
1,500 ms	–30 ms	49 ms
1,750 ms	–30 ms	61 ms
2,000 ms	–17 ms	81 ms
2,250 ms	–22 ms	91 ms
2,500 ms	–12 ms	106 ms
2,750 ms	–20 ms	123 ms
3,000 ms	–9 ms	155 ms
3,250 ms	–41 ms	157 ms
3,500 ms	–24 ms	209 ms

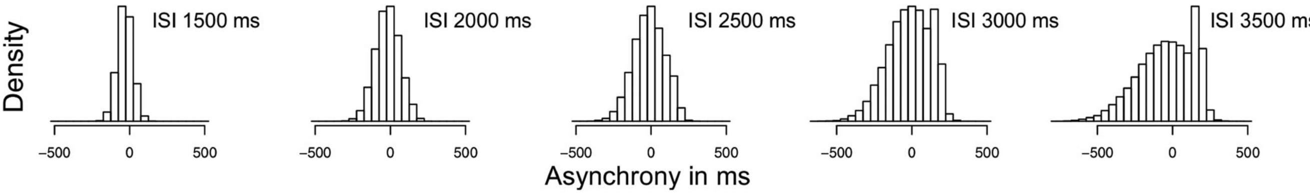


Fig. 6 Distribution of the simulated timing asynchronies for a subset of the ISI levels, generated according to the procedure described in the Simulation of Timing Asynchronies section

the estimators used in this specific case were not of great importance; what mattered was that the values of μ_P and σ_P should be likely values of the constant error and timing variability. Reasonable values for μ_R, σ_R, and λ_R were generated by fitting an ex-Gaussian distribution to the simple auditory reaction time data from a study by Löwgren, Bååth, Lindgren, Sahlén, and Hesslow (2014). The parameter values used for the ex-Gaussian distribution were $\mu_R = 157$ ms, $\sigma_R = 12.5$, and $\lambda_R = 0.031$ ms, which are in agreement with response distributions found in the literature (e.g., Ulrich & Stapf, 1984). Distributions of the simulated data using these parameter values are shown in Fig. 6 (cf. the actual asynchrony distributions from Repp & Doggett, 2007, in Fig. 8 below). For each of the 11 ISI levels, 500 batches of 90 timing responses each were simulated.

Comparison with the moment estimators Because the simulated data were not hierarchical, all data points from the same ISI level shared the same true parameter values, and only the nonhierarchical model and the moment estimators were compared. For each of the in total 500 × 11 = 5,500 batches, the Bayesian model was fit with the JAGS framework (Plummer, 2003), using 1,000 burn-in steps followed by 5,000 MCMC samples, and the resulting fits were used to calculate point estimates for μ_P and σ_P. Similarly, the moment estimators were also used to calculate point estimates for μ_P and σ_P. Figure 7 shows the mean differences between the estimated parameters and the true parameters, with a relative parameter estimate of 0 indicating no difference between the true parameter and the mean of the estimated parameters. Up to an ISI of 2,000 ms, both the Bayesian model and the moment estimators performed similarly, but from an ISI of 2,000 ms the moment estimators increasingly underestimated the true values of μ and σ. At an ISI of 3,500 ms, the mean differences between the Bayesian estimates and the true parameter values were 1.6 ms for both μ and σ, but for the moment estimators the mean differences were 16.6 ms for μ and 27.8 ms for σ.

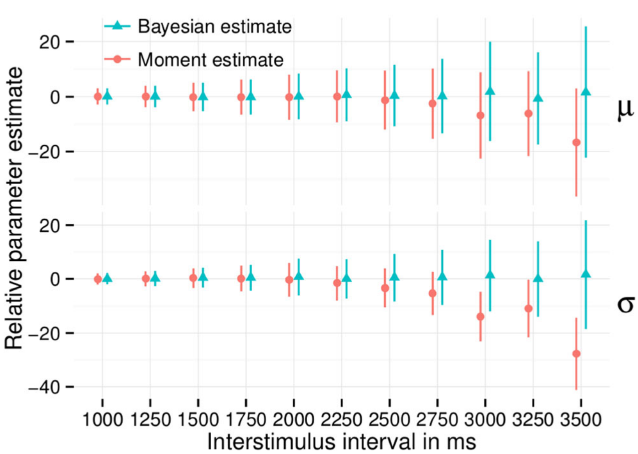


Fig. 7 Mean errors of the moment estimates and the Bayesian estimates compared to the actual parameter values. The error bars show the *SD*s of the estimates

Reanalyzing the data of Repp and Doggett (2007)

In a study by Repp and Doggett (2007), finger-tapping data were collected from eight musicians and 12 nonmusicians synchronizing to isochronous sound sequences, using ISI levels ranging from 1,000 to 3,500 ms.[2] The distributions of timing asynchronies for a subset of those ISI levels are shown in Fig. 8. Notice that the distributions of asynchronies for the long ISI levels exhibit the same pattern shown in Fig. 1; that is, large numbers of responses occur around 200 ms after the stimulus onset.

The data of Repp and Doggett (2007) were reanalyzed by calculating point estimates of the constant error and timing variability for all participants and ISI levels, using both the moment estimators, as in the original article, and the hierarchical Bayesian model. This model was fit separately to the data from the musicians and the data from the nonmusicians with the JAGS framework, using 1,000 burn-in steps followed by 15,000 MCMC samples. The differences between the moment and Bayesian estimates are shown in Figs. 9 and 10. On the basis of the simulation study, the moment estimates should have a tendency to underestimate the constant error and timing variability, and the results of the present analysis support this notion, because the estimates of the Bayesian model are higher than the moment estimates at the long ISI levels. At short ISI levels, this underestimation will be negligible, but at longer ISI levels it will have more of an impact. Whether an analysis would benefit from avoiding underestimation, then, would depend on the ISI range of the study and whether underestimation would impact the conclusions of the study. The Bayesian model estimates and the moment estimates start to diverge when the

[2] Due to a technical error, the timing asynchronies are shifted +15 ms in the original data (Repp, 2008). In the subsequent analysis, this shift has been corrected.

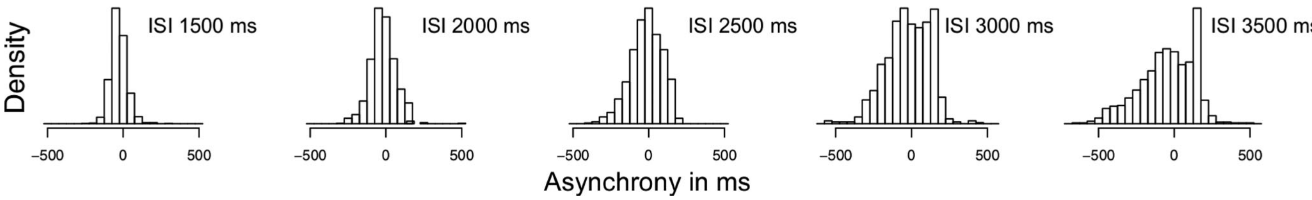


Fig. 8 Distribution of the timing asynchronies from Repp and Doggett (2007) for a subset of the ISI levels

ISI is longer than 2,500 ms, both for the simulated data and for the data from Repp and Doggett. Thus, one might want to consider using a method that avoids underestimation when analyzing a data set that includes ISIs slower than 2,500 ms.

Except for resulting in better parameter estimates, how does the interpretation of Repp and Doggett's (2007) data change when Bayesian estimates are used rather than the moment estimates? One example of what the Bayesian estimates change, relative to the moment estimates, is the interpretation of the source of the reactive responses. Mates et al. (1994) argued that the reason that reactive responses start occurring at long ISIs is a qualitative change in participants' response strategy, due to participants trying to minimize synchronization error. When the ISIs are long enough that predictive responses result in a large expected synchronization error, a better strategy might be to react to the stimulus tone, since this would result in an expected synchronization error smaller than around 200 ms, the average auditory reaction time. Alternatively, Repp and Doggett (2007) argued that reactive responses are not due to a change in response strategy, but rather to participants tending to produce reactive responses when failing to produce a predictive response long enough after the stimulus tone that a reactive response is possible. To evaluate this possible explanation, they used each participant's estimated constant error and timing variability to predict the percentage of reactive responses, under the assumption that the predictive responses would be normally distributed. The percentages of predicted reactive responses were then compared with the actual percentages of reactive responses at the different ISI levels, labeling all responses later than 100 ms as reactive. The predicted percentage of reactive responses was found to be similar to the actual percentage, and Repp and Doggett concluded that "no special strategy of reacting to the tones needs to be assumed" (p. 371). The predicted percentage was, however, slightly lower than the actual percentage, and this difference could still be explained by, for example, a change in response strategy.

A reason for the slightly lower estimates of the percentage of reactive responses might be that the constant error and timing variability were underestimated due to the use of the moment estimators. Using the Bayesian estimates to predict the number of reactive responses yielded a much closer correspondence with the actual percentage of reactive responses, especially at slow ISIs, as is shown in Fig. 11. Consequently, the Bayesian estimates change the interpretation of the data to more strongly support Repp and Doggett's (2007) explanation than their original analysis based on moment estimates.

The power of a hierarchical model

In order to investigate the utility of a hierarchical model, the data from the musicians in Repp and Doggett (2007) were

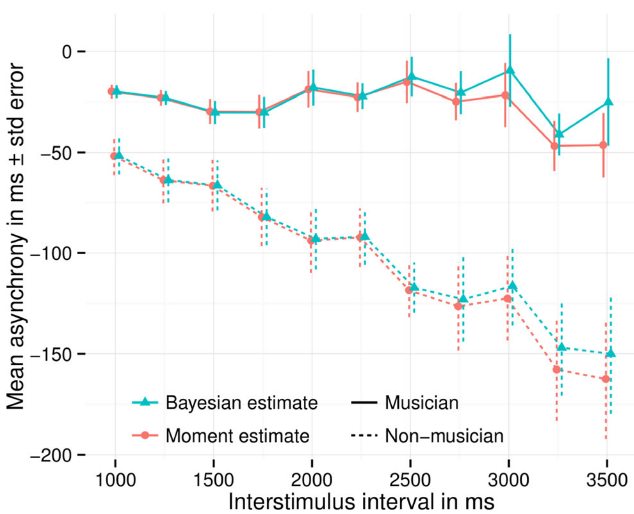


Fig. 9 Grand means of the sample mean estimates and the Bayesian estimates of the mean asynchronies for the musicians and nonmusicians in Repp and Doggett (2007)

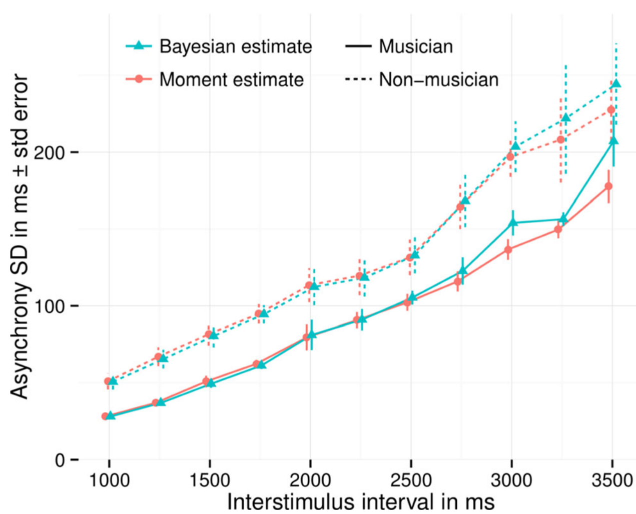


Fig. 10 Grand means of the sample *SD* estimates and the Bayesian estimates of asynchrony *SD*s for the musicians and nonmusicians in Repp and Doggett (2007)

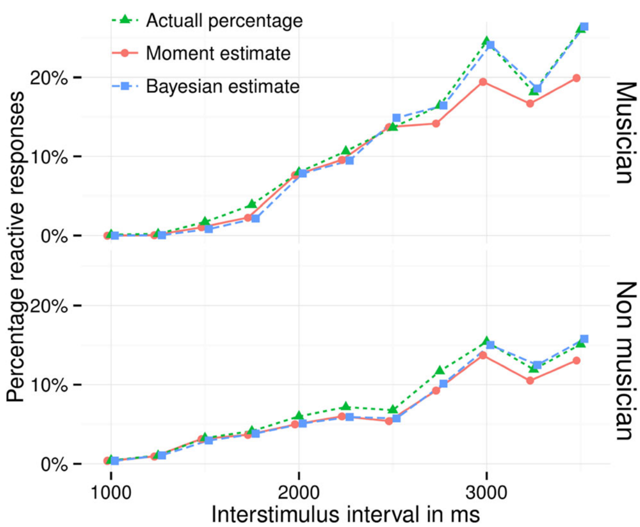


Fig. 11 Actual numbers of reactive responses for the musicians and nonmusicians in Repp and Doggett (2007), compared to the predicted numbers of reactive responses using the moment estimates and the Bayesian estimates

used. Four partitions of the data set were formed, such that each of the four tapping trials at each ISI level for each participant was randomly assigned to a partition. For each partition, the constant error and tapping variability were estimated using the moment estimators, the nonhierarchical model, and the hierarchical model. The Bayesian models were fit with the JAGS framework, using 1,000 burn-in steps followed by 15, 000 MCMC samples. For each estimation method, this yielded 4 (number of partitions) × 8 (number of participants) × 11 (number of ISI levels) = 352 estimates of constant error and tapping variability. In the simulation study, it was possible to compare the estimated parameter values with the actual parameter values, but here there were no "true" parameter values. An alternative would be to compare the parameter estimates from the partitions, which are estimated using only one fourth of the available data, with estimates that used the whole data set. The estimates calculated using the whole data set are assumed to be closer to the true parameters, and by comparing these with the estimates calculated using the partitioned data, it was possible to evaluate how well the three estimation methods retrieve the whole-data estimates. The whole-data estimates were estimated using the nonhierarchical Bayesian model.

