

**How musical rhythms entrain the human brain:
Clarifying the neural mechanisms of sensory-motor entrainment to
rhythms**

Tomas Lenc

A dissertation submitted for the degree of Doctor of Philosophy

MARCS Institute for Brain, Behaviour, and Development
Western Sydney University, Australia

September 2020

Acknowledgements

These last four special and exciting years of my life would never happen, and would mean nothing without the people who stood beside me.

Most importantly, I am immensely grateful to my primary supervisor Sylvie Nozaradan, for all her support as a mentor, but also as a caring friend. Thank you Sylvie for all the inspiring thoughts and debates, for giving me the opportunity and stimulation to grow, for always looking out for me, and for believing in me even in times when I did not. You are and always will be a great role model for me as a researcher and as a person.

I would like to thank my supervisors Peter and Manuel, for their endless support and encouragement. Thank you for the constructive discussions, for sharing your great experience, and for always being patient, kind and friendly. Indeed, I could have never imagined having a better supervisory panel.

My gratitude goes to all the friends and colleagues at MARCS, particularly Felix (for making and talking music), Bronson (for teaching me how to be cool and use Twitter), and everyone from Music Cognition and Action. Thank you for every smile, joke, encouragement, for listening to me complain, and for participating in my endless and boring experiments.

I am deeply indebted to the MARCS admin team for their patience with my terrible organization skills and silly questions and to the MARCS tech team, especially Donovan (for never hesitating to construct new toys for us).

My thanks go to all the great and inspiring people from NOCIONS, including Andre (for welcoming me and supporting me), Arthur (for every scientific debate and for being a caring friend), and Dominika (for all your positive energy). I am forever grateful to Jerome for his immense care and support in Belgium.

The ideas fundamental for this thesis did not emerge from nowhere, but were a result of all the years I spent making music with my closest friends. Thank you Samo, Tibi, Hajty, Lenka, Maria M., Maria R., Michal, and all the amazing people who shaped me not only as a musician, but also as a person.

I want to thank my girlfriend Hana and her beautiful family for allowing me to become a part of their lives and to feel the love and warmth of their home.

I also want to thank my siblings, Petra, Lukas and Tadeas for always having my back and being my closest mates.

My final, and greatest gratitude goes to my parents who have always loved me and supported my dreams. I would never be where I am without you.

Statement of Authentication

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.



Tomas Lenc

Table of Contents

List of Tables	vi
List of Figures.....	vii
Abbreviations	ix
Abstract	x
1 Introduction.....	2
1.1 Defining the phenomenon	3
1.1.1 Musical meter as perception of a nested set of periodic pulses	3
1.1.2 Functions of musical meter	7
1.1.2.1 Does meter perception rely on dynamic attending, expectation, entrainment models or predictive coding?	8
1.1.3 Primacy of one pulse (beat) in meter	15
1.2 Measuring the phenomenon.....	19
1.2.1 Indirect measures	19
1.2.1.1 Ratings	19
1.2.1.2 Other indirect behavioral measures	21
1.2.1.3 Functional magnetic resonance imaging (fMRI)	22
1.2.1.4 Mismatch negativity (MMN) studies	23
1.2.2 Direct measures.....	25
1.2.2.1 Meter as transformation	26
1.2.2.2 The need for a better method to capture periodic contrast	27
1.2.2.3 Frequency-tagging approach to measure contrast at meter periodicities ..	31
1.2.2.4 Contrast affected by differentiation of events	31
1.2.2.5 Meter contrast affected by arrangement of events	38
1.2.2.6 Contrast affected by fine time locking	57
1.3 Nature of the phenomenon	69
1.3.1 Transformation from sound to behavior	69
1.3.2 Transformation in the brain	70
2 Study 1: Attention affects overall gain but not selective contrast at meter frequencies in the neural processing of rhythm	75

2.1	Abstract	76
2.2	Introduction.....	77
2.3	Materials and methods	80
2.3.1	Participants.....	80
2.3.2	Auditory stimuli	81
2.3.3	Frequency-tagging analysis	82
2.3.4	Models of subcortical auditory processing.....	83
2.3.5	Early auditory responses	85
2.3.6	Experimental design and procedure.....	86
2.3.7	EEG recording and preprocessing.....	89
2.3.8	Overall EEG response magnitude	91
2.3.9	Relative EEG response at meter frequencies	92
2.3.10	Behavioral analyses	94
2.3.11	Statistical analyses.....	94
2.4	Results	95
2.4.1	Behavioral results	95
2.4.2	Overall EEG response magnitude	98
2.4.3	Relative EEG response at meter frequencies	98
2.4.4	Comparison of EEG responses with models of subcortical auditory processing 101	
2.4.5	Comparison of higher-level and early auditory EEG responses.....	102
2.5	Discussion	103
2.5.1	Wide range of low-level and higher-level processes in meter perception.....	103
2.5.2	Dissociation between overall gain and selective contrast at meter periodicities 105	
2.5.3	Robust responses at meter periodicities even with low meter contrast in the input 106	
2.5.4	Evidence for robust meter processing complementary to MMN studies of passive listening.....	109
2.5.5	Conclusions.....	110
2.6	Supplementary Material.....	111

2.6.1	Control analysis of higher-level and early auditory EEG responses excluding the highest (5 Hz) frequency.....	111
2.6.2	EEG oscillatory alpha activity.....	116
3	Study 2: Neural tracking of the musical beat is enhanced by low-frequency sounds	119
3.1	Abstract	120
3.2	Significance Statement.....	121
3.3	Introduction.....	122
3.4	Results	124
3.5	Discussion.....	129
3.5.1	Low-tone benefit in syncopated rhythm	131
3.5.2	Critical role of temporal attention.....	132
3.5.3	Conclusion	133
3.6	Materials and methods	134
3.7	Supplementary Information	138
3.7.1	Main Experiment	138
3.7.2	Control Experiment 1: Effect of sound intensity	143
3.7.2.1	Materials and Methods	143
3.7.2.2	Results	144
3.7.3	Control Experiment 2: Effect of behavioral task.....	146
3.7.3.1	Materials and Methods	146
3.7.3.2	Results	147
4	Study 3: Neural and behavioral evidence for frequency-selective context effects in rhythm processing in humans	150
4.1	Abstract	151
4.2	Introduction.....	152
4.3	Materials and Methods	154
4.3.1	Participants.....	154
4.3.2	Data and code availability	155
4.3.3	Auditory stimulation.....	155
4.3.4	Stimulus analysis	157
4.3.5	Experimental design and procedure.....	161
4.3.6	EEG recording and preprocessing.....	162

4.3.7	Frequency-domain analysis of EEG response	164
4.3.8	Tapping analysis	166
4.3.9	Head movement analysis	168
4.3.10	Statistical analyses.....	168
4.4	Results	169
4.4.1	Tapping	169
4.4.2	Frequency-domain analysis of EEG.....	170
4.5	Discussion	174
4.5.1	No one-to-one mapping between sensory input and perception	175
4.5.2	Evidence against evoked responses passively tracking low-level acoustic features of the rhythmic input	177
4.5.3	Context effect is short-lived in neural activity but long-lasting in behavior	178
4.5.4	Conclusion	179
4.6	Supplementary Materials	180
4.6.1	Supplementary Results: Control analyses of EEG data	180
5	Discussion and Perspectives	185
5.1	Future perspectives	191
5.1.1	Functional anatomy of the transformation	193
5.1.2	Role of learning and training	196
5.1.3	Further examining short-term context effects	197
5.1.4	Meter phase	198
5.2	Conclusions.....	199
	References.....	200

Note that Studies 2 and 3 have been published in peer-reviewed journals. Figures, and tables in these chapters have been relabeled relative to their position in the dissertation.

Study 2: Lenc, T., Keller, P. E., Varlet, M., and Nozaradan, S. (2018). Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proc. Natl. Acad. Sci.* 115, 8221–8226. doi:10.1073/pnas.1801421115.