Figure 12 shows the mean absolute errors of the three estimation methods applied to the partitioned data, as compared to the estimates based on the whole data. The mean errors of the hierarchical estimates are consistently lower than the errors of the two other estimation methods. Averaged over the ISI levels, the medians[3] of the absolute errors of the hierarchical Bayesian estimates were 12% less than for both the nonhierarchical Bayesian and the moment estimates. Because a hierarchical model benefits from there being many participants in the data set, this better performance of the hierarchical model occurred in spite of only eight participants being included in the analysis. For a data set with even fewer participants, or for participants that perform very differently from each other, a hierarchical model is not likely to improve the estimates much over a nonhierarchical model. However, for the common case in which many participants have comparable performance, using a hierarchical model will likely result in better estimates.

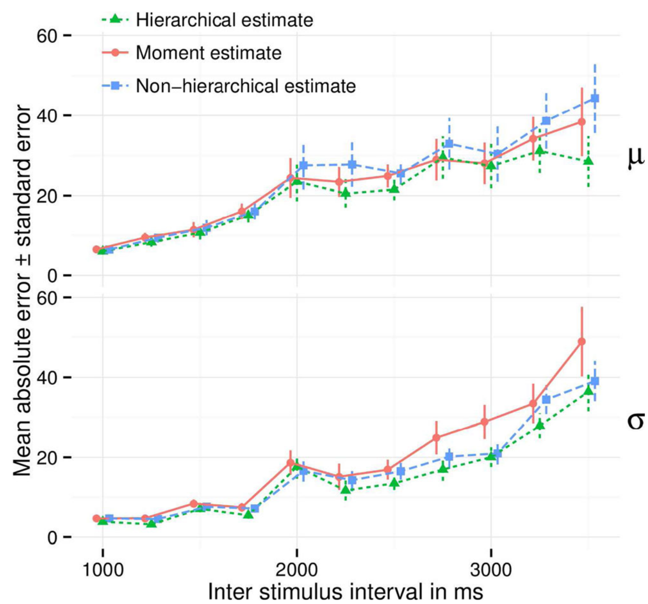


Fig. 12 Mean absolute errors of the three estimators applied to the partitioned data from Repp and Doggett (2007), as compared to the estimates using the full data set. The errors of the hierarchical Bayesian estimates are consistently lower than the errors of the moment estimates and of the nonhierarchical Bayesian estimates

Discussion

In studies dealing with sensorimotor synchronization (SMS) and rhythm production, two of the main parameters of interest are constant error and timing variability. It is common to estimate these parameters by calculating the sample mean and *SD*, but using these moment estimators is problematic in two respects. First, the moment estimators tend to underestimate the constant error and timing variability at long ISIs. This is due to a tendency of participants to overshoot the target interval and instead to react to the target stimulus, resulting in a left-skewed and bimodal distribution for both the stimulus-to-response asynchronies and interresponse intervals. Second, when a data set includes many participants, moment estimators fail to model the hierarchical structure of the data, and as a result all available data are not used when estimating the parameters.

[3] The median was used here as the measure of central tendency because the distributions of relative errors were heavily right skewed.

The Bayesian model presented in this article addressed the first problem by treating the predictive timing responses as a partially latent variable and by modeling the timing asynchronies using a right-censored normal distribution. The second problem was addressed by modeling the hierarchical structure of the data. The model was compared with the moment estimators and was shown to be less biased toward low estimates of constant error and timing variability and to yield more accurate estimates when applied to a hierarchical data set with multiple participants. It was also shown that the Bayesian estimates changed the interpretation of the data from a study by Repp and Doggett (2007).

The focus of the evaluation of the Bayesian model was on contrasting it with the moment estimators. This choice was made because moment estimators are arguably the most common estimators of constant error and timing variability in the literature. The case made against using these estimators when dealing with SMS data could also be made against other methods that do not consider the skewness and hierarchical structure of such data. Some methods other than Bayesian modeling could also address these two issues. A right-truncated normal distribution could, for example, be fit to the data from an SMS task using the approach described by Ulrich and Miller (1994), or the hierarchical structure could be modeled using mixed-effects modeling (Baayen et al., 2008). The combination of these two approaches is, however, more straightforward in a Bayesian framework.

All comparisons between the Bayesian model and the moment estimators were made using point estimates generated by the Bayesian model. This was done to facilitate the comparison with the moment estimators. Although it is certainly possible to use the Bayesian model in this way, it disregards the much more useful approach of using the posterior probability distributions of the parameters for inference. In many cases, the latter approach would make more sense. Why go through the trouble of estimating point values and analyzing them when it is possible to directly analyze the distributions of the hyperparameters already specified in the model? Using the hierarchical version of the model, it is also possible to make inferences regarding the population timing variability by using the posterior probability distribution for the m_σ parameter, and regarding the population constant error by using the posterior probability distribution of the μ_μ parameter. In order to compare two groups of participants, the data of each group could be fit separately using the hierarchical Bayesian model, and then the credible differences between the group parameters could be investigated. One of the main advantages of doing a full Bayesian analysis is that all model parameters are estimated, including measures of uncertainty, so that comparisons and inferences can readily be made regarding any parameter or generated quantities.

Because the model is implemented in the flexible modeling language JAGS, it will be straightforward to extend it. A possible extension would be to include additional predictor variables in the model, allowing the timing variability and constant error to vary not only by ISI, but also by, for example, participant group or task condition. Another extension would be to introduce a functional dependency on the timing variability or constant error between ISI levels. For example, one could assume that the scalar property (Gibbon, Church, & Meck, 1984) holds for timing variability by adding the assumption that the asynchrony *SD* increases linearly as a function ISI. The supplementary text describes how one could add such an assumption to the hierarchical model. That text also describes how one could model a correlation in a participant's timing performance between ISI levels. The purpose of the model presented here was to model the distributional properties of SMS data. In doing that, it did not consider the time series properties of the data, such as the serial dependency of the responses. A further extension of the model would be to combine it with a time series model of SMS, such as the one developed by Vorberg and Wing (1996). Because that model does not separate predictive responses from reactive responses, it should, like the moment estimators, be biased toward low estimates of timing variability when the timing responses include reactive responses.

Author Note This research was supported by the Linnaeus environment Thinking in Time: Cognition, Communication and Learning, Swedish Research Council Grant Number 349-2007-8695. I thank Bruno H. Repp and Rebbecca Doggett for providing their data for reanalysis. In addition, I thank Geoffrey R. Patching and Christian Balkenius for useful comments on various aspects of this work.

References

Andrews, M., & Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology, 66,* 1–7. doi:10.1111/bmsp.12004

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition, 48,* 66–79. doi:10.1006/brcg.2001.1304

Bååth, R., & Madison, G. (2012). The subjective difficulty of tapping to a slow beat. In E. Cambouropoulos, C. Tsourgas, P. Mavromatis, & C. Pastiadis (Eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition* (pp. 82–85). Jyväskylä, Finland: European Society for the Cognitive Sciences of Music. **Retrieved from icmpc-escom2012.web.auth.gr/?q=node/67**.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Chen, Y., Ding, M., & Kelso, J. (1997). Long memory processes ($1/f^\alpha$ type) in human coordination. *Physical Review Letters, 79,* 4501–4504. doi:10.1103/PhysRevLett.79.4501

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex, 45,* 119–130. doi:10.1016/j.cortex.2007.09.008

Craigmile, P. F., Peruggia, M., & van Zandt, T. (2010). Hierarchical Bayes models for response time data. *Psychometrika, 75,* 613–632. doi:10.1007/s11336-010-9172-6

Development Core Team, R. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review, 15,* 1209–1217. doi:10.3758/PBR.15.6.1209

Forsman, L. J., Madison, G., & Ullén, F. (2009). Neuroticism is correlated with drift in serial time interval production. *Personality and Individual Differences, 47,* 229–232. doi:10.1016/j.paid.2009.02.020

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis, 1,* 515–534. doi:10.1214/06-BA117A

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

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

Gottsdanker, R. (1982). Age and simple reaction time. *Journal of Gerontology, 37,* 342–348. doi:10.1093/geronj/37.3.342

Green, J. T., Ivry, R. B., & Woodruff-Pak, D. S. (1999). Timing in eyeblink classical conditioning and timed-interval tapping. *Psychological Science, 10,* 19–23. doi:10.1111/1467-9280.00100

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

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. New York, NY: Springer.

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology, 69,* 382–386. doi:10.1037/h0021740

Holm, L., Ullén, F., & Madison, G. (2011). Intelligence and temporal accuracy of behaviour: Unique and shared associations with reaction time and motor timing. *Experimental Brain Research, 214,* 175–183. doi:10.1007/s00221-011-2817-6

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society A, 186,* 453–461. doi:10.1098/rspa.1946.0056

Kruschke, J. K. (2011a). *Doing Bayesian data analysis*. Amsterdam, The Netherlands: Elsevier.

Kruschke, J. K. (2011b). Introduction to special section on Bayesian data analysis. *Perspectives on Psychological Science, 6,* 272–273. doi:10.1177/1745691611406926

Lee, M. D., & Wagenmakers, E.-J. (2014). Bayesian cognitive modeling: A practical course. Cambridge University Press.

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *BioScience, 51,* 341. doi:10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

Löwgren, K., Bååth, R., Lindgren, M., Sahlén, B., & Hesslow, G. (2014). *The cerebellum, timing and age: Eyeblink conditioning, isochronous interval tapping and prism adaptation in children and adults*. (Manuscript in preparation).

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Madison, G., Forsman, L., Blom, B., Karabanov, A., Ullén, F., & Blom, O. (2009). Correlations between intelligence and components of serial timing variability. *Intelligence, 37,* 68–75. doi:10.1016/j.intell.2008.07.006

Mates, J., Müller, U., Radil, T., & Pöppel, E. (1994). Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience, 6,* 332–340. doi:10.1162/jocn.1994.6.4.332

Miyake, Y., Onishi, Y., & Pöppel, E. (2004). Two types of anticipation in synchronization tapping. *Acta Neurobiologiae Experimentalis, 64,* 415–426.

Moore, G. P., & Chen, J. (2010). Timings and interactions of skilled musicians. *Biological Cybernetics, 103,* 401–414. doi:10.1007/s00422-010-0407-5

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance, 37,* 58–71. doi:10.1037/a0020747

Plummer, M. (2003, March). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*. Article presented at the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria. Retrieved from www.ci.tuwien.ac.at/Conferences/DSC-2003/Drafts/Plummer.pdf

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12,* 969–992. doi:10.3758/BF03206433

Repp, B. H. (2006). Musical synchronization. In E. Altenmüller, M. Wiesendanger, & J. Kesselring (Eds.), *Music, motor control and the brain* (pp. 55–76). Oxford, UK: Oxford University Press.

Repp, B. H. (2008). Metrical subdivision results in subjective slowing of the beat. *Music Perception, 26,* 19–39. doi:10.1525/mp.2008.26.1.19

Repp, B. H., & Doggett, R. (2007). Tapping to a very slow beat: A comparison of musicians and nonmusicians. *Music Perception, 24,* 367–376. doi:10.1525/mp.2007.24.4.367

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

Robert, C. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. Heidelberg, Germany: Springer.

Roeckelein, J. E. (2008). History of conceptions and accounts of time and early time perception research. In S. Grondin (Ed.), *Psychology of time* (pp. 1–50). Bingley, UK: Emerald.

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika, 68,* 589–606. doi:10.1007/BF02295614

Schulze, H. (1992). The error correction model for the tracking of a random metronome: Statistical properties and an empirical test. In M. Françoise, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 275–286). Dordrecht, The Netherlands: Kluwer.

Stevens, L. T. (1886). On the time-sense. *Mind, 11,* 393–404.

Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Pearson.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123,* 34–80. doi:10.1037/0096-3445.123.1.34

Ulrich, R., & Stapf, K. H. (1984). A double-response paradigm to study stimulus intensity effects upon the motor system in simple reaction time experiments. *Perception & Psychophysics, 36,* 545–558. doi:10.3758/BF03207515

van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7,* 424–465. doi:10.3758/BF03214357

Vorberg, D., & Wing, A. (1996). Modeling variability and dependence in timing. In H. Heuer & S. W. Keele (Eds.), *Handbook of perception and action* (Vol. 2, pp. 181–262). San Diego, CA: Academic Press.

Supplementary Material

Estimating the Distribution of Sensorimotor Synchronization Data: A Bayesian Hierarchical Modeling Approach

Rasmus Bååth

One advantage of working within a Bayesian statistical framework is that it is relatively straight forward to modify and extend model definitions. What follows are four possible modifications of the models presented in the main paper. Example implementation of these model using R and JAGS can be found here: `https://github.com/rasmusab/bayes_timing`

1. Extending the Model with Informative Priors

One advantage with a Bayesian approach is that prior knowledge about task performance can be incorporated into the analysis. A subjective Bayesian model is feasible in the case where there exists sufficient prior information regarding the parameters. This is often the case regarding SMS studies where many published papers include descriptive statistics of the distribution of constant error and timing variability at different tempi (for a comprehensive review see Repp and Su, 2013; Repp, 2005). Prior information can be incorporated in the hierarchical model by replacing the vague top-level priors with distributions that can be made more or less informative depending on the strength of the prior information. What follows is a modification of the hierarchical model presented in the paper that enables the inclusion of prior information in the analysis.

The prior on the group mean μ_μ is a normal distribution with parameters $\mu_{\mu,\mu}$ and $\sigma_{\mu,\mu}$. The prior on the mean group SD m_σ is a log-normal distribution with parameters $\mu_{m.\sigma}$ and $\sigma_{m,\sigma}$. To facilitate the use of informative priors this prior is reparameterized to be specified by its arithmetic mean $m_{m,\sigma}$ and SD $s_{m,\sigma}$. As proposed by Gelman, 2006, the SD parameters s_σ and σ_σ are given half-Cauchy priors with parameter $s_{s,\sigma}$ and $s_{\sigma,\mu}$ respectively. It is straightforward to be informative regarding the half-Cauchy priors as the scale parameter defines the median of the distribution (Lunn et al., 2012). The full specification of model is then:

$$
\begin{aligned}
Y_{ij} &\sim \text{Right-Cenc-Normal}\left(\mu_j, \sigma_j, c_j\right) \\
\mu_j &\sim \text{Normal}\left(\mu_\mu, \sigma_\mu\right) \\
\sigma_j &\sim \text{Log-Normal}\left(\mu_\sigma, \sigma_\sigma\right) \\
\mu_\mu &\sim \text{Normal}\left(\mu_{\mu,\mu}, \sigma_{\mu,\mu}\right) \\
\sigma_\mu &\sim \text{Half-Cauchy}\left(s_{\sigma,\mu}\right) \\
\mu_\sigma &= \log\left(m_\sigma\right) - \sigma_\sigma^2/2 \\
\sigma_\sigma &= \sqrt{\log\left(s_\sigma^2/m_\sigma^2 + 1\right)} \\
m_\sigma &\sim \text{Log-Normal}\left(\mu_{m,\sigma}, \sigma_{m,\sigma}\right) \\
s_\sigma &\sim \text{Half-Cauchy}\left(s_{s,\sigma}\right) \\
\mu_{m,\sigma} &= \log\left(m_{m,\sigma}\right) - \sigma_{m,\sigma}^2/2 \\
\sigma_{m,\sigma} &= \sqrt{\log\left(s_{m,\sigma}^2/m_{m,\sigma}^2 + 1\right)}
\end{aligned}
$$

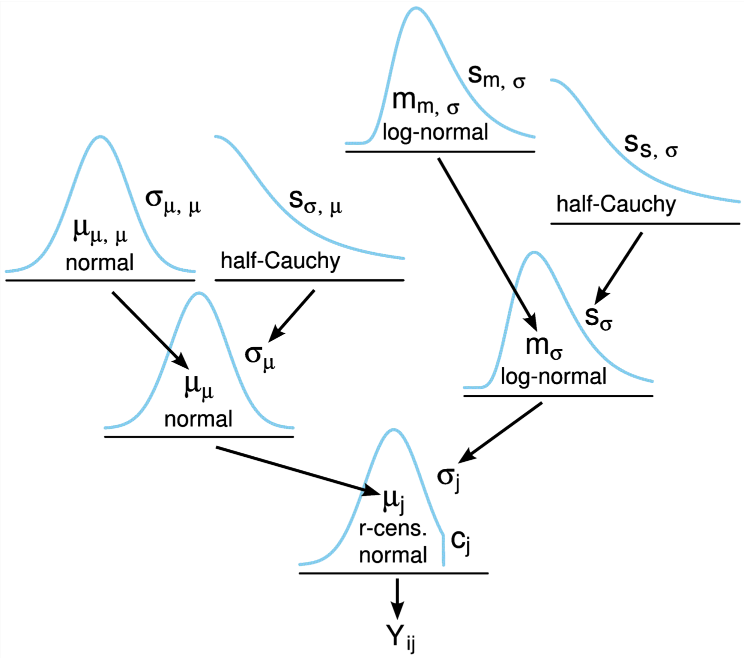


Figure 1: A diagram of the informative hierarchical model.

2. Extending the Model by Adding a Functional Dependency Between ISI levels.

The hierarchical model presented in the paper currently allows that data from one participant informs the parameters of all other participants due to the hierarchical structure of the model. Data from one ISI level does not inform parameters for other ISI levels, however. A dependency between ISI levels can be introduced in many ways, where one possibility is to introduce a functional dependency between ISI levels for the parameters at the group level. Below is a modification of the hierarchical model where there group mean (μ_μ) and the group standard deviation (m_σ) is assumed to depend linearly on the ISI level. Here k indexes the different ISI levels with ISI_k being the ISI in ms at level k.

$$
\begin{gathered}
Y_{i,j,k} \sim \text{Right-Cenc-Normal}\left(\mu_{j,k}, \sigma_{j,k}, c_{j,k}\right) \\
\mu_{j,k} \sim \text{Normal}\left(\mu_{\mu,k}, \sigma_{\mu,k}\right) \\
\sigma_{j,k} \sim \text{Log-Normal}\left(\mu_{\sigma,k}, \sigma_{\sigma,k}\right) \\
\mu_{\mu,k} = \beta_{\mu,0} + \text{ISI}_k \cdot \beta_{\mu,ISI} \\
\sigma_{\mu,k} \sim \text{Uniform}\left(0, \text{ISI}_k/2\right) \\
\mu_{\sigma,k} = \log\left(m_{\sigma,k}\right) - \sigma^2_{\sigma,k}/2 \\
\sigma_{\sigma,k} = \sqrt{\log\left(s^2_{\sigma,k}/m^2_{\sigma,k} + 1\right)} \\
m_{\sigma,k} = \beta_{\sigma,0} + \text{ISI}_k \cdot \beta_{\sigma,ISI} \\
s_{\sigma,k} \sim \text{Uniform}\left(0, \text{ISI}_k/4\right)
\end{gathered}
$$

Where the regression coefficients $\beta_{\mu,0}$, $\beta_{\mu,ISI}$, $\beta_{\sigma,0}$, $\beta_{\sigma,ISI}$ could be given vague priors. Care has to be taken so that $m_{\sigma,k}$ will not take negative values. This can be done by shifting the ISI values so that the shortest ISI level is at the zero and constraining $\beta_{\sigma,0}$, $\beta_{\sigma,ISI}$ to take on only positive values. As an example, data from twelve participants from Bååth and Madison (2012) was used to fit this model. Figure 2 and Figure 3 show the median posterior for the mean asynchrony and asynchrony SD. The colored circles show the group mean ($\mu_{\mu,k}$ in green, $m_{\sigma,k}$ in red), with the colored bars showing one and two SDs ($\sigma_{\mu,k}$ in green, $s_{\sigma,k}$ in red), and the gray circles showing the estimates for each participant. In the case where the assumed functional dependency between ISI levels corresponds well with the data then this modification of the hierarchical model will allow for better informed estimates than if data from each ISI level was estimated on its own.

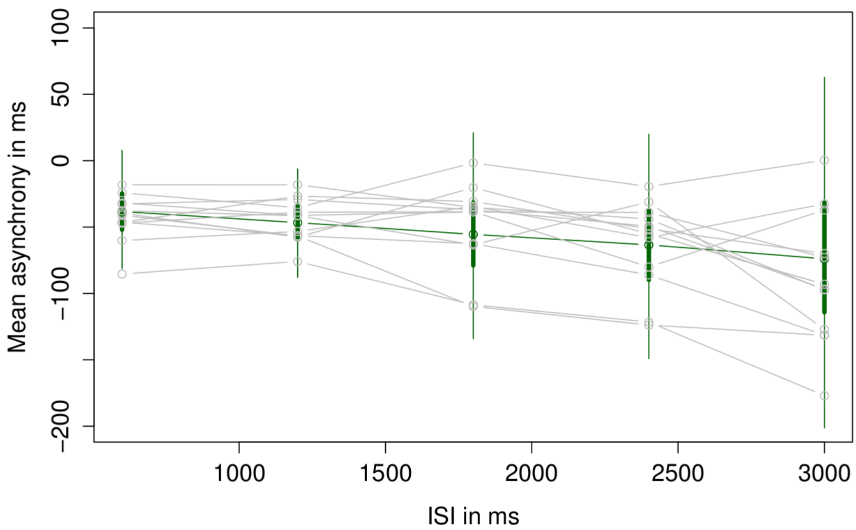


Figure 2: Estimated group mean (green) with individual estimates in grey from the hierarchical model with a functional dependency between ISI levels.

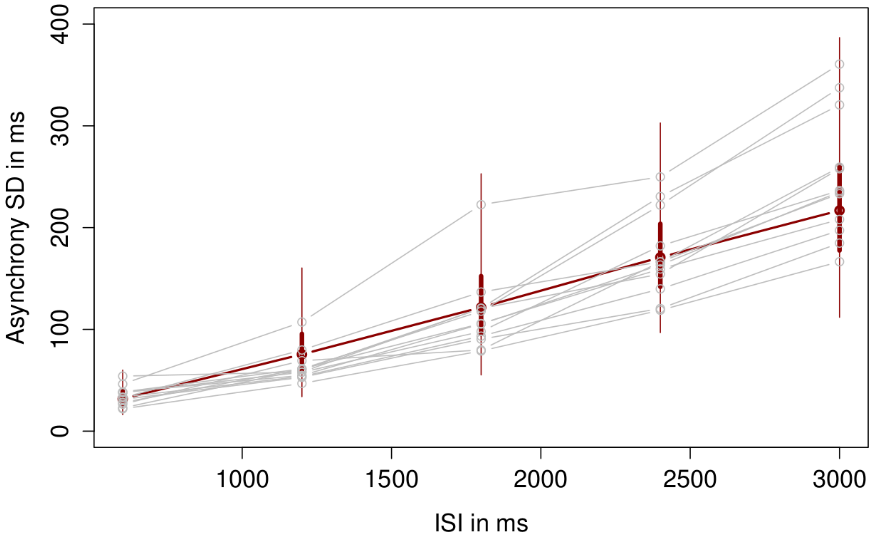


Figure 3: Estimated group SDs (red) with individual estimates in grey from the hierarchical model with a functional dependency between ISI levels.

3. Extending the Model by Modeling the Correlation between Timing Performance at Different ISI levels.

An alternative way of introducing dependencies between ISI levels is to model parameters at the participant level as coming from a multivariate normal distribution. This type of dependency can capture patterns such as that participants with large timing variability at the 600 ms ISI level also tend to have a relatively large variability at the 1200 ms ISI level. This, without imposing a functional dependency between ISI levels. Here there many options, one or more participant level parameters could be given a multivariate normal distribution, and the parameters of the multivariate normal distributions could in turn be assumed to dependent on the ISI level. Below is a modification of the hierarchical model presented in the paper where the logarithms of the participant level asynchrony standard deviations, $\log(\sigma_{j,k})$, are modeled as being distributed as a multivariate normal distribution. Here k again indexes the different ISI levels with ISI_k being the ISI in ms at level *k*, and *n* being the total number of ISI levels. Now $\mu_{\sigma,1\ldots n}$ is a vector of means and Σ_σ is an *n* by *n* covariance matrix. A default non-informative prior to use for Σ_σ could be an Inverse-Wishard distribution with parameters I_n, a *n* by *n* identity matrix, and *n* degrees of freedom.

$$
\begin{gathered}
Y_{i,j,k} \sim \text{Right-Cenc-Normal}\left(\mu_{j,k}, \sigma_{j,k}, c_{j,k}\right) \\
\mu_{j,k} \sim \text{Normal}\left(\mu_{\mu,k}, \sigma_{\mu,k}\right) \\
\log\left(\sigma_{j,k}\right) \sim \text{Multi-Normal}\left(\mu_{\sigma,1..n}, \Sigma_\sigma\right) \\
\mu_{\mu,k} \sim \text{Uniform}\left(-\text{ISI}_k/2, \text{ISI}_k/2\right) \\
\sigma_{\mu,k} \sim \text{Uniform}\left(0, \text{ISI}_k/2\right) \\
\mu_{\sigma,k} \sim \text{Uniform}\left(\log(1), \log\left(\text{ISI}_k/2\right)\right) \\
\Sigma_\sigma \sim \text{Inv-Wishard}\left(I_n, \text{df}{:}\, n\right)
\end{gathered}
$$

Figure 4 shows posterior draws from the $\text{Multi-Normal}\left(\mu_{\sigma,1..n}, \Sigma_\sigma\right)$ distribution that resulted from fitting the model above to finger tapping data from the 30 participants in Bååth and Madison (2012). There is some positive correlation visible, indicating that participants that had a high variability at one ISI level tended to have a relatively high variability at other ISI levels. This correlation also seems strongest between adjacent ISI levels.

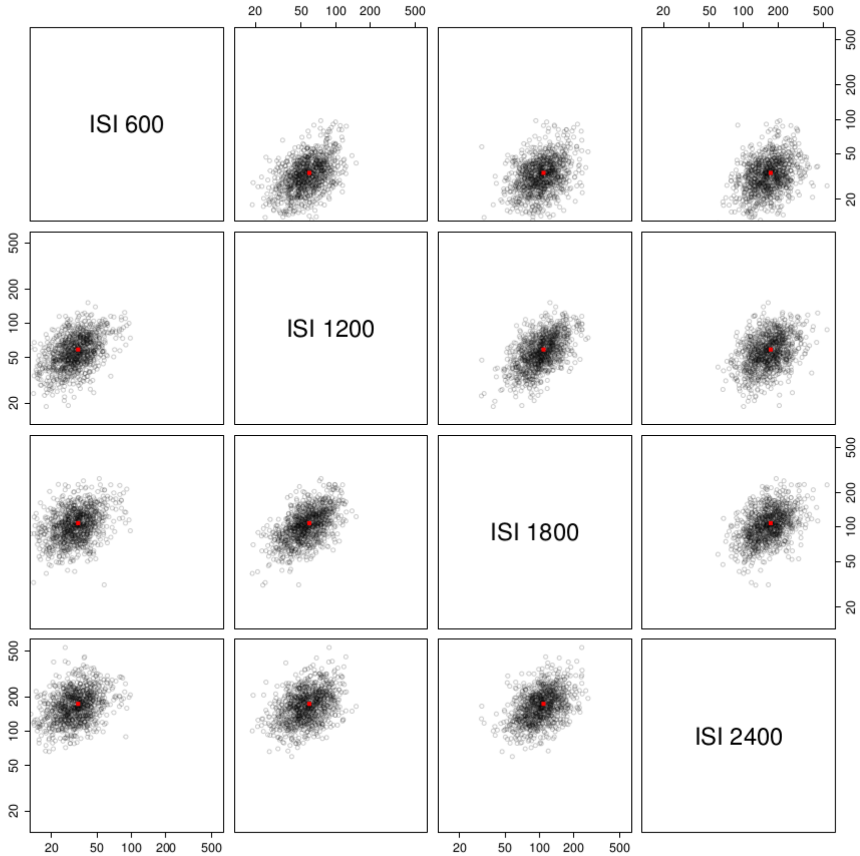


Figure 4: The estimated correlation structure shown as a sample of 1000 draws from the posterior distribution of the Multi-Normal$\left(\mu_{\sigma,1..n}, \Sigma_{\sigma}\right)$ distribution. The red squares marks the marginal means of the posterior.