Study 3: Lenc T, Keller PE, Varlet M, Nozaradan S (2020) Neural and behavioral evidence for frequency-selective context effects in rhythm processing in humans. *Cereb. Cortex. Commun.* doi:10.1093/texcom/tgaa037

List of Tables

Table 2.1. Comparison of mean z-scored amplitude at meter-related frequencies between higher-level EEG responses and models of auditory subcortical processing using one-sample t-tests.	99
Table 2.S1. Comparison of mean early auditory response amplitude averaged across all sidebands against zero using non-parametric Wilcoxon signed rank tests.	112
Table 2.S2. Control analysis of mean z-scored amplitude at meter-related frequencies without taking the highest frequency (5 Hz) into account. Comparison between higher-level EEG responses and models of auditory subcortical processing using one-sample t-tests.	113
Table 2.S3. Comparison of mean z-scored amplitude at meter-related frequencies between early auditory EEG responses and models of auditory subcortical processing using one-sample t-tests.	114
Table 2.S4. Control analysis of mean z-scored amplitude at meter-related frequencies without taking the highest frequency (5 Hz) into account. Comparison between early auditory EEG responses and models of auditory subcortical processing using one-sample t-tests.	115
Table 3.S1. Descriptive statistics for the main experiment.	142
Table 3.S2. Descriptive statistics for Control Experiment 1.	145
Table 3.S3. Descriptive statistics for Control Experiment 2.	148
Table 4.1. Prominence of meter-related frequencies in the EEG response compared between the first and all subsequent segments, separately for the two sequence directions, and for musicians (N = 16) and non-musicians (N = 16).	172

List of Figures

Figure 1.1. Representation of the rhythmic stimulus and perceived meter.	5
Figure 1.2. Issues of an adaptive oscillator when tracking pulse in a rhythmic pattern.	11
Figure 1.3. Multiple realizability of pulse representation.	28
Figure 1.4. Difficulties of time-domain approaches to measure contrast.	30
Figure 1.5. Using frequency-tagging to measure contrast in an isochronous sequence of events.	34
Figure 1.6. Contrast created by differences in kernel shape.	35
Figure 1.7. Illustration of the baseline subtraction method.	35
Figure 1.8. Contrast at one pulse periodicity in a rhythmic sequence consisting of a repeating rhythmic pattern.	37
Figure 1.9. Different methods to assess contrast at metric periodicities using abstract representation of rhythmic sequences.	41
Figure 1.10. From grid representation of a rhythmic sequence to continuous modulated signal.	45
Figure 1.11. Response shape can attenuate or enhance contrast at specific periodicities.	49
Figure 1.12. Simulation of contrast enhancement in a continuously modulated signal.	53
Figure 1.13. Results of meta-analysis across 100 simulated experiments, testing for spurious effects.	55
Figure 1.14. Example of uncertainty when assigning taps to individual events in a rhythmic pattern constructed on an isochronous grid.	60
Figure 1.15. Examples of tapping data.	61
Figure 1.16. Sensitivity of different methods to systematic contrast locked onto a metric pulse.	62
Figure 1.17. Importance of taking tapping force (i.e. amplitude) into account when assessing contrast at meter periodicities.	64
Figure 1.18. Sensitivity of different methods to changes in systematic alignment and tapping period.	66
Figure 2.1. Stimulus design and higher-level EEG responses.	88
Figure 2.2. Diagram showing dissociation between higher-level and early auditory EEG responses.	89

Figure 2.3. Characteristics of the higher-level EEG responses.	96
Figure 2.4. Characteristics of the early auditory EEG responses.	97
Figure 2.5. Comparison of higher-level EEG responses with models of auditory subcortical processing.....	100
Figure 2.6. Comparison of prominence of meter-related frequencies in the higher-level and early auditory EEG responses.	102
Figure 2.S1. EEG power at the alpha frequency elicited across the different tasks at posterior channels.....	117
Figure 3.1. Spectra of the acoustic stimuli (processed through the cochlear model) and EEG responses (averaged across all channels and participants; $n = 14$; shaded regions indicate SEMs; see Morey, 2008).	125
Figure 3.2. Grand average topographies ($n = 14$) of neural activity measured at meter-related (left column) and meter-unrelated (right column) frequencies for the unsyncopated and syncopated rhythm conveyed by low or high tones.	128
Figure 3.3. Effect of tone frequency on the selective enhancement of EEG activity at beat- and meter-related frequencies.	129
Figure 3.S1. Tapping responses.	141
Figure 4.1. Illustration of the sequence generation method.....	157
Figure 4.2. Cochlear model, EEG, and tapping spectra.....	164
Figure 4.3. Mean z-scored amplitudes at meter-related frequencies in the cochlear model, EEG, and tapping response.....	173
Figure 4.4. Topographies of the mean EEG amplitude at meter-related frequencies.....	174
Figure 4.S1. Analysis of the stimulus sequences used in the EEG session.....	181
Figure 4.S2. Evolution of syncopation scores across segments, assuming different metrical interpretations of the sequences (nested pulses at rates of {2,4}, {3,6}, and {2,6} events).	182
Figure 4.S3. Analysis of inter-tap interval (ITI) error across sequence directions and segments.	182

Abbreviations

EEG	Electroencephalogram
MEG	Magnetoencephalogram
ERP	Event-related potential
MMN	Mismatch negativity event-related potential
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
SNR	Signal-to-noise ratio
fMRI	Functional magnetic resonance imaging
BOLD	Blood-oxygen-level-dependent signal

Abstract

When listening to music, people across cultures tend to spontaneously perceive and move the body along a periodic pulse-like meter. Increasing evidence suggests that this ability is supported by neural mechanisms that selectively amplify periodicities corresponding to the perceived metric pulses. However, the nature of these neural mechanisms, i.e., the endogenous or exogenous factors that may selectively enhance meter periodicities in brain responses to rhythm, remains largely unknown. This question was investigated in a series of studies in which the electroencephalogram (EEG) of healthy participants was recorded while they listened to musical rhythm. From this EEG, selective contrast at meter periodicities in the elicited neural activity was captured using frequency-tagging, a method allowing direct comparison of this contrast between the sensory input, EEG response, biologically-plausible models of auditory subcortical processing, and behavioral output. The results show that the selective amplification of meter periodicities is shaped by a continuously updated combination of factors including sound spectral content, long-term training and recent context, irrespective of attentional focus and beyond auditory subcortical nonlinear processing. Together, these observations demonstrate that perception of rhythm involves a number of processes that transform the sensory input via fixed low-level nonlinearities, but also through flexible mappings shaped by prior experience at different timescales. These higher-level neural mechanisms could represent a neurobiological basis for the remarkable flexibility and stability of meter perception relative to the acoustic input, which is commonly observed within and across individuals. Fundamentally, the current results add to the evidence that evolution has endowed the human brain with an extraordinary capacity to organize, transform, and interact with rhythmic signals, to achieve adaptive behavior in a complex dynamic environment.

1 Introduction

1.1 Defining the phenomenon

To investigate any perceptual phenomenon, the most important prerequisite is its definition. Defining the basic properties of a phenomenon can subsequently guide the ways we measure its correlates in behavior, or in neural activity. An illustrative example from a different perceptual domain could be defining visual categorization as a process that provides (i) different responses to different sensory inputs (discriminates) and (ii) provides identical response to physically different sensory inputs that contain same higher-level category (generalizes). Once such definition is established, a behavioral paradigm can be developed, or a certain measure of neural activity can be meaningfully interpreted as related (or not) to the perceptual phenomenon in question (Rossion et al., 2020).