4. Extending the Model to work with Interresponse Intervals.

The model described in this paper model the stimulus-to-response asynchronies in an SMS task. When the timing responses are self-paced, as in the synchronization-continuation paradigm (Stevens, 1886), there is no referent tone onset and the interresponse intervals, the time difference between consecutive timed responses, are instead the focus of the analysis. The model described in the this paper can be modified to accommodate interresponse interval data by changing the right-censored normal distribution to a normal distribution and by modifying the prior distribution parameters. For the non-hierarchical model a proposal would be to use:

$$I_i \sim \text{Normal}(\mu, \sigma)$$
$$\mu \sim \text{Uniform}(T/k, k \cdot T)$$
$$\log(\sigma) \sim \text{Uniform}\left(\log(1), \log\left(\frac{k}{2} \cdot T\right)\right)$$

Where T is the target interval and k is a constant that is large enough so that the prior distributions include all reasonable values of μ and σ.

References

Bååth, R., & Madison, G. (2012). The subjective difficulty of tapping to a slow beat. *In Proceedings of the 12th international conference on music perception and cognition* (pp. 82–85).

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Chapman and Hall/CRC.

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006-2012). *Psychonomic bulletin & review*, 20(3), 403–52. doi: 10.3758/s13423-012-0371-2

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6), 969–992. doi: 10.3758/BF03206433

Stevens, L. T. (1886). On the time-sense. *Mind*, 11(43), 393 – 404.

Paper VI

Bååth, R. (Submitted) Working memory, memory for musical rhythms, and rhythm perception.

Working Memory, Memory for Musical Rhythms, and Rhythm Perception

Rasmus Bååth

Abstract

This study investigated the relation between auditory working memory, rhythmic timing performance, and memory for musical rhythms. Thirty-six participants were asked to perform each of a digit span task, a finger tapping task and a rhythm memory task. A moderately positive correlation was found between auditory working memory capacity – as measured by the digit span task – and memory capacity for musical rhythms. However, rhythmic timing performance and memory capacity correlated only weakly. Furthermore, the influence of memory capacity on rhythmic timing performance showed no interaction with the interval length of the sequences to which participants synchronized. This suggests that working memory capacity does not play an integral role in rhythm production.

1 Introduction

Time perception can be characterized as the tracking and experience of perceptual events over time. Such a view emphasizes the connection between time perception and memory, as a coherent memory trace of past events appears to be a requirement for perceiving time. It has even been suggested (e.g., by Lewis & Miall, 2006) that time perception is solely dependent on memory traces in working memory.

One aspect of time perception is rhythm perception: the experience of temporal patterns. While both time and rhythm perception are supramodal processes – auditory, visual, and tactile stimuli can all result in time percepts (Hanson, Heron, & Whitaker, 2008) – auditory stimuli tend to dominate over other type of stimuli in the temporal domain (Ortega, Guzman-Martinez, Grabowecky, & Suzuki, 2014). This proves true for rhythm perception, where auditory stimuli take precedence over visual stimuli, and for sensorimotor synchronization where participants are better at synchronizing rhythmic responses to auditory stimuli (Barakat, Seitz, & Shams, 2015; Glenberg, Mann, Altman, Forman, & Procise, 1989; Repp & Penel, 2002). Here, *sensorimotor synchronization* refers to rhythmic coordination of perception and action, where the prototypical sensorimotor synchronization task involves synchronizing finger taps to the beat of a metronome sequence (Repp, 2005).

Given the seeming reliance of time perception on a memory component, and the dominance of auditory stimuli in rhythm perception, *auditory working memory* is a potentially relevant component of rhythm perception. One of the most influential models of working memory is found in A. D. Baddeley and Hitch (1974). It describes working memory as a multi-component system; the component that implements auditory working memory is the called the *phonological loop*. Auditory working memory capacity is commonly measured using a *digit span* task (A. Baddeley, 2000; Hester, Kinsella, & Ong, 2004), in which participants listen to sequences of digits and try to correctly recall as long sequences as possible. Musicians have been shown to have a larger working memory capacity than non-musicians (George & Coch, 2011) and there is a positive correlation between working memory capacity and musical ability (Hansen, Wallentin, & Vuust, 2012). Saito (2001) also found a moderate positive correlation between auditory working memory capacity and performance in a combined rhythm memory and rhythm reproduction task[1] .

Auditory working memory and rhythm perception both have temporal limitations. In Baddeley and Hitch's model of working memory, the phonological loop is assumed to hold acoustic information for up to two seconds (A. Baddeley, 2000). Pöppel (2004) has argued for a two to three second window for temporal integration: events that fall within this window can be united into one percept without effort. Rhythm perception requires the integration of sound events over time and temporal limits of rhythm perception have been modeled as temporal limits in short term memory (Gilden & Marusich, 2009; Grondin, Laflamme, & Mioni, 2015). The proposed limits of rhythm perception range from a little over one second to a couple of seconds. Grondin (2012) notes that when the interstimulus interval (ISI) between consecutive metronome sounds becomes longer than around 1.3 seconds, participants become significantly worse at representing and reproducing the rhythm. The

[1]Note that this task is presented as a rhythm memory task. However, the rhythms that the participants were required to remember were of low complexity. The task score was based on how well participants could reproduce the rhythms by tapping on a computer keyboard. In my opinion, this makes the task more akin to a sensorimotor synchronization task than a rhythm memory task.

experienced difficulty of synchronizing to a metronome also increases significantly as the ISI of the pacing sequence approaches two seconds (Bååth & Madison, 2012). The temporal limit of *subjective rhythmization* – the subjective grouping of monotone metronome sequences – is in the vicinity of a two-second ISI (Bolton, 1894; Bååth, 2015b). Participants' synchronization to a metronome rhythm becomes more variable as the tempo gets slower, but there is no evidence for a "slower limit" of sensorimotor synchronization beyond which point participants are unable to synchronize (Repp, 2006). However, when asked to tap a self-paced regular rhythm as slowly as possible, participants tend to tap at an interval of around 2.5 seconds (McAuley, Jones, Holub, Johnston, & Miller, 2006). This is sometimes called the *slow motor tempo task* and has been employed both as a measure of the a slower limit of rhythm perception (McAuley et al., 2006) and as a measure of auditory working memory capacity (Drake & El Heni, 2003).

The current study investigates the relation between auditory working memory, sensorimotor synchronization performance, and memory for rhythms. The capacity to memorize rhythms is measured using a novel rhythm span task taken from (Schaal, Banissy, & Lange, 2014). The literature on rhythm perception and memory capacity allows a number of predictions to be made. A first prediction – also made by (Schaal et al., 2014) – is that auditory working memory capacity and the capacity to memorize rhythms are positively correlated, possibly as the result of a common underlying mechanism. Second, both a larger auditory and a larger rhythm memory capacity should be related to better performance in a sensorimotor synchronization task. Third, the relation between memory capacity and synchronization performance should be stronger when the rhythms to be synchronized to are relatively slower. As synchronization to slow rhythms requires longer time intervals to be retained and reproduced, a longer memory span should be more beneficial when synchronizing to a slower rhythm. Similarly, there might be a positive correlation between slow motor tempo – as a measure of a slower limit of rhythm production – and memory span.

2 Method

2.1 Participants

Thirty-six participants were recruited through public advertizing (17 women, mean age: 29 years, age SD: 13 years). Twenty-three reported having experience playing a musical instrument where the mean number of years of regular practice was 14 (SD = 13).

2.2 Material

A session consisted of four subtasks: a rhythm span task, a digit span task, a sensorimotor synchronization task, and a slow motor tempo task with overt counting. Participants were tested individually in a quiet room. The rhythm span task, the digit span task and the sensorimotor synchronization task began the session, where the order of presentation of these tasks was randomized. The session concluded with the slow motor tempo task. A session lasted on average 45 minutes. All tasks were presented on a computer running the Ubuntu operating system. Audio was presented through a pair of closed headphones in the sensorimotor synchronization task, while in in the digit and rhythm span tasks audio was presented though a pair of multimedia speakers.

2.2.1 Rhythm span task

Participants were asked to listen to two short rhythm sequences of equal length and judge whether they were identical or different. Depending on participant's performance sequences got longer or shorter. A participant who was able to remember and correctly compare longer sequences received a higher *rhythm span* score. The task was modeled after that described by Schaal et al. (2014), which in turn is modeled after the pitch span task described by Williamson and Stewart (2010). The task used identical auditory stimuli and the only difference with respect to Schaal et al. was slight changes in the visual presentation.

The sequences ranged from two to ten beats where one to three notes were played each beat. They were played at 60 beats per minute using 70 ms long triangle wave sounds with a frequency of 440 Hz. Figure 1 shows an example of a pair of four beat sequences where the rhythm differs.

In each trial, the participant was presented with a pair of sequences, in turn, where it was randomized whether the rhythms were to be identical or different. They were separated by a two second pause. The participant was then asked to indicate whether the sequences were identical or different by pressing the right or the left Control key on a computer keyboard. The length of sequences followed a two-up, one-down staircase procedure: The length of the sequences increased after two correct responses and decreased after one incorrect response. The task ended after the sequence length had reversed direction eight times. The final rhythm span score was calculated by taking the mean

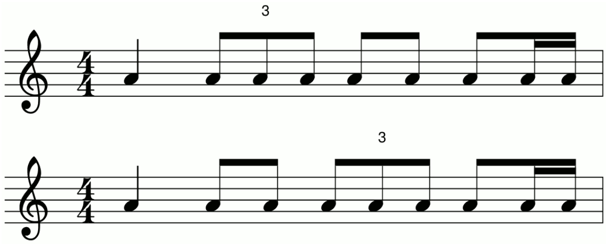


Figure 1: An example of a pair of rhythm sequences from the rhythm span task. In this example the correct reponse would be to indicate that the sequences are different.

of the sequence length on the last six trials where the sequence length reversed direction.

2.2.2 Digit span task

This task was identical to the forward digit span task described in the Wechsler Adult Intelligence Scale IV (Wechsler, 2008), with the one difference: the sequences of digits were pre-recorded and played back to the participants using a computer instead of being read by the experimenter. Participants were asked to listen to sequences of digits and repeat them back, in the same order. The first sequence was two digits long and two sequences were presented at each sequence length level. The task terminated when the participant could not remember either of the sequences at the current length level correctly or after two sequences of length nine had been presented. Sequences consisted of the digits “one” to “nine”, with no digit being repeated, and were read at a rate of one digit per second. The final digit span score was calculated as the number of correctly recalled sequences; the maximum attainable score was sixteen.

2.2.3 Sensorimotor synchronization task

In this task participants were asked to synchronize finger taps to isochronous metronome sequences. They were instructed to start as soon as a sequence began and to continue to tap until it ended. A custom-built tapping board, consisting of a piezoelectric sensor mounted on a 5 cm^2 piece of corrugated fiberboard, recorded the finger taps (see Bååth, 2011 for details). Participants tapped with their index finger, their hand resting on a foam cushion. The stimuli consisted of isochronous sequences of 440 Hz square wave tones of 20 ms, where each sequence was 31 tones long. They were presented at four tempi, corresponding to tone ISIs of 500, 1000, 2000 and 3000 ms. An Arduino microcontroller was used both for generating the sounds and registring the taps.The task was divided into three blocks of four trials each, one for each ISI level. The order of the trials within each block was randomized. Participants were instructed to tap along to each tone sequence, to start tapping as soon as the sequence began, and to stop tapping when the sequence stopped. Participants were requested not to subdivide the beat in any way, for example, by covert counting or by moving their body.

2.2.4 Slow motor tempo task with overt counting

The same apparatus was used as in the sensorimotor synchronization task with the difference that the finger tapping was self paced. Prior to each trial, participants were instructed to tap a regular beat that was as slow as possible, while still maintaining a regular beat. Participants were asked to refrain from subdividing the taps in any way. To avoid covert subdivision, the participants were instructed to count aloud with each tap, starting from one. These instructions conform to those described by (McAuley et al., 2006), with the addition of the overt counting. The task consisted of three trials of 15 taps each.

2.3 Analysis

For the sensorimotor synchronization task, the first four taps in every trial were discarded, so as to use only those taps where the participants had had some time to synchronize to the sequence. For each tap, the tone-to-tap asynchrony was calculated as the time difference between the tone and the tap, where a negative asynchrony indicates that the tap preceded the tone. The asynchrony SD was used as a measure of timing variability and was estimated for each participant and ISI level using the Bayesian hierarchical method described in (Bååth, 2015a). This method was used, instead of the conventional sample SD, as the Bayesian method has been shown to yield more accurate estimates of timing variability when participants synchronize to slow tempo sequences. Timing variability is here used as the measure of performance in the sensorimotor synchronization task where a low timing variability is taken to mean high synchronization performance.

A technical fault with the tapping board meant that data from four participants in the sensorimotor-synchronization task was lost, as well as that from one participant in the slow motor tempo task. Data from both tapping tasks was excluded for one participant who, after the experiment, admitted to having subdivided the beat covertly. Data was excluded from two participants in the sensorimotor synchronization task as analysis suggested that, in many of the trials, they tapped on the off-beat: the time points in between the tones. However, the exclusion or inclusion of this data does not change the result of the experiment as neither estimates, confidence intervals, nor p-values differ substantially depending on whether this data is excluded or retained.

The slow motor tempo for each participant was estimated by first calculating the median intertap interval for each trial, then taking the mean of the three trial medians. The median was used rather than the mean because some participants were found to produce a small number of inter-tap intervals deviating greatly from the norm.

Statistical analysis was performed using the statistical computing environment *R* (R Core Team, 2012). Relationships between the main measures were assesssed using Pearson product-moment correlation, except in the case of timing variability. As timing variability was measured at four different ISIs for each participant, a linear mixed-effects model was used to assess how timing variability changed as a function of the other measures. Mixed-effects model analyzes were performed using the package *lme4* (Bates, Mächler, Bolker, & Walker, 2014).

3 Results

The following measures were calculated for each participant: A digit span score, a rhythm span score, a slow motor tempo and a timing variability at the four ISI levels. Slow motor tempo was calculated as the average intertap interval and timing variability as the asynchrony SD.