1.1.1 Musical meter as perception of a nested set of periodic pulses

The aim of this thesis is to examine perceptual organization of musical rhythm. Rhythm in general is conveyed by dynamic changes in physical features of the sensory input, in other words modulation of a stimulus property over time (e.g. a steady pure tone does not provide any rhythmic information, whereas amplitude- or frequency-modulated tone does). While such modulations are by definition continuous, it is often useful to think about rhythm as temporal intervals between discrete sound events defined by transient feature changes, such as tone onsets (indeed, timing of such transient changes is well represented in the auditory system, Hamilton et al., 2018; Daube et al., 2019). The continuous and discrete descriptions of rhythmic structure are closely related, and one or the other will be used according to context. The discrete description is particularly useful to describe rhythmic stimuli as patterns and sequences (Honing and Bouwer, 2018). Rhythmic patterns correspond to ordered sets of time intervals defined by a series of sound events. In music, rhythmic structure is often based on repetition of the same pattern, or concatenation of different patterns into longer rhythmic sequences.

Importantly, the temporal structure of musical rhythms is not random, but often contains a restricted set of time intervals (Desain and Honing, 2003; Jacoby and McDermott, 2017; Roeske et al., 2020). The way humans perceive and respond to rhythmic patterns is based on two fundamental higher-level phenomena. One of them is categorical perception of

intervals. That is, humans map the continuous space of time intervals in a rhythmic pattern onto a limited set of categories (including, but not restricted to, small integer ratios) (Jacoby and McDermott, 2017). At the same time, slight deviations from the perceptual categories are not lost (unlike in other perceptual domains, Goldstone and Hendrickson, 2009) but contribute to perception of expressive timing used by musicians to convey aesthetic information (Desain and Honing, 2003; Honing and Bouwer, 2018). As if the categorical prototype induced by the stimulus was then used as an internal reference point to judge the deviation of the sensory input from this prototype.

The second phenomenon relevant to perceptual organization of musical rhythm constitutes the main topic of the current thesis. Even though rhythmic patterns in music are rarely isochronous, and often not exactly periodic, they typically induce *perception of a nested set of periodic pulses* (a pulse is defined as a series of regularly-spaced points in time). Importantly, these perceived pulses are time-locked onto the temporal structure of the rhythmic sequence (often in a non-trivial way, see below). Similarly to categorical perception of time intervals, while a pattern of intervals is mapped onto regular pulses, the information about the pattern is not lost, but perceptually organized with respect to the internal representation of these pulses (Povel and Essens, 1985; Honing and Bouwer, 2018). In musical terminology, this form of perceptual organization is referred to as *musical meter* (Yeston, 1976; Cohn, 2020).

A particular perceived meter is described by two parameters (see Figure 1.1). One defines the periods for all pulses in the set. Because the set must be nested, the periods of longer pulses must be integer multiples of *all* shorter periods in the set. Another parameter is needed to define the phase, i.e. the alignment of the metric structure with the rhythmic pattern. The meter phase can be described by a single value due to the nesting restriction, which constrains the relative phases between individual pulses.

In some cases, periods and phase of meter can be defined in absolute units of time, such as seconds. However, for patterns that are artificially constructed on a fast isochronous grid (where any interval in the pattern can be defined by an integer number of grid intervals), meter periods and phase can be defined by the number of grid points (in fact, such “grid” method to construct rhythmic patterns is commonly used in most modern music production software). Such relative representation is in line with recent developments of meter theory

(Cohn, 2015, 2020) and provides the required flexibility to classify meters across units and scales.

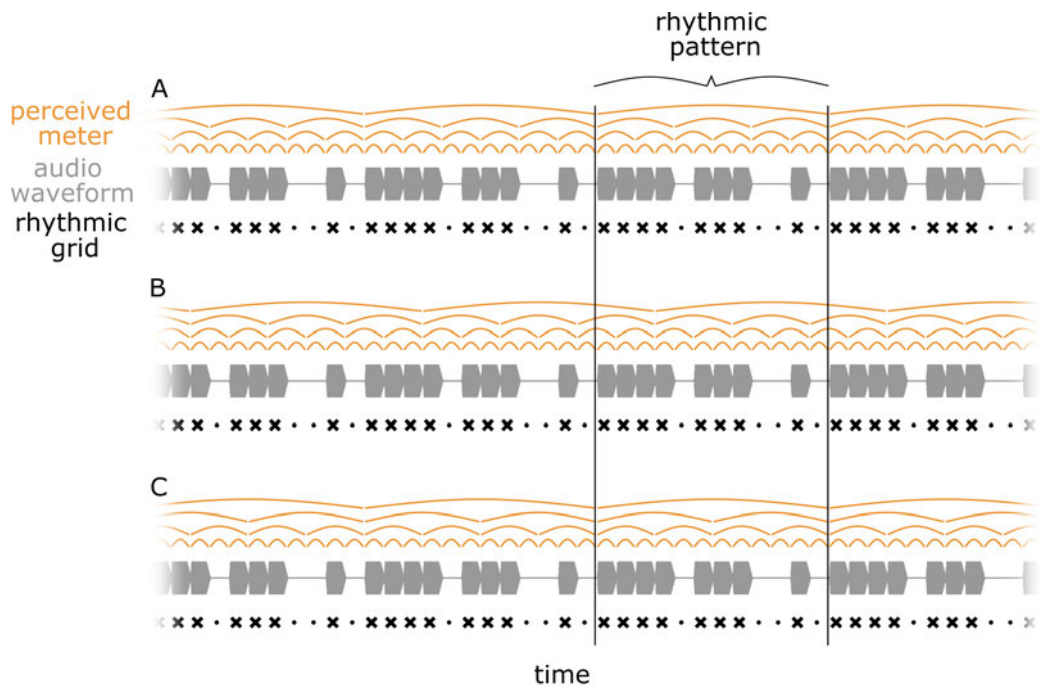


Figure 1.1. Representation of the rhythmic stimulus and perceived meter. The audio waveform of the rhythmic stimulus is shown in grey. The time intervals between tone onsets can be taken to represent a discrete sequence of time intervals. An abstracted representation of the sequence is depicted on an isochronous grid (in black), as a series of equally-spaced points, which can be either sound events (“x”) or silent events (“.”). Such “grid representation” can only be used when all intervals in the sequence can be expressed as integer multiples of a fast “grid” interval. Note that the rhythmic sequence consists of a seamless repetition of the same rhythmic pattern. Importantly, the same rhythmic sequence is shown here with three different perceived meters. Meter is shown as yellow arches that mark discrete time points, thus defining periodic pulses (starting from the slowest pulse on the top, and embedded faster pulses towards the bottom). **(A)** The perceived meter consists of pulses with periods of 1, 2, 4, and 12 grid points, and relations between these pulse periods can be coded recursively as {2,2,3}. The alignment of the metric structure with the starting point of the repeating pattern (i.e. meter phase) can be coded as 0, because the slowest pulse in the meter starts exactly with the first pattern event. **(B)** The depicted meter has identical periods {2,2,3} as in A, but is shifted to the right by 3 events (i.e. phase = +3). Thus, the pulses are aligned with different events in the pattern. **(C)** The meter has different periods, corresponding to 1, 3, 6, and 12 grid events, i.e. pulse periods with relations {3,2,2} and phase 0.

Theoretically, a large number of possible pulses could be perceived for a given rhythmic sequence. However, certain stimulus-meter combinations are unlikely to occur. One reason are physiological constraints that determine absolute limits of possible perceived pulse periods. For instance, the fastest possible metric pulse has been determined to be around 100 ms (see, e.g. London, 2004; Repp, 2005; Polak, 2018). Although a limit for the slowest pulse arguably exists, it is less sharply defined (for further discussion, see Repp and Su, 2013). In addition, a number of authors have proposed that there exists a preferred pulse period peaking around 600 ms and this preference monotonically decreases towards the absolute limits (Parncutt, 1994; van Noorden and Moelants, 1999; London 2004). Acknowledging the existence of such constraints is crucial to distinguish meter from other perceptual phenomena such as pitch perception. However, it is also important to note that a number of studies addressing these constraints have not clearly dissociated temporal limits of internally represented pulses that are used as a time reference (i.e. meter, as defined in the current thesis) from biomechanical limits of movement and from stimulus properties (e.g. van Noorden and Moelants, 1999; Parncutt, 1994). In other words, one should keep in mind that metric pulse with a faster (or slower) period than the tapped one can still be internally represented (see e.g. Repp 2008). Similarly, even though an auditory metronome with a very slow period does not induce internal representation of a metric pulse with the same period (e.g. when tapping 1:1 with the metronome, Repp and Doggett, 2007), this does not imply that such a slow metric pulse cannot be induced by an auditory input with more temporal information (such as in 1:N tapping with a metronome, Repp 2010a). These issues are further discussed in section 1.2.2.6.

Another source of bias towards perception of a particular meter may be the actual temporal structure of modulations (or intervals) in the rhythmic stimulus. To satisfy the time-locking constraint from the definition of meter, the pulse periods need to be related to the intervals within the rhythmic sequence. Even though this relation can be complex (Polak et al., 2016), here, the main focus is on a conventional case where the metric pulses and rhythmic intervals form integer ratios.

Moreover, the modulations in the stimulus may contain *periodic contrast at particular frequencies*. As discussed further in section 1.2.2, periodic contrast is present in a signal when a particular feature repeats at time points precisely separated by a stable interval

(period) and is different from features at other time points. A number of studies tried to link acoustic periodicities in the stimulus to the periods and phase of the perceived metric structure (Longuet-Higgins and Lee, 1984; Povel and Essens, 1985; Palmer and Krumhansl, 1990; Parncutt, 1994; McAuley and Semple, 1999; Eck, 2003; Toiviainen and Snyder, 2003; Hannon et al., 2004; Bouwer et al., 2018). These influential studies, along with ideas taken from research on interval timing (McAuley and Jones, 2003; Grube et al., 2010), have encouraged the (often implicit) assumption that meter is somehow an objective property of the stimulus that needs to be “extracted” by the brain (Kotz et al., 2018). However, the view of meter as a sound property is at odds with approaches emphasizing the subjective nature of meter (Agmon, 1990; Gabrielsson, 1993). Along this line, several authors have recently argued in favor of the subjective nature of meter, pointing out the flexibility in mapping between the rhythmic input and perceived meter across genres and cultures (London et al., 2017; van der Weij et al., 2017). For instance, stereotypical patterns from afro-cuban music (e.g. tresillo, clave, cascara) readily induce perception of a specific meter in listeners familiar with the genre, despite the lack of unambiguous acoustic cues. Similarly, the phase of the induced meter can be flexibly aligned with the prominent sound events for some genres (e.g. techno), but not others (e.g. swing, ska, reggae, mazurka, or even backbeat in rock music) (Merchant et al., 2015a).

These examples show that meter is arguably not an objective property of the acoustic stimulus, and is highly sensitive to top-down intention, learning and context (Penhune and Zatorre, 2019). While the acoustic structure of a rhythmic stimulus can constrain, bias, and even prevent induction of meter perception, the perceived meter cannot be determined solely on the basis of physical structure of the stimulus, rather, it is a higher-level perceptual phenomenon. Thus, the same rhythmic pattern can be mapped onto different metric pulses, while different patterns can be mapped onto the same set of pulses (Nozaradan et al., 2017a). The problems of approaches that assume one-to-one mapping between acoustic structure of the stimulus and perceived meter are further discussed in the following section.

1.1.2 Functions of musical meter

The definition of meter as a nested set of pulses that are time-locked onto the rhythmic input is sufficient to develop measures of meter in the brain and behavior (as discussed in

section 1.2.2). However, an important additional question concerns the function of these pulses (what do they serve for?). Based on prior literature, three basic functions of musical meter can be proposed. Firstly, perception of meter has a phenomenological component; the same rhythmic pattern simply sounds different when a different meter is perceived even though the physical stimulus is identical (Garner and Gottwald, 1968; Repp, 2007; Honing and Bouwer, 2018). The second function is related to time perception; timings of sound events are not only perceived with respect to each other, but with respect to the internally generated time points, defined by the perceived metric pulses (Povel and Essens, 1985; Large and Jones, 1999; McAuley and Jones, 2003). The third, and most important function is that meter perception drives movement (Toiviainen et al., 2010; Janata et al., 2012; Burger et al., 2014, 2018). This movement can be 1:1 with one of the pulses in the perceived meter (e.g. head bobbing, foot tapping). But perceived meter also provides temporal scaffolding for much more complex forms of movement, enabling precise temporal coordination between an individual and a rhythmic stimulus (dance) and between individuals (music performance) *even when the rhythmic patterns within the sensory input and behavioral output are not identical, nor periodic* (Keller and Repp, 2005; Repp and Su, 2013; Heggli et al., 2019). A simple example involves a string quartet, where each musician must produce a complex series of movements that rarely precisely repeat (i.e. are not strictly periodic). At the same time, these movements must be precisely coordinated in time with the acoustic input the musician receives, which comprises a complex (again, rarely strictly periodic) sound sequences of the whole ensemble.

1.1.2.1 Does meter perception rely on dynamic attending, expectation, entrainment models or predictive coding?

Over the last decades, a number of functional descriptions beyond those described above have been proposed as fundamental to meter perception, particularly dynamic attending and expectation (Jones and Boltz, 1989; Drake et al., 2000; Vuust et al., 2018). These have strongly influenced the neuroscientific methods and hypotheses developed to study the phenomenon (Cameron and Grahn, 2016; Honing and Bouwer, 2018). Contrary to these views, I argue that processes such as dynamic attention and expectation may not be fundamental to meter perception, even though they could be definitely related in certain

contexts (see e.g. van der Weij et al., 2017). Therefore, it is important to review the work that led to development and wide acceptance of these functional descriptions, and point out their weaknesses.

An important finding of early investigations was that perception and reproduction of a single time interval (i.e. defined by two successive sound events) is often not equally accurate, but depends on context. For instance, the same interval is represented more accurately if it is directly preceded by multiple repetitions of the same interval (Drake and Botte, 1993; Miller and Auley, 2005). On the other hand, accuracy decreases when intervals in the preceding context are variable (Zeni and Holmes, 2018), and the estimate is biased towards the mean of the recently encountered intervals (Jazayeri and Shadlen, 2010; Cicchini et al., 2012). To explain these phenomena, two competing models were proposed, both assuming that time is measured with respect to an internal clock (Barnes and Jones, 2000; Pashler, 2001; McAuley and Jones, 2003). For *interval-based timing* models, the internal clock would work like an hourglass, i.e. being able to re-start at any arbitrary time point (Keele et al., 1989; Ivry and Hazeltine, 1995). The time interval between two sensory events could then be compared to the reference interval stored in the timer. Because of sensory and internal noise, the reference interval could be considered a statistical estimate that can be biased, but also made more precise by integrating previously observed intervals (Drake and Botte, 1993; Jazayeri and Shadlen, 2010; Cicchini et al., 2012; van Rijn, 2016). On the other hand, *beat-based timing* models suggest that the internal timer works like an oscillator, with period and phase set by the context sequence of intervals (Schulze, 1978; Povel and Essens, 1985). The timing of events in the sensory input is then compared to the time points defined relative to the instantaneous phase of the oscillator. In fact, the two systems can be both conceptualized as oscillators, whereby timing is estimated according to the phase of the oscillator at which the sensory event arrives. The crucial distinction is that an interval timer resets its phase completely with each sensory event, but beat-based timer is assumed to keep oscillating without correcting the phase at all. This leads to dissociable predictions of the two models about effects of prior rhythmic context on perception of time intervals, which were tested in a number of behavioral studies that yielded mixed results (Schulze, 1978; Keele et al., 1989; McAuley and Kidd, 1998; Pashler, 2001). In an elegant study, McAuley and Jones (2003) fitted a continuous phase- and period-correction parameter directly to behavioral data, showing that neither pure interval nor beat-based

model can explain effect of isochronous context sequence on single interval perception. Rather, the most parsimonious internal clock was characterized by a partial phase and period correction (the so called “entrainment model”). Similar lack of clear dissociation in brain networks involved in interval and beat-based timing was found in neuropsychological studies (Cope et al., 2014) (but see Breska and Ivry, 2018 for recent evidence of dissociation between the two timing systems).