Summary statistics for these measures are presented in Table 1. The distributions of these measures were found to be positively skewed and were therefore log-transformed in the subsequent statistical analysis. Both the median rhythm span and median digit span scores conformed to median scores reported in other studies on similar populations (Salthouse & Saklofske, 2010; Schaal et al., 2014).

A positive correlation was found between the rhythm span and digit span scores ($r(34) = 0.48$, 95% CI: [0.18, 0.70], $p = 0.003$). Figure 2 shows the relation between these measures, including marginal distributions.

The relationship between timing variability and rhythm span score was investigated using a linear mixed-effects model with log asynchrony SD as the outcome variable and ISI, log rhythm span score, and the interaction between ISI and rhythm span score as the predictor

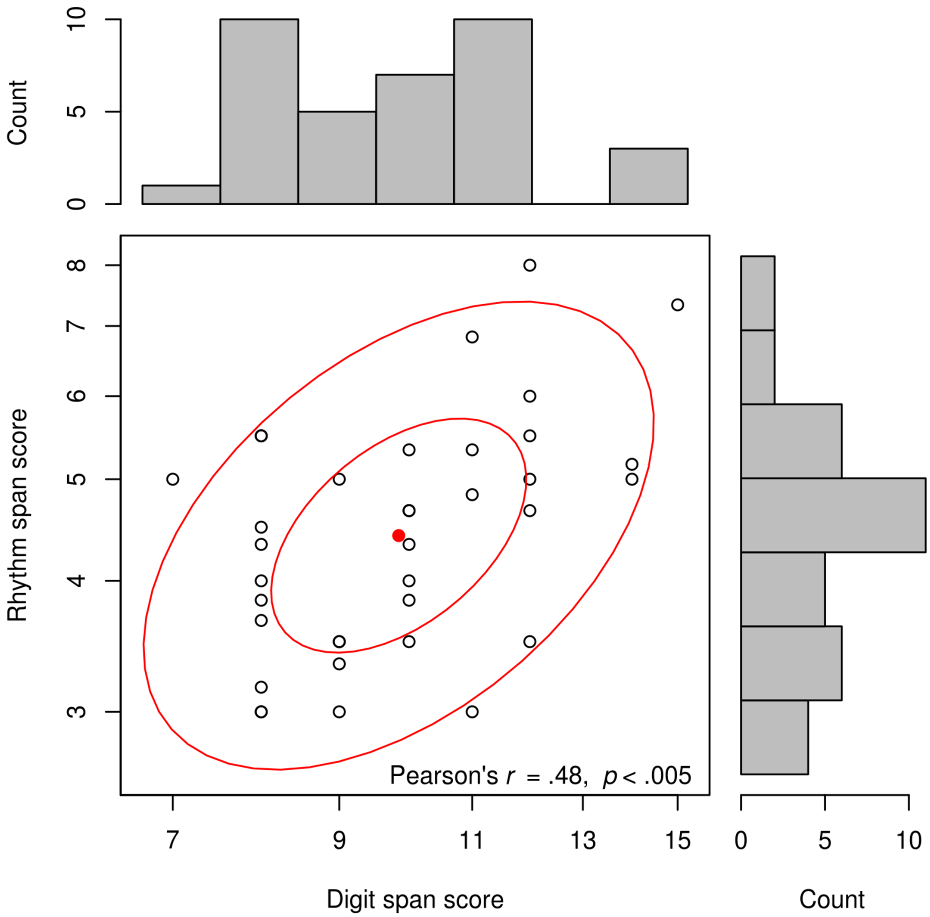


Figure 2: The relation between digit span score and rhythm span score. The ellipses reflect the correlation one and two SD out from the mean, here shown as the filled circle.

Measure	Median	25% quantile	75% quantile	SD
Rhythm span score	4.6	3.5	5.2	1.2
Digit span score	10	8	11	2.0
Slow motor tempo	2719 ms	2124 ms	3207 ms	1046ms
Asynchrony SD				
ISI 500 ms	25 ms	23 ms	30 ms	6.1 ms
ISI 1000 ms	39 ms	34 ms	44 ms	19 ms
ISI 2000 ms	112 ms	87 ms	137 ms	32 ms
ISI 3000 ms	216 ms	204 ms	236 ms	44 ms

Table 1: Summary statistics for the main measures

variables. The rhythm span score and ISI variables were standardized prior to the regression analysis. As each participant contributes four data points – one for each ISI level – the intercept and ISI effect were treated as random effects by participant. Table 2 shows the estimated regression coefficients. There was a statistically significant effect of ISI and rhythm span score, where participants with a large rhythm span score tended to have lower timing variability. However, no substantial interaction effect was found between ISI and rhythm span score.

Coefficient	Estimate	95% CI	p
Intercept	4.26	[4.20, 4.33]	-
ISI	0.83	[0.80, 0.87]	$< .001$
log(rhythm span score)	-0.088	[-0.15, -0.024]	.011
ISI × log(rhythm span score)	0.0050	[-0.029, 0.039]	.77

Table 2: Estimated coefficients for the linear mixed-effects model with rhythm span score as predictor.

Figure 3 shows the effect of ISI and rhythmspan score. The two regression lines were obtained from the coefficients in Table 2 by plugging in the 25% and 75% quantiles of the rhythm span score. The results show a strong effect of ISI, a small constant effect of rhythm span score, but no substantial interaction effect: i.e., the effect of rhythm span score does not change significantly between ISI levels.

The relationship between timing variability and digit span score was investigated again using a linear mixed-effects model, identical to the one described above but with digit span score as the predictor variable. As with the analysis of the rhythm span score, digit span score shows a small but statistically significant effect on timing variability, where participants with a high digit span score tend to have lower timing

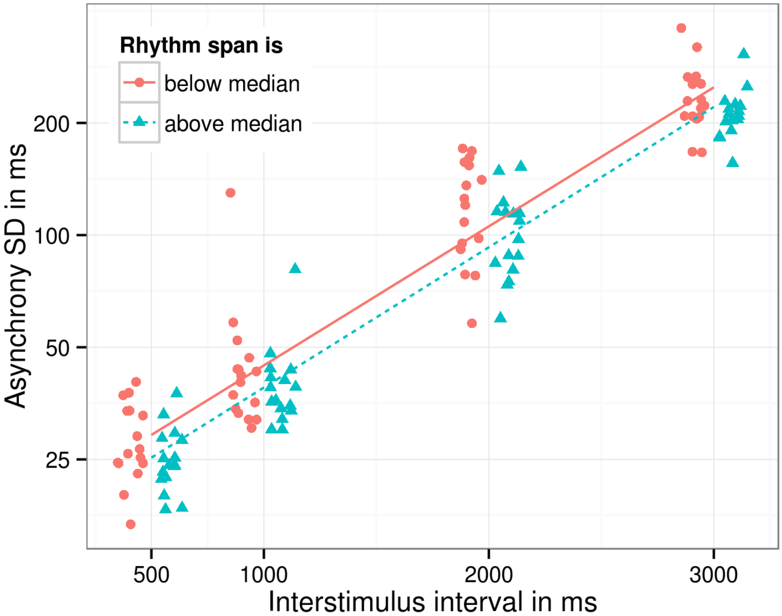


Figure 3: Effect of ISI and rhythm span score on asynchrony SD as estimated using a linear mixed-effects model.

variability. No substantial interaction effect was found between ISI and digit span score. Table 3 shows the estimated regression coefficients. Figure 4 – made made in the same way as Figure 3 – shows the effect of ISI and digit-span score.

Coefficient	Estimate	95% CI	*p*
Intercept	4.26	[4.20, 4.33]	-
ISI	0.83	[0.80, 0.87]	< .001
log(Digit span score)	-0.077	[-0.14, -0.012]	.027
ISI × log(Digit span score)	-0.0026	[-0.037, 0.031]	.88

Table 3: Estimated coefficients for the linear mixed-effects model with digit span score as predictor.

Slow motor tempo showed close to no correlation with digit span score ($r(33) = 0.034$, 95% CI: [-0.30, 0.36], p = .84) and a weak positive correlation with rhythm span score ($r(33) = 0.35$, 95% CI: [-0.018, 0.59], $p = 0.063$). Figure 5 shows the distribution of participants' slow motor tempo. A linear mixed-effects model found no statistically significant

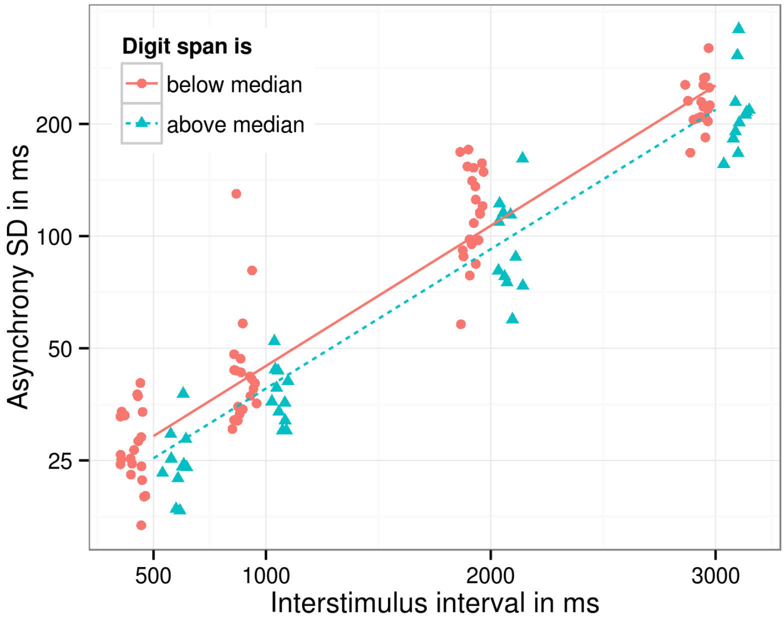


Figure 4: Effect of ISI and digit span score on asynchrony SD as estimated using a linear mixed-effects model.

effect of slow motor tempo on timing variability.

4 Discussion

The current study examined the relationship between auditory working memory, sensorimotor synchronization performance, and memory capacity for rhythms. Auditory working memory was measured using a standard digit span task (Wechsler, 2008), sensorimotor synchronization performance was measured as the timing variability in a finger tapping task, and memory capacity for rhythms was measured using the rhythm span task described in Schaal et al. (2014). Participants were also given a novel slow motor tempo task with overt counting, which aims at measuring a slower limit of rhythm perception. A number of predictions were made regarding the relationship between these measures, based on the current literature on rhythm perception and working memory.

A first prediction was that there would be a positive correlation between auditory working memory capacity and capacity to memorize

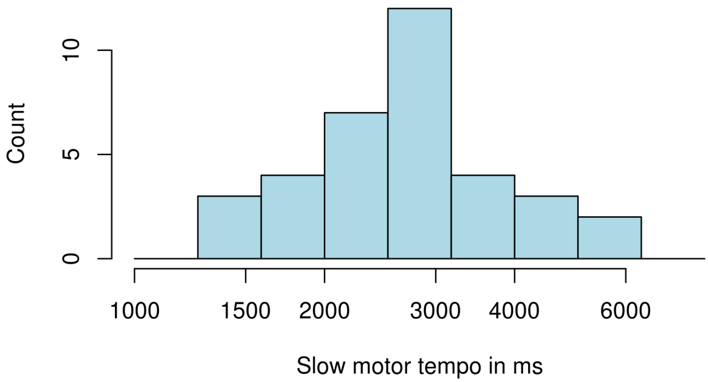


Figure 5: Distribution of participants' slow motor tempo.

rhythms. A correlation of 0.48 was found, which is generally considered a moderate positive correlation (Taylor, 1990). To put this in perspective, compare this result with the difference in median rhythm span score between musicians and non-musicians in the study by Schaal et al. (2014): 4.5 and 3.8 respectively. In this study, the median rhythm span score for the group with a digit span score equal to or above the median was 4.9; that for the group with a digit span score below the median was 3.8. Compared to the difference between musicians and non-musicians (0.67), the difference between the high and low digit span groups is considerably larger (1.1). This suggests that working memory capacity is as strong a predictor of memory capacity for rhythms as being an active musician. That result could be considered surprising, given that musicians spend much of their time practicing rhythms.

A second prediction was that both a larger auditory and a larger rhythm memory capacity should correlate with lower timing variability in a sensorimotor synchronization task. While the effect of rhythm span score on timing variability *was* statistically significant, the effect was not strong. The asynchrony SD and rhythm span score variables were log transformed prior to being entered into the regression analysis, making it difficult to interpret the resulting coefficients (shown in Table 2) directly. However, because the predictors were standardized, it is possible to compare the magnitude of the coefficients. One SD increase in ISI, corresponding here to an increase of 1109 ms, is predicted to increase the log asynchrony SD by 0.83. The effect of rhythm span score is comparably much smaller: a one SD increase in log rhythm span score is predicted to decrease the log asynchrony SD by 0.088. the same decrease in timing variability that would be predicted by decreasing the ISI by 117bms. For example, at an ISI of 1000 ms the predicted asynchrony SD for the group having a below median rhythm span score is 45 ms, for the group having an above median score it is 39 ms. The difference is only 6 ms, which could be a considered a small decrease in timing variability.

The effect of auditory working memory on timing variability was very similar to the effect of rhythm memory capacity, namely, a small but statistically significant effect of digit span score on asynchrony SD, with a larger digit span score predicting a smaller asynchrony SD. Given that sensorimotor synchronization performance is positively correlated with other capacities such as fluid intelligence (Madison, Forsman, Blom, Karabanov, & Ullén, 2009) and simple reaction time (Holm, Ullén, & Madison, 2011), this small effect of memory capacity on sensorimotor synchronization performance weighs against the idea that auditory working memory is an integral part of rhythm perception. Rather, given that working memory is also related to fluid intelligence (Engle, Tuholski, Laughlin, & Conway, 1999), the relation between sensorimotor synchronization and memory capacity can be explained, in part, by that they both correlate with other capabilities.