These studies have shown that using oscillators to model perception of single intervals and isochronous sequences can be powerful. However, it involves significant challenges when modeling time processing in non-isochronous rhythmic patterns (see Figure 1.2). This is because a single oscillator with linear phase-response function would be constantly desynchronized by sound events separated by intervals different than integer multiples of its internal period. One way to solve this issue is to use a rigid oscillator (i.e. beat-based model with no phase or period correction) (Povel and Essens, 1985). However, the phase and period of such oscillator need to be set a priori, after first assessing its alignment with the acoustic structure of the *whole* rhythmic sequence. Moreover, such model fails when slight timing fluctuations typical for human performance are considered (Large and Palmer, 2002). To avoid the a priori setting, a bank of oscillators (tuned to a range of frequencies) with amplitude dynamics and phase correction can be used. This way, only oscillators tuned to frequencies that yield high correlation with periodic contrasts given by modulations in the physical stimulus remain active (Large, 2000a; Tomic and Janata, 2008; Todd and Lee, 2015a). These models have been influential in the field, leading to the assumption that meter perception is directly driven by the modulations in the sensory input, and is only induced when these modulations create prominent contrast at meter-related periodicities (Povel and Essens, 1985; Jones and Boltz, 1989). However, as noted above, these models cannot account for a range of genres where acoustic modulations in the stimulus cue different meter parameters than are readily perceived by the listeners (e.g. reggae, ska), or where the modulations are ambiguous (e.g. tresillo, clave).

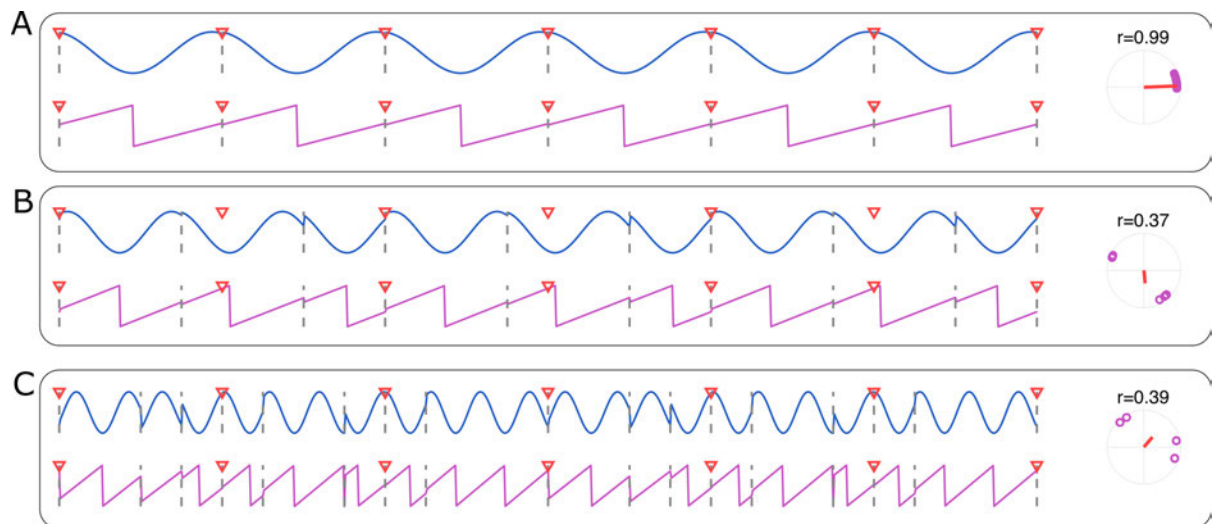


Figure 1.2. Issues of an adaptive oscillator when tracking pulse in a rhythmic pattern. Oscillator with linear phase (50%) and period (10%) correction was modeled based on Obleser et al. (2017). The phase of the oscillator over time is plotted in magenta, and waveform of the oscillator (cosine of the phase) is plotted in blue. Grey vertical dashed lines represent sound onset times. Red triangles depict positions of the putative metric pulse. The extracted phases of the oscillator across the pulse positions are plotted on the right. The phase consistency across pulse positions (mean vector length r) is shown on the top of the circular plot. In case of perfect tracking of the pulse, the phase should be identical across all pulse positions, i.e. $r=1$. The response to the first 4 pattern cycles were discarded to remove the transient response of the oscillator. **(A)** Isochronous rhythmic input. The oscillator precisely synchronizes with the pulse positions, which coincide one-to-one with the isochronous stream of sound onsets. **(B)** Tresillo rhythm, typical for afro-cuban music and often used in popular music (Cohn, 2016). The pulse typically perceived for this pattern does not systematically coincide with the sound onsets. Still, the onsets elicit phase and period correction of the oscillator, resulting in poor tracking of the perceived pulse positions. This shows that a single adaptive linear oscillator cannot faithfully track metric pulse in a complex rhythmic pattern. **(C)** “Complex” rhythm, used in a number of previous studies (Nozaradan et al., 2016b, 2018; Lenc et al., 2018). Assuming a pulse period and phase typically tapped by human participants shows that the oscillator does not faithfully synchronize to the perceived pulse.

A softer version of this view is that the perceived meter does not have to constantly align with the acoustic structure of the stimulus, but still needs to be first directly induced by this acoustic stimulus (Large and Palmer, 2002; Honing and Bouwer, 2018). In other words, once induced by a stimulus with regular acoustic structure, meter perception can persist over time even if the pattern of modulations in the stimulus temporarily changes towards cueing different meter periods (in musical terminology “hemiola”) or phase (in musical terminology “syncopation”). While this notion supports mechanistic oscillator models of meter perception and may hold in many cases, it is not a rule. In many genres the acoustic

modulations never clearly support the meter readily perceived by experienced listeners (London et al., 2017). Moreover, while models using a bank of oscillators with amplitude dynamics can be used to model selection of meter periods, the *phase-resetting issue remains*. A way around this is to use a nonlinear phase-response function (Large and Palmer, 2002). This way, the oscillator can “ignore” events that do not occur within a narrow range on its cycle, but still correct to small timing deviations in the input. Such assumption is closely related to models implementing dynamic attending theory (DAT). DAT proposes that limited processing resources can be selectively allocated in time, thus resulting in enhanced processing of sensory input at some time points, and suppressed processing at other time points (Jones and Boltz, 1989; Jones et al., 2002). They suggest that such “expectations” about upcoming stimulus timing are induced through mechanistic entrainment of internal oscillators to acoustic modulations in the sensory input (thus providing a link to oscillator models of timing). An allocation of resources to such anticipated time points would lead to better stimulus processing and adaptive behavior including enhanced auditory and visual sensitivity (Jones et al., 2002; Hurley et al., 2018; Bouwer et al., 2020b), better discrimination, memory encoding (Johndro et al., 2019; Hickey et al., 2020), and faster reaction times (Escoffier et al., 2010; Bolger et al., 2013, 2014) to stimuli aligned with the pulses induced by prominent periodic modulations in the stimulus sequences. While DAT was originally introduced as a more general theory, it has significantly influenced theories of meter perception (Fitzroy and Sanders, 2015; Honing and Bouwer, 2018; Bouwer et al., 2020b), to the point where some authors defined meter perception as “entrained attention” (London, 2004).

There are multiple additional reasons why ideas from DAT have been adopted by so many researchers of meter perception. Firstly, research on musical rhythm has been predominantly focused on a subset of genres (Western classical and popular music), whereby periods and phase of the perceived meter are closely related to the acoustic structure of rhythmic patterns (for further discussion, see London et al., 2017). In such environment, metric pulses represent a reliable source of expectations, i.e. they statistically predict when subsequent sound events will occur (Palmer and Krumhansl, 1990). At the same time, mechanistic oscillator models commonly used in time-perception and DAT studies fit nicely to such stimuli, thus encouraging generalization of findings from simple rhythmic stimuli to meter perception in complex rhythmic patterns. Moreover, a number of

recent neurophysiological studies suggested that neural networks responding to acoustic stimulation could be described as self-sustaining weakly coupled oscillators that can entrain (i.e. systematically align their phase) to temporal modulations of the acoustic input (Doelling et al., 2019; Lakatos et al., 2019; Poeppel and Assaneo, 2020). This provides direct mechanistic link to functional interpretations of neural oscillations known from low-level neurophysiology, typically involving excitability fluctuations and information transfer between brain regions (Draguhn and Buzsaki, 2004; Lakatos et al., 2005).

Thus, the picture emerging from the literature reviewed above could be summarized as follows: (i) if acoustic modulations in the stimulus contain prominent contrast at particular periodicities, this entrains neural oscillations such that their phases align to time points where acoustic modulations systematically (and periodically) occur. (ii) Because oscillations represent fluctuations in neural excitability, there should be enhanced processing of stimuli presented at these periodic time points, and suppressed processing of misaligned stimuli. (iii) Thus, the phases of neural oscillations can be thought of as embodying expectations of prominent acoustic events.

The question is whether such a model can be generalized to the domain of musical rhythm and meter perception. On one hand, ideas of expectation are ubiquitous in music theory of meter, as illustrated in the work of Lerdahl and Jackendoff (1983): "*The listener's cognitive task is to match the given pattern of phenomenal accent as closely as possible to a permissible pattern of metrical accentuation*" (p. 18). In their terminology, phenomenal accents represent modulation structure of the acoustic stimulus. *Metrical accents* are time points internally defined by the perceived metric pulses. The number of nested metric pulses that coincide at a particular time point determines (i) *how strongly a prominent sound event is expected* at that time point and also (ii) the amount of perceptual prominence of such event. On the other hand, most of these music-theoretic approaches do not capture the wealth of musical traditions and genres around the globe, as already suggested in the sections above (Polak et al., 2016; London et al., 2017). Indeed, applying ideas of DAT to swing, ska, reggae, chanson, mazurka, clave, tresillo, i.e. rhythmic stimuli where prominent sound events are typically not aligned with the perceived metric pulses, would mean that perception of most prominent sound *events would be largely suppressed*. Similarly, some empirical evidence about the distribution of resources across time points predicted by DAT entrainment models (Bolger et al., 2013, 2014; Kunert and Jongman,

2017) is not compatible with perceptual salience of different time points as defined by music-theoretic models of meter (Lerdahl and Jackendoff, 1983; Palmer and Krumhansl, 1990; Large, 2000a; Large and Palmer, 2002). Consequently, taking dynamic attention as fundamental to meter perception can lead to contradictions. For instance, London (2004) proposes a definition of meter as entrained selective attention, but at a different point emphasizes that “we do not ignore the non-emphasized events, we organize them” (p. 14). Hence, using the term “attention” when referring to meter may be more misleading than useful.

In musical contexts, rhythmic patterns with sound events aligned and misaligned from the perceived metric pulses are often cyclically repeated (Witek, 2017). Therefore if the behavioral goal is to align preferential processing to time points where sound events are expected, mechanistically relying on the perceived metric pulses would be statistically inefficient and disadvantageous. Instead, there is evidence that attention may be allocated in time based on memory of a repeating rhythmic pattern alone (Bouwer et al., 2020b; Schirmer et al., 2020). In light of these results, the current point is not to claim that meter perception cannot provide temporal reference for dynamic allocation of attention (see, e.g. Breska and Deouell, 2016; van der Weij et al., 2017). However, *there may not be a one-to-one relationship between the perceived meter and dynamic attentional fluctuations* (Hurley et al., 2018). Therefore, dynamic attention and expectation may not be fundamental processes defining meter perception, contrary to what has been previously proposed (London, 2004). Perhaps refraining from using these terms, as is done in this thesis, would lead to a more parsimonious theory and hypotheses about meter processing.

Similar reasoning can be used for conceptualization of meter in the framework of predictive coding. This framework has been previously used to explain how rhythmic stimuli can give rise to feeling of pleasure (Vuust and Witek, 2014; Vuust et al., 2014, 2018; Koelsch et al., 2019; Matthews et al., 2020), and why certain rhythmic inputs elicit spontaneous movement, or urge to move (Witek et al., 2014b; Witek, 2017). Originally proposed by Karl Friston as a general theory of brain function, the predictive coding theory relies on the idea that the brain predicts the causes of sensations based on a comparison between the actual sensory input and previous knowledge (Friston, 2005). Previous experience would thus be the basis for the continuous generation of internal predictive models to be compared to the sensory input, and these models would be gradually updated to minimize prediction error.

Yet, when applied to meter perception, the predictive coding theory may show very similar issues as those already discussed for DAT, internal oscillators and expectations. Particularly, if the goal is error minimization, it is not clear why the system would consistently generate incorrect predictions when stimulated for example with a repeated pattern that contains consistent deviations from the metric pulse template (London et al., 2017; Witek, 2017). Instead, successful prediction could simply be achieved through memorizing the pattern of inter-onset intervals (Bouwer et al., 2020b; Schirmer et al., 2020). Indeed, it has been recently shown that high predictive accuracy can be achieved by formalizing meter as a learned generative model of rhythm while allowing high expectation of events to be formed at arbitrary positions with respect to the metric pulses (van der Weij et al., 2017). In other words, as for DAT, *there may not be a one-to-one relationship between the perceived meter and high predictive accuracy*. Nevertheless, in order for the system to calculate the prediction error of an event with respect to an internal pulse-like metric template at all, the *pulses need to be internally represented in the first place*. While the question of how the internal representation of pulses is used for prediction is definitely a topic worth investigating, the aim of the current thesis is more fundamental, as it investigates how the metric pulses are internally represented. For these reasons, the terminology of predictive coding is not adopted when discussing meter perception in the rest of this thesis.

1.1.3 Primacy of one pulse (beat) in meter

Until this point, the current discussion has been based on the definition of meter as a *set* of perceived pulses. However, it is important to note that music-theoretic and psychological literature often takes a different approach, where it is assumed that one pulse is somehow special to the listener (Lerdahl and Jackendoff, 1983; Drake et al., 2000; Toiviainen and Snyder, 2003; London, 2004; Honing and Bouwer, 2018). This special pulse has been referred to as the *tactus*, *beat*, *referent* (or simply *pulse*), and should correspond to the pulse that the listener spontaneously synchronizes to when asked to move along with the rhythmic pattern (e.g. by tapping their finger, or foot). The listener is assumed to determine the pulse as the periodicity that is most prominently cued by the acoustic modulations in the stimulus (Jones and Boltz, 1989; Drake et al., 2000). At the same time, characteristics of

the listeners such as their age (Drake et al., 2000) or cultural experience (Drake and El Heni, 2003) can bias the choice of the beat.

The fact that many authors have assumed a special status of one particular periodicity may be a result of multiple factors. Firstly, while the widespread use of finger-tapping to investigate rhythm processing provides simple means to assess induced pulses, it inherently limits the participant to synchronize only with one pulse at a time (van Noorden and Moelants, 1999; Drake et al., 2000; Snyder and Krumhansl, 2001; McKinney and Moelants, 2006; Martens, 2011; Nozaradan et al., 2012; Large et al., 2015; Rajendran et al., 2020). In contrast, recent experiments capturing spontaneous movement of the whole body reveal that humans readily synchronize to multiple metric pulses simultaneously with distinct body parts (Toiviainen et al., 2010; Burger et al., 2013, 2014). Secondly, the focus on a single pulse also comes from psychology and neuroscience research, where a single oscillator is often used to model time perception (Povel and Essens, 1985; McAuley and Jones, 2003) or the response of a neural population (Doelling and Poeppel, 2015; Assaneo and Poeppel, 2018; Zoefel et al., 2018; Doelling et al., 2019; Lakatos et al., 2019; Zalta et al., 2020). Yet, it is not clear what the selection of one pulse as the beat means functionally, besides the fact that the listener spontaneously taps that pulse. In other words, the status of other pulses is often not specified, i.e. whether they are still perceived, and if so, how they differ from the beat (Cohn, 2014).