A third prediction was that the relation between memory capacity and synchronization performance would be stronger when synchronizing to slower sequences. This prediction was based on that synchronization to slow sequences requires longer time intervals to be retained and reproduced, and that a long auditory working memory span would be advantageous for retention of long intervals. The temporal span of working memory has been suggested to be between two and three seconds (A. Baddeley, 2000; Pöppel, 2004), with large individual differences in working memory capacity (Just & Carpenter, 1992); therefore, the effect of working memory span was expected to be especially pronounced for the sequences with an ISI of 2000 and 3000 ms. No such interaction effect was found. As Figure 4 shows, the estimated advantage of having a large working memory span was constant over all ISI levels, this was also the case for rhythm span. Again, the results do not support the view that working memory is integral to sensorimotor synchronization.

The results of the slow motor tempo task showed no substantial correlation with the other measures. One reason might be that the instructions for the slow motor tempo task were open to interpretation, leaving participants free to approach the task in many different ways. A participant might choose to focus on tapping at a very slow tempo, resulting in a more variable response, or on responding consistently, requiring the participant to tap at faster tempo. While slow motor tempo did not show any substantial correlation with working memory capacity, the majority of participants had slow motor tempi in the range of 2000 to 3000 ms, which is also the temporal region of a suggested temporal span of working memory (A. Baddeley, 2000; Pöppel, 2004).

In conclusion, the results suggest that auditory working memory – as measured by a forward digit span task – and memory capacity for rhythms are related. Indeed, a high working memory capacity is as strong a predictor of rhythm memory capacity as extensive musical experience, if not stronger. Auditory working memory and memory capacity for rhythms are also related to sensorimotor synchronization performance, albeit weakly. The influence of memory capacity on synchronization performance shows no interaction with sequence tempo, suggesting that auditory memory capacity does not play an integral role in rhythm production. This is in line with models of rhythm perception, such as that of (Large, 2010), according to which rhythm perception does not depend on an explicit memory component.

5 Acknowledgment

This research was supported by the Linnaeus environment *Thinking in Time: Cognition, Communication and Learning*: Swedish Research Council grant 349-2007-8695. I would like to thank Christian Balkenius and Joel Parthemore for helpful comments on this manuscript.

References

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation - Advances in Research and Theory*, *8*(C), 47–89. doi: 10.1016/S0079-7421(08)60452-1

Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current biology*, *25*(2), R60–1. doi: 10.1016/j.cub.2014.12.011

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv e-prints*. Retrieved from `http://arxiv.org/abs/1406.5823`

Bolton, T. (1894). Rhythm. *The American Journal of Psychology*, *6*(2), 145–238. Retrieved from `http://www.jstor.org/stable/1410948`

Bååth, R. (2011). Construction of a Low Latency Tapping Board. *LUCS minor*, *17*.

Bååth, R. (2015a). Estimating the distribution of sensorimotor synchronization data: A Bayesian hierarchical modeling approach. *Behavior research methods*. doi: 10.3758/s13428-015-0591-2

Bååth, R. (2015b). Subjective Rhythmization: A Replication and an Assessment of Two Theoretical Explanations. *Music Perception*.

Bååth, R., & Madison, G. (2012). The subjective difficulty of tapping to a slow beat. In *Proceedings of the 12th international conference on music perception and cognition* (pp. 82–85). Retrieved from `www.sumsar.net/papers/ICMPC_2012_rasmus_baath.pdf`

Drake, C., & El Heni, J. B. (2003). Synchronizing with Music: Intercultural Differences. *Annals of the New York Academy of Sciences*, *999*, 429–437. doi: 10.1196/annals.1284.053

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, a. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology. General*, *128*(3), 309–331. doi: 10.1037/0096-3445.128.3.309

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094. doi: 10.1016/j.neuropsychologia.2011.02.001

Gilden, D. L., & Marusich, L. R. (2009). Contraction of time in attention-deficit hyperactivity disorder. *Neuropsychology*, *23*(2), 265–269. doi: 10.1037/a0014553

Glenberg, A. M., Mann, S., Altman, L., Forman, T., & Procise, S. (1989). Modality effects in the coding reproduction of rhythms. *Memory & Cognition*, *17*(4), 373–383. doi: 10.3758/BF03202611

Grondin, S. (2012). Violation of the Scalar Property for Time Perception Between 1 and 2 Seconds: Evidence From Interval Discrimination, Reproduction, and Categorization. *Journal of experimental psychology. Human perception and performance*, *38*(4), 880–890. doi: 10.1037/a0027188

Grondin, S., Laflamme, V., & Mioni, G. (2015). Do not count too slowly: evidence for a temporal limitation in short-term memory. *Psychonomic bulletin & review*, *22*(3), 863–8. doi: 10.3758/s13423-014-0740-0

Hansen, M., Wallentin, M., & Vuust, P. (2012). Working memory and musical competence of musicians and non-musicians. *Psychology of Music*. doi: 10.1177/0305735612452186

Hanson, J. V. M., Heron, J., & Whitaker, D. (2008). Recalibration of perceived time across sensory modalities. *Experimental Brain Research*, *185*(2), 347–352. doi: 10.1007/s00221-008-1282-3

Hester, R. L., Kinsella, G. J., & Ong, B. (2004). Effect of age on forward and backward span tasks. *Journal of the International Neuropsychological Society : JINS*, *10*(4), 475–481. doi: 10.1017/S1355617704104037

Holm, L., Ullén, F., & Madison, G. (2011). Intelligence and temporal accuracy of behaviour: unique and shared associations with reaction time and motor timing. *Experimental brain research.*, *214*(2), 175–83. doi: 10.1007/s00221-011-2817-6

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review*, *99*(1), 122–149. doi: 10.1037/0033-295X.99.1.122

Large, E. (2010). Neurodynamics of music. In M. Riess Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception* (pp. 201–231). Springer. doi: 10.1007/978-1-4419-6114-3

Lewis, P. a., & Miall, R. C. (2006, September). Remembering the time: a continuous clock. *Trends in cognitive sciences*, *10*(9), 401–6. doi: 10.1016/j.tics.2006.07.006

Madison, G., Forsman, L., Blom, O., Karabanov, A., & Ullén, F. (2009). Correlations between intelligence and components of serial timing variability. *Intelligence*, *37*(1), 68–75. doi: 10.1016/j.intell.2008.07.006

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *Journal of experimental psychology. General*, *135*(3), 348–67. doi: 10.1037/0096-3445.135.3.348

Ortega, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2014). Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability. *Attention, perception & psychophysics*, *76*(5), 1485–502. doi: 10.3758/s13414-014-0663-x

Pöppel, E. (2004). Lost in time: A historical frame, elementary processing units and the 3-second window. In *Acta neurobiologiae experimentalis* (Vol. 64, pp. 295–301). Retrieved from `http://www.ane.pl/pdf/6428.pdf`

R Core Team. (2012). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from `http://www.r-project.org`

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*(6), 969–992. doi: 10.3758/BF03206433

Repp, B. H. (2006). Rate Limits of Sensorimotor Synchronization. *Advances in Cognitive Psychology*, *2*(2), 163–181. doi: 10.2478/v10053-008-0053-9

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of experimental psychology. Human perception and performance*, *28*(5), 1085–1099. doi: 10.1037/0096-1523.28.5.1085

Saito, S. (2001). The phonological loop and memory for rhythms: An individual differences approach. *Memory (Hove, England)*, *9*(4), 313–322. doi: 10.1080/09658210143000164

Salthouse, T. a., & Saklofske, D. H. (2010). Do the WAIS-IV Tests Measure the Same Aspects of Cognitive Functioning in Adults Under and Over 65? In *Wais-iv clinical use and interpretation* (pp. 217–235). doi: 10.1016/B978-0-12-375035-8.10008-4

Schaal, N. K., Banissy, M. J., & Lange, K. (2014). The Rhythm Span Task: Comparing Memory Capacity for Musical Rhythms in Musicians and Non-Musicians. *Journal of New Music Research*, *44*(April 2015), 3–10. doi: 10.1080/09298215.2014.937724

Taylor, R. (1990). Interpretation of the Correlation Coefficient: A Basic Review. *Journal of Diagnostic Medical Sonography*, *6*(1), 35–39. doi: 10.1177/875647939000600106

Wechsler, D. (2008). *Wechsler adult intelligence scale - Fourth Edition (WAIS - IV)*. San Antonio, TX: NCS Pearson.

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem. *Memory (Hove, England)*, *18*(6), 657–69. doi: 10.1080/09658211.2010.501339

Paper VII

Bååth, R., Tjøstheim, T., Lingonblad M. (Submitted) The role of executive control in rhythmic timing at different tempi.

The Role of Executive Control in Rhythmic Timing at Different Tempi

Rasmus Bååth, Trond Arild Tjøstheim, Martin Lingonblad

Abstract

We investigated the role of attention and executive control in rhythmic timing, using a dual-task paradigm. The main task was a finger tapping task in which participants were asked to tap their index finger in time with metronome sequences. The tempo of the sequences ranged from 600 ms to 3000 ms between each beat. The distractor task, chosen so as to engage executive control processes, was a novel covert n-back task. When the tempo was slow, simultaneous performance of the tapping and n-back tasks resulted in significant performance degradation in both tasks. There was also some dual-task interference at the fast tempo levels, however, the magnitude of the interference was much smaller in comparison. The results suggests that, when the tempo is sufficiently slow, performing rhythmic timing demands attentional resources and executive control. This accords with models of time perception that assume that different timing mechanisms are recruited at different time scales. It also accords with models that assume a dedicated mechanism for rhythm perception and where rhythm perception is assumed to have a slower limit.

Like visual perception can be divided into such subcategories as color perception, motion perception, and depth perception, so too can time perception. Some aspects of time perception are interval timing, temporal motor coordination, rhythm perception, and meter perception. It is possible to further subdivide time perception by modality and time scale. Much debated is whether all aspects of time perception share a common mechanism and, if not, what aspects of which mechanisms they do share. One influential class of models assumes that timing is governed by a pacemaker-accumulator type mechanism (Ivry, 1996), while more recent theoretical development are dynamical systems models that assume that timing and rhythm perception depend on oscillatory neural circuits (Large, 2010; Large & Jones, 1999). The former have been used successfully to model interval timing but has not proven a good model of responses to more complex stimuli such as musical rhythms, while the latter have been used to model rhythm and meter perception but have not been applied to interval timing (Grondin, 2010). The two mechanisms – pacemaker-accumulator type and oscillatory based – need not stand in opposition; models exist that incorporate both (Teki, Grube, & Griffiths, 2011).

Some have suggested that time perception relies on different mechanisms, depending on time scale. P. a. Lewis and Miall (2006) report evidence that different neural mechanism are responsible for timing intervals shorter versus longer than one second. The timing of sub-second intervals has been termed *automatic timing* and that of supra-second intervals has been termed *cognitive timing*. These terms reflect that automatic timing recruits circuits within the motor system and auditory cortex, while cognitive timing depends more on circuits within the prefrontal and parietal cortices (P. A. Lewis & Miall, 2003). Interval timing is but one aspect of time perception, and a three second window has been suggested as the limit of temporal integration (Mates, Müller, Radil, & Pöppel, 1994; Pöppel, 2004). For rhythmic timing, for example, synchronizing finger taps to a metronome sequence, there is evidence supporting a shift in mechanisms between a one- and a two second interstimulus interval (ISI), i.e., the time interval between each beat in a rhythmic sequence, where a short ISI implies a fast tempo and vice versa. The *Weber fraction* – a measure of relative timing error – increases markedly from synchronizing to a one-second ISI to synchronizing to a two-second ISI (Grondin, 2012) and so does the perceived difficulty of synchronizing (Bååth & Madison, 2012). A related notion is the slower limit of rhythm perception, suggested to lie between a 1.5 second and a 3 second ISI (Repp, 2006).

Brown (1997) hypothesizes that the mechanism responsible for rhythmic timing above a two second ISI requires more attentional and executive resources , support for which has been presented by Miyake, Onishi, and Pöppel (2004). They showed that participants' ability to synchronize to a metronome sequence while simultaneously performing a memory task is more impaired at ISIs above two seconds compared to shorter ISIs. However, a similar study by Holm, Ullén, and Madison (2013) showed no evidence of any interaction effect between the ISIs of the sequences and whether participants performed an executive function distractor task or not. Both studies used a dual-task setup: a standard experimental paradigm that aims to discern whether two tasks depend on the same limited cognitive capacity, such as executive control or short term memory (Pashler, 1994). One of the tasks – sometimes referred to as the *main task* – is the task under study; the second task – here called the *distractor task* – is assumed *a priori* to tax a certain cognitive capacity. Participants are either asked to perform the main task and the distractor task simultaneously or to perform solely the main task. Performance between the two conditions is then compared. If the distractor task interferes with performance in the main task, this is taken to indicate that both tasks rely – at least to some extent – on the same limited cognitive capacity.

Our study investigated whether rhythmic timing requires more attentional resources when the tempo is slow compared to when it is fast, where a fast tempo is loosely defined as an ISI shorter than 1500 ms and a slow tempo as an ISI longer than 1500ms. In keeping with Miyake et al. (2004) and Holm et al. (2013), we used a dual-task paradigm, with a rhythmic timing task as the main task and a distractor task selected to require attentional resources and executive control.

More specifically, the main task was a sensorimotor synchronization task in where participants were asked to tap their index finger in time with metronome sequences. The tempo of the sequences included ISIs of 600 ms to 3000 ms. The distractor task was a novel variation on the n-back task. The n-back task was chosen because it is commonly used to assess executive function (Baddeley, 2003; Kane, Conway, Miura, & Colflesh, 2007) and because its design facilitates straightforward varying of attentional resource and executive control demands (Chatham et al., 2011; Smith & Jonides, 1999). The difficulty with using the standard n-back task in a dual-task setup is that it requires participants to make responses throughout the task, either verbally or by key press. These motor responses might well interfere with the motor responses in the sensorimotor synchronization task, making it difficult to infer whether any task interference is due to attentional interference or motor interference. Therefore, a novel variant of the *n-back task* was used, here called the *covert n-back task*, where the participant makes no overt responses during the task.