A related view is that perception of a single pulse is induced first, and other pulses in the meter are defined by mentally accenting every N th pulse, thus creating groups of stronger and weaker time points. This view is closely related to the concept of “metrical accent” discussed in the section above, which assumes that metric pulses which coincide with one another elicit stronger expectation of prominent sound events (Lerdahl and Jackendoff, 1983). Such a view is typical for music theory, mainly due to the use of a notational system developed in the 18th-century for music that was closely related to metered poetry (characterized by a stream of accented and unaccented syllables) (Cohn, 2015). This has led to the notion of meter as a “pattern of strong and weak beats”. This conceptualization may be driven by the fact that music conductors are restricted to conveying a single pulse (similar to finger tapping paradigms) and marking slower pulses through spatial accentuation of individual pulse elements. In the same way, a metronome can temporally deliver only a single pulse and slower metric pulses must be cued by changing a feature of

individual pulse sounds (e.g. pitch). The assumption of a single pulse with regular pattern of accents was readily adopted by cognitive scientists (Palmer and Krumhansl, 1990; Large and Snyder, 2009; Bouwer and Honing, 2015; Honing and Bouwer, 2018), allowing to make connections with metrical phonology where a stream of syllables is recursively grouped based on their accentuation (Cooper and Meyer, 1963; Fitch, 2013; Kotz et al., 2018). However, such a view is not parsimonious for multiple reasons. Firstly, if slower pulses in a meter emerge from accenting (grouping) the beat, the beat must be the fastest perceived pulse. This is not compatible with the definition of beat as the pulse people commonly tap, as the fastest pulse within a meter can be faster than biomechanical limits of finger tapping (Repp, 2005; Polak, 2018). Thus, an ad-hoc concept of “beat subdivision” must be used to retain the model, leading to qualitatively different nature of pulses in a meter, including (i) the beat, (ii) slower beat groupings, and (iii) faster beat subdivisions. However, to what extent such theoretical categorization is useful remains unclear. Because the slower pulses emerge as a byproduct of accentuation, when considering only one slower level, the listener can synchronize to strong (accented) as opposed to weak (unaccented) beat points to track the slower pulse. However, when additional hierarchical levels are present, it is not clear which points are weak and strong at the global level. Instead, there may be points with various amounts of accentuation. How to get from this “pattern of accents” to slower metric pulses is not straightforward. Instead, a reversed process may be more plausible, i.e. thinking of the accentuation of a time point as derived from the number of coinciding metric pulses (Cohn, 2020).

To sum up, the terminology used to describe meter has become convoluted with terms such as accent, beat, tactus, bar, subdivision, grouping, pattern, strong, weak, contributing to conflation with distinct music-theoretic and perceptual phenomena. For instance, using the term “grouping” (Fitch, 2013; Kotz et al., 2018) is not appropriate, as meter should be clearly separated from other forms of musical grouping (Lerdahl and Jackendoff, 1983). Similarly, the term “accent” may be problematic, as it rigidly links expectations to time points defined by metric pulses, which may not hold across musical genres and traditions (see section 1.1.2.1) (London et al., 2017; van der Weij et al., 2017). Note that this confusion was also present in my early work, by using the terminology of “grouping” and “subdivision”, and by separating beat and meter processing (see Study 2, section 3).

Instead, new analytical approaches treat all the constituent pulses in a meter as equivalent (Cohn, 2020). While the number of pulses within the perceived meter can vary (and perhaps in certain cases, only one pulse can be perceived), conceptualizing such difference as more quantitative (one, two, or N pulses) than qualitative (beat-grouping-subdivision dichotomy) constitutes a better analytical tool to represent and categorize meters. This way, the focus can stay on the most important property of meter: the relations amongst the perceived pulses (meter periods) and their relation to the rhythmic stimulus (meter phase), as shown in Figure 1.1. It may seem that some distinctions discussed above primarily concern terminology. For example, grouping/subdivision vs. independent pulse can refer to the same psychological phenomenon without clear functional implications. Yet, establishing a clear way to describe, represent and classify meters facilitates development of coherent theoretical frameworks that can be used to design experimental paradigms and analysis methods within psychology and neuroscience.

The following section first discusses paradigms and measures that have been previously used to assess meter processing in the brain and behavior, while pointing out their strengths and weaknesses. Subsequently, a new conceptualization of **meter as transformation** is introduced, showing its direct relationship with the fundamental definition of meter that has been developed so far. Then, implications of this new conceptualization of meter are discussed, especially in terms of development of new methods to appropriately capture the mapping from rhythmic sound input to perception and behavior.

1.2 Measuring the phenomenon

Typically, two kinds of measures have been utilized to measure processes relevant for meter perception in behavior and brain activity. (1) Indirect methods rely on measures that should give different results when meter with particular parameters (pulse periods and phase) is induced, compared to when meter is not induced or has different parameters than assumed. These measures predominantly rely on functional definition of meter as discussed in section 1.1.2 (i.e. “what does meter do?”). (2) Direct measures are capable of directly tracking the representation of metric pulses in brain or behavior. They are based on a fundamental definition of the perceptual phenomenon. It is important to note that the term “direct” does not imply that we can directly measure meter perception in a phenomenological sense. Perception is by definition subjective, and thus can only be measured indirectly, e.g. through correlates in behavior or brain activity.

1.2.1 Indirect measures

1.2.1.1 Ratings

The most straightforward way to assess meter perception would be to ask the participant: “Do you feel meter?”. Variations of this question have been used across studies, e.g. “How much did this sound have a beat?”, “How easy was it to feel a beat?”, or “How clear was the beat?” (Grahn and Rowe, 2009; Matthews et al., 2020). Others have instructed participants to answer this question based on a subjective prediction of the difficulty to move along with a single metric pulse (beat) when listening to the rhythmic input (Henry et al., 2017; Bouwer et al., 2018). However, it is not clear what exactly such ratings measure. Firstly, the concepts of beat and meter are often quite difficult to explain to musically naïve participants, and extensive practice is required. Secondly, one could argue that the way these questions are formulated invites participants to rate properties of the rhythmic input instead of internal representation of meter. Indeed, some studies directly encouraged such strategies during instructions (Bouwer et al., 2018). Claiming that these ratings provide a good measure of beat (meter) *perception* because they correlate with syncopation-score models (Henry et al., 2017) is not convincing. Indeed, to calculate a syncopation score by assessing how well the sound events align with the metric pulses, *the meter needs to be*

internally represented in the first place (Povel and Essens, 1985). A participant could easily learn to rate patterns that align less with her internal representation of meter as “having less beat”. Still, if asked to tap to these patterns, this participant may demonstrate robust internal representation of pulse (e.g. Tranchant et al., 2016 show some dissociation between ratings of “beat salience” and tapping performance). Importantly, rating measures do not reveal the periods and phase of the perceived meter, thus providing limited insights into the perceptual phenomenon and how it relates to input features, context and learning (Nozaradan et al., 2017a).