If the distractor task should be found to impair rhythm timing more at slow tempi than at fast tempi then this would accord with models that assume different timing mechanisms being recruited depending on time scale (P. A. Lewis & Miall, 2003). It would also accord with models that assume a dedicated rhythm-perception mechanism and a slower limit for rhythm perception, for example, Large's (2008) proposed resonance model of rhythm perception.

1 Method

1.1 Participants

Twenty-four participants were recruited via public advertising (11 women and 13 men, mean age: 27 years, SD: 6 years). Seventeen participants reported having experience playing a musical instrument and the mean reported number of years of regular practice was 13 (SD: 10).

1.2 Material

The main task was a sensorimotor synchronization task. A covert response n-back task was used as distractor task.

1.2.1 Sensorimotor synchronization task

Participants were asked to synchronize finger taps to isochronous metronome sequences. They were to start as soon as a sequence started and continue until the sequence ended. They were requested not to subdivide the beat in any way, for example, by covert counting or by moving their body. A custom-built tapping board consisting of a piezo-electric sensor mounted on 5cm^2 corrugated fiberboard recorded the timing of the finger taps (see Bååth, 2011 for details). Participants tapped with their index finger, their hand resting on a plastic foam cushion. The stimuli consisted of isochronous sequences of 440 Hz square wave tones of 20 ms, where each sequence was 45 seconds long. Sequences were presented at five tempi with ISIs of 600, 897, 1342, 2006, and 3000 ms, selected so as to be equidistant on a log scale. An Arduino microcontroller generated the sounds and registered the taps.

1.2.2 Covert response N-back task

Participants were asked to perform a visuospatial 2-back task. The visual stimuli was modeled after (Jaeggi et al., 2007). It consisted of a white 3×3 grid on a black background, with a white fixation cross in the middle and a blue square in one of eight outer grid positions (see Figure 1). The blue square changed position every 2150 ms, including 700 ms to fade in and 700 ms to fade out. These time intervals were chosen so that the presentation of the blue square would not regularly coincide with stimuli in the sensorimotor synchronization task.

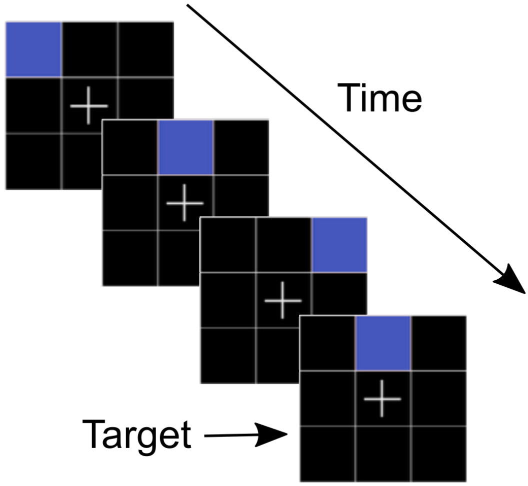


Figure 1: The stimuli presented in the covert response 2-back task.

A given stimulus presentation constituted a target if the blue square's current position was the same as two positions back. The square's position was randomized so that, on average, half the presentations were targets. Instead of responding overtly to each target, participants were instructed to count the number of targets silently and report the total at the conclusion of each trial. This variation of the n-back task was used as the the responses during the 2-back task could otherwise interfere with the motor part of the sensorimotor synchronization task. Trials were 47 seconds long: slightly longer than the sensorimotor synchronization task trials. The distractor task was implemented in the Java programming language using the Processing framework (Reas & Fry, 2007).

1.3 Procedure

Participants were tested individually in a quiet room. Sessions began with a number of practice trials. First a sensorimotor synchronization trial at 600ms ISI, then an n-back only trial, and finally a trial where the two tasks were presented simultaneously. After this the participant was given four n-back-only-trials to establish baseline performance. For the sensorimotor-synchronization task, audio was delivered through a pair of closed headphones. The n-back distractor task used a 27″ monitor positioned 50 cm from the participant.

The experiment proper consisted of four blocks of five sensorimotor synchronization trials, one for each of the five ISI levels. The order of the trials within each block was randomized. Either the first and third or the second and fourth blocks included the n-back distractor task and whether or not a participant started with a distractor block was also randomized. Each participant performed 20 trials, four at each ISI level, where two included the distractor task and two were without the distractor task.

1.4 Analysis

The first three taps in every sensorimotor-synchronization trial were discarded to use only those taps where participants had time to synchronize to the sequence. For each tap, tone-to-tap asynchrony was calculated as the time difference between the tone and the tap, a negative asynchrony indicating that the tap preceded the tone and vice versa. Asynchrony SD was taken as a measure of timing variability and it was estimated for each participant and ISI level using the Bayesian hierarchical method described in (Bååth, 2015). This method was used instead of the conventional sample SD, as it has been shown to yield more accurate estimates of timing variability when participants synchronize to slow sequences. Timing variability is here used as the measure of performance in the sensorimotor synchronization task with low variability taken to indicate high performance. As a second measure of timing performance, we used the coefficient of variation, a measure of timing variability relative to the ISI, calculated for each participant and condition as the asynchrony SD divided by the ISI.

Statistical analysis was performed using the statistical computing environment *R* (R Core Team, 2012). Because timing variability was measured at five different ISIs for each participant, a linear mixed-effects model was used to asses how timing variability changed as a function of ISI and distractor condition. Mixed-effects model analyses were performed using the package *lme4* (Bates, Mächler, Bolker, & Walker, 2014).

2 Results

The dependence of timing variability on ISI and distractor condition – control or n-back – was investigated by fitting a linear mixed-effects model, using $\log_e$ asynchrony SD as the outcome variable and ISI, distractor condition, and the interaction between ISI and distractor condition as the predictor variables. The ISI was standardized prior to fitting the model and the asynchrony SD was $\log_e$ transformed, as

Coefficient	Estimate	95% CI	p
Intercept	4.25	[4.20, 4.36]	-
ISI	0.74	[0.69, 0.78]	< .001
Distractor condition	0.35	[0.29, 0.40]	< .001
ISI × distractor condition	0.11	[0.060, 0.17]	.< .001

Table 1: Estimated coefficients for the mixed-effects model with ISI and distractor condition as predictors.

it was found to have a right skewed distribution. Table 1 reports the resulting parameter estimates. Figure 2 shows $\log_e$ asynchrony SD as a function of ISI, with superimposed regression lines from the mixed-effects model.

The effect of both ISI and distractor condition on asynchrony SD was statistically significant, as was the interaction effect, where the difference between the control and the n-back condition increased with longer ISI. For example, the mean difference in $\log_e$ asynchrony SD between the control and the n-back condition was more than three times as large at the 3000 ms ISI level compared to the 600 ms ISI level. This interaction effect can also be seen when looking at the difference between each participant's asynchrony SD under the two conditions. Figure 3 shows how the difference increases as a function of ISI; a positive difference means that the timing variability was higher in the n-back than in the control condition. In this and all subsequent figures, error bars show 95% confidence intervals (CI) calculated as 1.96 × standard error.

The effect of the distractor condition can be seen in other measures of timing performance. Figure 4 shows the mean coefficient of variation as a function of ISI and distractor condition; the difference between the two distractor conditions increases with longer ISIs. Another measure of timing performance is the percentage of reactive responses (Miyake et al., 2004; Repp & Doggett, 2007), defined as the percentage of responses that overshot the target tones by more than 100 ms. Figure 5 shows very few reactive responses at 600 and 897 ms ISIs. For longer ISIs, the percentage of reactive responses was greater in the n-back condition.

Timing performance decreased under the n-back condition and so did performance in the n-back task at slower tempi. Baseline n-back performance was calculated for each participant as the mean number of errors made in the four n-back only trials. The difference between each participant's baseline performance and performance during the

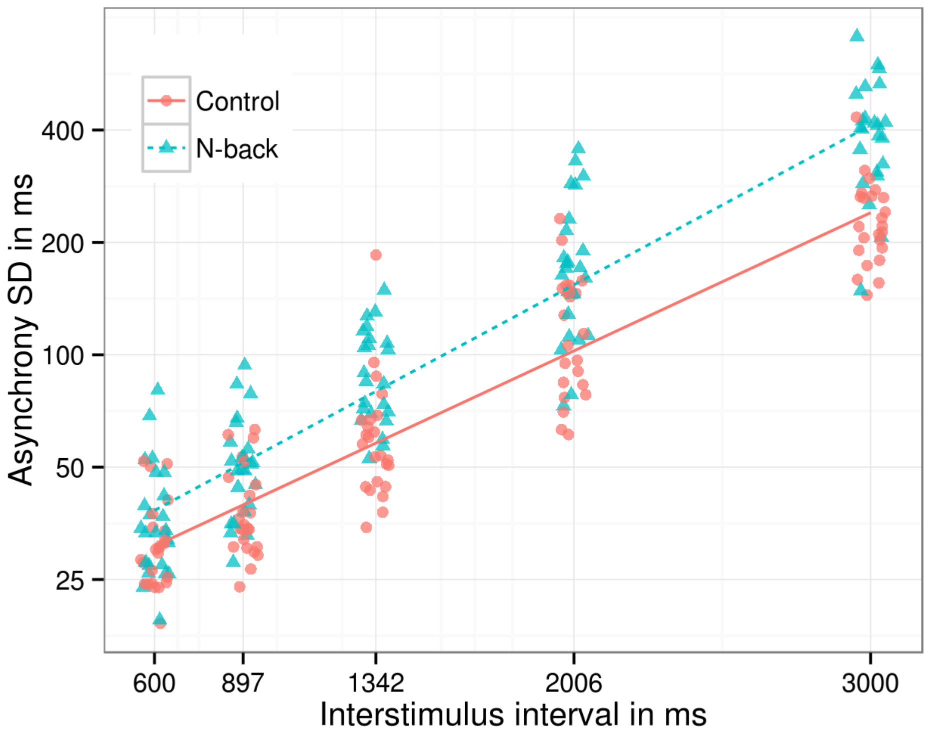


Figure 2: Mean timing variability as measured by asynchrony SD for all participants and ISI levels. The regression lines show the results of the mixed-effects model analysis.

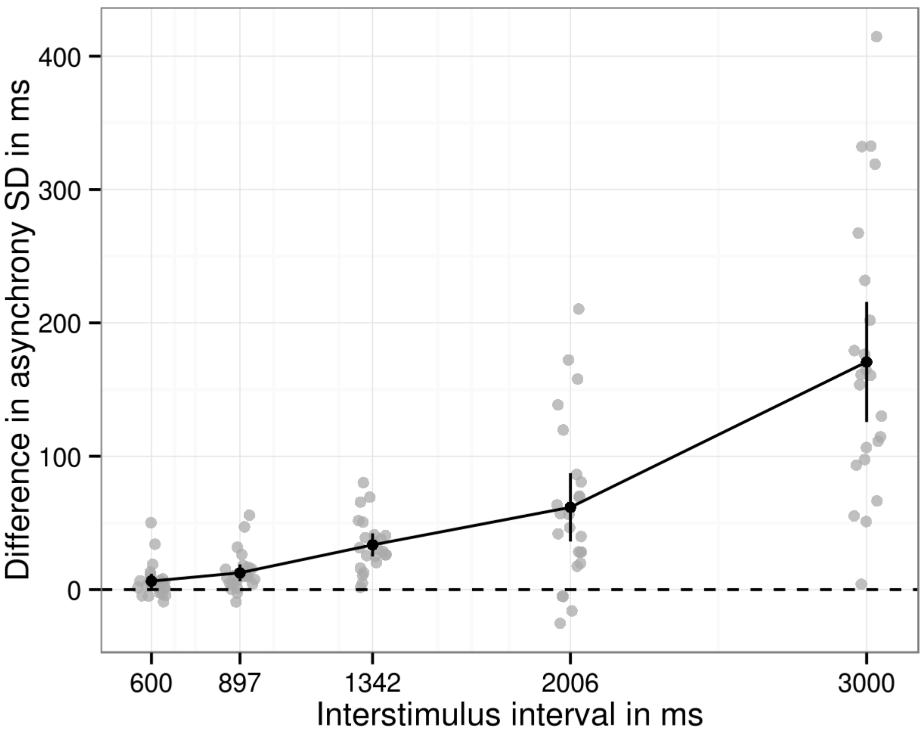


Figure 3: Difference between asynchrony SD under the control and n-back conditions for each participant and ISI level. The connected points show the grand means. The error bars show 95% CIs.

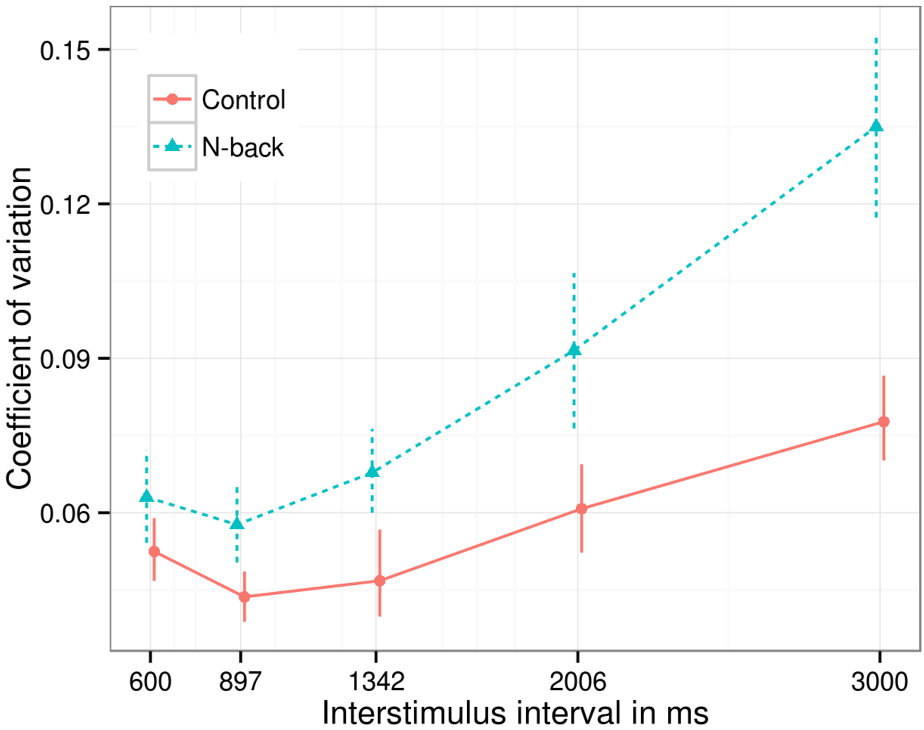


Figure 4: Mean coefficient of variation in the control and n-back condition. The error bars show 95% CIs.

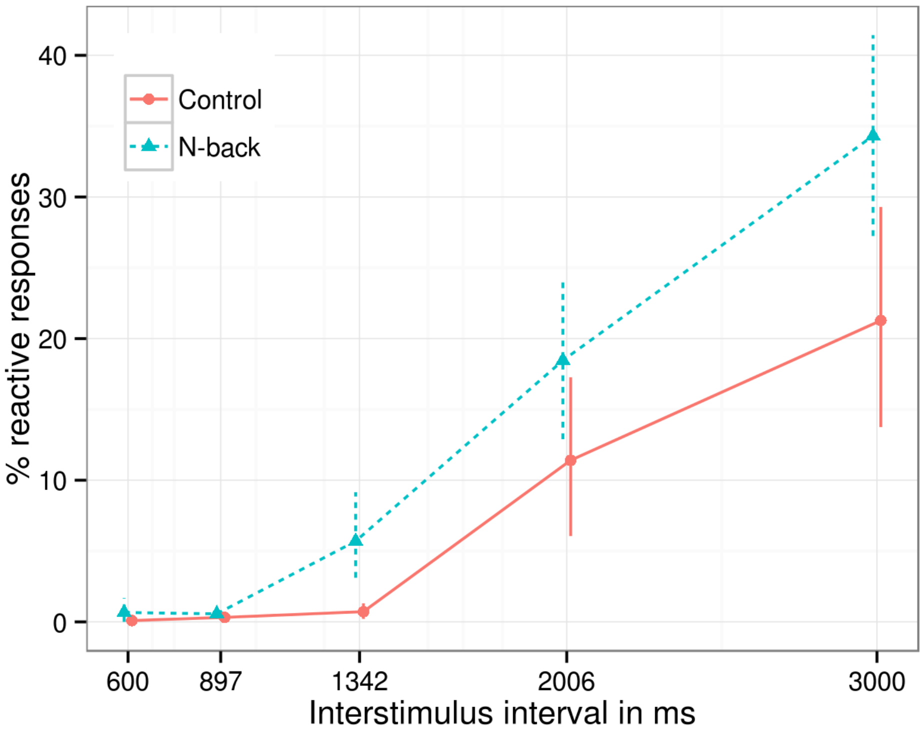


Figure 5: Mean percentage of reactive responses in the control and n-back conditions. The error bars show 95% CIs.

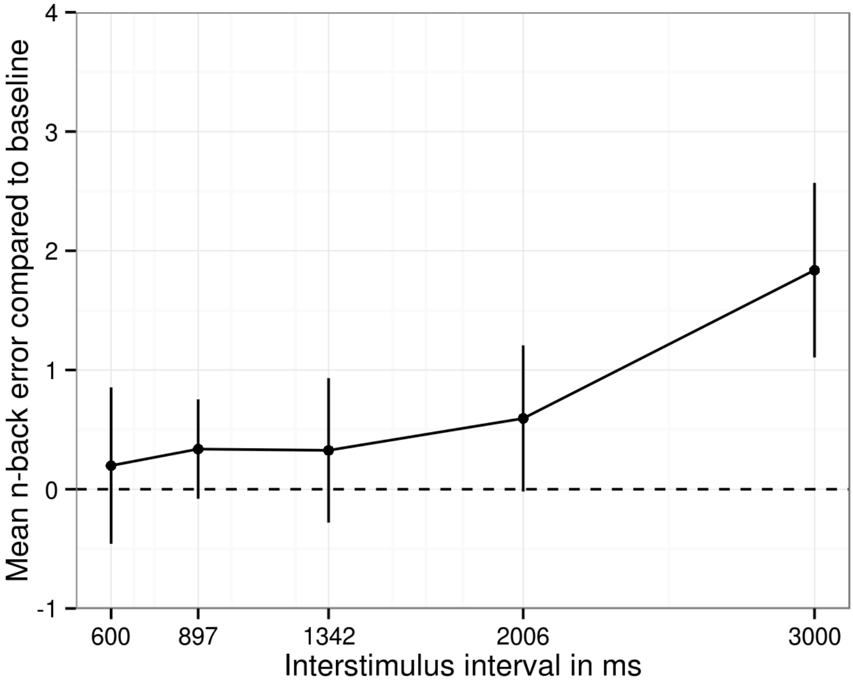


Figure 6: Mean n-back error above the baseline error in the n-back only trials. Error bars show 95% CIs.

experiment proper was then calculated for each ISI level. Figure 6 show the mean n-back error compared to baseline. The difference was statistically significantly different from zero at the 3000 ms ISI level (one sample t-test, M = 1.8, t(22) = 5.2, $p < 0.001$). For shorter ISIs, average n-back performance was less than one error above baseline.

3 Discussion

Many models of human timing and time perception have been proposed. One important way in which they differ is whether they posit a single, overreaching mechanism for timing or assume that timing recruits different mechanisms depending on the nature of the task. Regarding rhythmic timing, it has been proposed that different mechanisms are responsible depending on the tempo (Grondin, 2012). Relevant here is the notion of a slower limit of rhythm perception, a proposed temporal boundary where perceiving and synchronizing to a rhythmic sequence goes from being effortless and automatic to requiring attention and executive control (Repp, 2006). The present study used a dual-task setup to investigate whether rhythmic timing requires more attentional resources at slow tempi compared to comfortable tempi. The main task was a sensorimotor synchronization task where participants tapped their finger in time with metronome sequences and the distractor task was a covert response n-back task.

The results point towards rhythmic timing requiring more attentional resources at slow tempi. At the slowest tempo – at an inter-stimulus interval (ISI) of 3000 ms – performance of the tapping task and n-back task simultaneously resulted in a significant performance degradation in both tasks. The fastest tempo – at an ISI of 600 ms – also saw dual-task interference, however, the magnitude of interference was much lower in comparison. It is difficult to identify a particular tempo at which dual-task interference becomes significant. Looking at the different performance measures, the largest increase in interference occurs between an ISI of 897 and 1342 ms for the log asynchrony SD, and between an ISI of 2006 and 3000 ms for the coefficient of variation, percentage of reactive responses, and number of errors in the n-back task. The results reflect the authors' own experience when piloting the experiment and participants' informal verbal reports: keeping the beat with a fast metronome while doing a 2-back task is easy; keeping the beat with a metronome that strikes every third second while doing a 2-back task is hard.

The results are consistent with those from the study by Miyake et al. (2004), who asked participants to perform a word-memory task and rhythmic tapping task. While Miyake et al. did not analyze timing variability, they found that participants produced more reactive responses when both tasks were performed simultaneously. As with the present study, the difference was not found at shorter ISIs but became pronounced at 1800 ms ISI.

The results are not consistent with a recent study by Holm et al. (2013), who asked participants to perform a rhythmic timing task under either a low or high cognitive load condition. They did *not* find an effect of cognitive load on timing performance, nor did they find an interaction between cognitive load and sequence tempo. The results may be due to the distractor task used. Under the low cognitive load condition in Experiment 1 in Holm et al. participants were asked to tap the rhythm on two buttons using the sequence (1, 2, 1, 2, ...). In experiment 2, participants instead used four buttons and the sequence (1, 2, 3, 4, 1, 2, ...). Under the high cognitive load condition, participants were instead asked to tap the rhythm in a random sequence. A possible reason for why no task interference was observed when participants synchronized at a slow tempo is because the distractor task is easier to perform at a slower compared to a faster tempo, i.e., the distractor task is not invariant to the sequence tempo. At 1000 ms a participant must make twice as many random decisions as at an ISI of 2000 ms. The cognitive load resulting from the timing task might indeed have been heavier at the slower tempi, but no interference effect was manifest, because the cognitive load resulting from the distractor task was lighter at the slower tempi.

In conclusion, the present study shows that, when the tempo is sufficiently slow, performing rhythmic timing demands attentional resources and involvement of executive control. These results accord with neural models of timing that suggest a dedicated, *automatic* timing mechanism for short intervals and a general, *cognitive* timing mechanism for longer intervals (P. A. Lewis & Miall, 2003). The results might also be explained though by a single timing mechanism that requires more cognitive resources at slower tempi. As shown in this study, rhythmic timing requires more cognitive resources the slower the tempo, and both attentional resources and executive control are presumably limited. Therefore, independent of whether rhythmic timing depends on one or several mechanisms, this study supports the view that rhythm perception and rhythmic timing have a slower limit.

Acknowledgement

This research was supported by the Linnaeus environment Thinking in Time: Cognition, Communication and Learning, Swedish Research Council grant number 349-2007-8695. T. Tjøstheim and M. Lingonblad contributed equally to this work. The authors would like to thank Frida Nelhans, David Sivén, Ryo Yamazaki, and Henrik Siljebråt for their help with the data collection, and Christian Balkenius and Joel Parthemore for helpful comments on this manuscript.

References

Bååth, R. (2011). Construction of a Low Latency Tapping Board. *LUCS minor, 17*. Retrieved from `http://www.sumsar.net/papers/ICMPC_2012_rasmus_baath.pdf`

Bååth, R. (2015). Estimating the distribution of sensorimotor synchronization data: A Bayesian hierarchical modeling approach. *Behavior Research Methods*. doi: 10.3758/s13428-015-0591-2

Bååth, R., & Madison, G. (2012). The subjective difficulty of tapping to a slow beat. In *Proceedings of the 12th international conference on music perception and cognition* (pp. 82–85). Retrieved from `www.sumsar.net/papers/ICMPC_2012_rasmus_baath.pdf`

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews. Neuroscience*, *4*(10), 829–839. doi: 10.1038/nrn1201

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv e-prints*. Retrieved from `http://arxiv.org/abs/1406.5823`

Brown, S. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & psychophysics*. Retrieved from `http://link.springer.com/article/10.3758/BF03205526`

Chatham, C. H., Herd, S. a., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R., & Friedman, N. P. (2011). From an executive network to executive control: a computational model of the n-back task. *Journal of cognitive neuroscience*, *23*(11), 3598–619. doi: 10.1162/jocn_a_00047

Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Attention, perception & psychophysics*, *72*(3), 561–82. doi: 10.3758/APP.72.3.561

Grondin, S. (2012). Violation of the Scalar Property for Time Perception Between 1 and 2 Seconds: Evidence From Interval Discrimination, Reproduction, and Categorization. *Journal of experimental psychology. Human perception and performance*, *38*(4), 880–890. doi: 10.1037/a0027188

Holm, L., Ullén, F., & Madison, G. (2013). Motor and executive control in repetitive timing of brief intervals. *Journal of experimental psychology. Human perception and performance*, *39*(2), 365–80. doi: 10.1037/a0029142

Ivry, R. B. (1996). The representation of temporal information in perception and motor control. *Current opinion in neurobiology*, *6*(6), 851–7. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/9000026`

Jaeggi, S. M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig, W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, affective & behavioral neuroscience*, *7*(2), 75–89. doi: 10.3758/CABN.7.2.75

Kane, M. J., Conway, A. R. a., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of experimental psychology. Learning, memory, and cognition*, *33*(3), 615–622. doi: 10.1037/0278-7393.33.3.615

Large, E. (2008). Resonating to musical rhythm: Theory and experiment. In S. Grondin (Ed.), *The psychology of time* (pp. 189–231). Emerald Group Publishing. doi: 10.1016/B978-0-08046-977-5.00006-5

Large, E. (2010). Neurodynamics of music. In M. Riess Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception* (pp. 201–231). Springer. doi: 10.1007/978-1-4419-6114-3

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

Lewis, P. A., & Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Current Opinion in Neurobiology*, *13*(2), 250–255. doi: 10.1016/S0959-4388(03)00036-9

Lewis, P. a., & Miall, R. C. (2006). Remembering the time: a continuous clock. *Trends in cognitive sciences*, *10*(9), 401–6. doi: 10.1016/j.tics.2006.07.006

Mates, J., Müller, U., Radil, T., & Pöppel, E. (1994). Temporal Integration in Sensorimotor Synchronization. *Journal of Cognitive Neuroscience*, *6*(4), 332–340. doi: 10.1162/jocn.1994.6.4.332

Miyake, Y., Onishi, Y., & Pöppel, E. (2004). Two types of anticipation in synchronization tapping. *Acta neurobiologiae experimentalis*, *64*(3), 415–26. Retrieved from `http://www.ane.pl/pdf/6428.pdf`

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, *116*(2), 220–244. doi: 10.1037/0033-2909.116.2.220

Pöppel, E. (2004). Lost in time: A historical frame, elementary processing units and the 3-second window. In *Acta neurobiologiae experimentalis* (Vol. 64, pp. 295–301). Retrieved from `http://www.ane.pl/pdf/6428.pdf`

R Core Team. (2012). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from `http://www.r-project.org`

Reas, C., & Fry, B. (2007). *Processing: a programming handbook for visual* *designers and artists*. The MIT Press.

Repp, B. H. (2006). Rate Limits of Sensorimotor Synchronization. *Advances in Cognitive Psychology*, *2*(2), 163–181. doi: 10.2478/v10053-008-0053-9

Repp, B. H., & Doggett, R. (2007). Tapping to a Very Slow Beat: A Comparison of Musicians and Nonmusicians. *Music Perception: An Interdisciplinary Journal*, *24*(4), 367–376. doi: 10.1525/mp.2007.24.4.367

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science (New York, N.Y.)*, *283*(5408), 1657–1661. doi: 10.1126/science.283.5408.1657

Teki, S., Grube, M., & Griffiths, T. (2011). A unified model of time perception accounts for duration-based and beat-based timing mechanisms. , *5*(January), 1–7. doi: 10.3389/fnint.2011.00090