Another way to assess perception of a metric pulse is to overlay an explicit sound representation of a pulse over the rhythmic pattern (Hannon et al., 2004). Participants then rate how well the overlaid pulse fits the rhythmic pattern for different pulse periods (and possibly also phases). The pulse parameters with greatest fit ratings are assumed to reflect the meter induced by the pattern when it is presented alone. This approach is superior to the rating method described above in the sense that it is informative about the periods and phase of the perceived meter. Thus, with appropriate instruction to the participants, it could yield important insights into meter perception. Yet, overlaying the pulse as another sound layer significantly changes the acoustic input to the participant, thus it remains unclear to what extent such fit-ratings are informative about meter perception when the rhythmic pattern is presented alone (Lenc et al., 2020). A similar approach that avoids this drawback may be to first present a rhythmic pattern in an encoding phase and then to test recognition memory for that pattern across different explicit metric contexts (i.e. with overlaid sound representation of different pulses). This procedure assumes that performance will be best when the overlaid pulse in the recognition phase matches the meter that was originally perceived during the encoding phase (Keller, 1999). Still, both approaches described above do not provide a measure of meter perception in real time. Moreover, to have a comprehensive test of meter perception, all plausible combinations of pulse periods and phases should be exhaustively tested, which may lead to an overly long procedure.

Recently, this approach has been converted into a behavioral test of beat perception (BAT, beat alignment test), where an isochronous auditory pulse (click track) is superimposed over a musical excerpt (Iversen and Patel, 2008; Harrison and Müllensiefen, 2018). However, instead of examining the parameters of the perceived meter, the aim is to measure the *general ability to perceive beat*. Thus, the test adaptively measures how close the pulse has

to be to the assumed beat positions to be still rated as aligned to the beat. Thus, the test *does not measure spontaneous induction of meter perception*, and instead assumes that the “correct meter parameters” are known in advance (but see Parncutt, 1994; McKinney and Moelants, 2006 for evidence of interindividual variability in meter perception). Furthermore, it needs to be determined to what extent performance in BAT relies on other processes than meter perception, such as auditory stream segregation (Pressnitzer et al., 2011) and integrative attending (Keller, 1999; Keller and Burnham, 2005). Additionally, the possibility of alternative strategies such as relying on asynchronies between the superimposed metronome clicks and prominent on-beat sounds in music (instead of internally generated pulse percept) must be eliminated.

1.2.1.2 Other indirect behavioral measures

Another type of indirect behavioral measures is based on the assumption that meter perception leads to better encoding of time intervals making up the rhythmic pattern. These tasks include detection of temporal distortions in rhythmic sequences (Trehub and Hannon, 2009; Hannon et al., 2011), as well as discrimination, recognition, and reproduction of rhythmic patterns (Povel and Essens, 1985; Fitch and Rosenfeld, 2007; Grahn and Brett, 2007). Similar to rating measures, these paradigms do not aim to measure parameters of the perceived meter (specifically, its periods and phase), but whether or not *a* meter is induced by certain patterns (Povel and Essens, 1985; Grahn and Brett, 2007; Grube and Griffiths, 2009) or in certain individuals (Grube et al., 2010; Dalla Bella et al., 2017; Vuvan et al., 2018).

A similar measure used across a range of studies involves inserting a deviant sound event into a repeating rhythmic pattern after listening to it for a while (Hannon and Trehub, 2005a, 2005b; Hannon et al., 2012b). Assuming prominent acoustic cues to metric pulses, the event can be added so that the whole rhythmic sequence from that point onwards becomes misaligned with the induced meter. If meter is perceived, a detection of such deviant should be better than detecting an added event that does not cause subsequent misalignment with the induced meter. However, such paradigm may not work with patterns that do not contain prominent cues to a specific meter, as the misalignment would not

produce tension between internal metric pulses and the acoustic input (Parncutt, 1994; Repp, 2007; Repp et al., 2008).

As many processes contribute to performance on these indirect tasks (Tranchant and Vuvan, 2015), the paradigms rely on comparisons with baseline (or across groups) to isolate meter perception. For instance, simply observing that a temporal deviation in a particular rhythmic pattern can be detected does not say anything about meter perception on its own. Indeed, in some cases (e.g. Grahn and Brett, 2007; Grube and Griffiths, 2009), the fact that a particular rhythmic pattern is poorly reproduced or recognized does not directly imply that it did not induce meter perception. Instead, internal encoding of the pattern may be less precise simply because it does not closely align with the *induced* meter (Povel and Essens, 1985).

1.2.1.3 Functional magnetic resonance imaging (fMRI)

The main assumption of prior neuroimaging studies has been that there is a network of regions involved specifically in meter perception (Merchant et al., 2015a). Generally, to isolate a network involved in a perceptual process, a typical neuroimaging paradigm contrasts brain activity when the process is presumably involved to baseline activity when the process is presumably absent. Importantly, all unrelated internal processes, as well as low-level features of the sensory input should be constant between the two conditions. However, a baseline condition with a strict control of low-level features has been difficult to achieve in studies of meter perception.

For instance, there are a number of widely cited papers showing involvement of motor areas during meter perception without overt movement (Grahn, 2012; Cameron and Grahn, 2016). However, closer inspection reveals that many of these results are based on contrasting listening to regular rhythms with rest (Chen et al., 2008a) or with listening to physically distinct stimuli such as isochronous sequences and patterns comprising non-integer interval ratios (Bengtsson et al., 2009; Kung et al., 2013).

Perhaps the best controlled baseline stimulus up to date has been developed in the seminal study of Grahn and Brett (2007), where regular rhythmic patterns (prominent acoustic cues to meter periodicities) were contrasted with complex patterns (little acoustic cues to meter), which were created by re-shuffling the constituent inter-onset intervals. Still, these

two stimulus categories differ in their modulation spectra, with simple patterns having prominent energy at meter periodicities. Grahn and Brett showed increased responses in a “beat network” (typically including areas such as auditory cortices, basal ganglia, premotor cortex, and supplementary motor area) for regular rhythms in contrast to complex rhythms. However, their interpretation of these regions being selectively involved in beat perception is problematic, as beat (and meter) perception is often induced even for rhythms with little acoustic cues (see section 1.1) (Chapin et al., 2010; Witek et al., 2014b; Large et al., 2015). Yet, Grahn and Brett did not directly measure beat perception for their stimuli, e.g. in a tapping task. Moreover, they observed no difference between complex rhythms and completely scrambled rhythms (non-integer interval ratios) where meter perception is not induced. Yet, with somewhat similar stimuli, Matthews et al. (2020) showed discrepant results, i.e., larger activations across the same brain regions for complex than completely scrambled rhythms, while other studies did not report clear differences between regular and complex rhythms during listening (Chen et al., 2008a; Bengtsson et al., 2009; Kung et al., 2013). These inconsistencies may be related to slight differences in stimuli between these studies, but also the fact that many of them (e.g. Grahn and Brett, 2007; Chen et al., 2008a; Kung et al., 2013) presented very short rhythmic stimuli where meter perception may not have time to develop (we do not dance to 5-s long songs).

Overall, the main issue of fMRI studies may be directly related to the low temporal resolution of the blood-oxygen-level-dependent (BOLD) response and thus its poor ability to directly capture temporal contrasts at meter-relevant timescales, as discussed below in section 1.2.2. Instead, fMRI studies are inherently limited to indirect measures that struggle to directly link the recorded signal to the perceptual phenomenon. Therefore, temporally resolved methods such as intracerebral EEG (see e.g. Nozaradan et al., 2016a; Herff et al., 2020) may provide more reliable insights into behaviorally-relevant transformations of rhythmic input across brain structures, and thus the functional anatomy of meter processing in humans.

1.2.1.4 Mismatch negativity (MMN) studies

Another prominent line of research utilized the high temporal resolution of noninvasive electrophysiological methods (EEG or MEG) to isolate human brain responses to regularity

violations in rhythmic patterns (for a review, see Honing et al., 2014). The most commonly used paradigm would aim to isolate the mismatch negativity ERP component (MMN) that has been extensively used in cognitive neuroscience (Näätänen et al., 2007; Sussman et al., 2014). MMN is considered to reflect auditory change detection with respect to a regularity established by the preceding context. The component is typically isolated from a difference wave obtained by subtracting transient EEG response to a “standard” event from the EEG response to a “deviant” event (MMN typically peaks around 150-250 ms after event onset). Importantly, low-level physical features across the standard and deviant must be identical, thus making sure that any difference in the EEG response can be only explained by different contexts in which the two events occur (i.e. with respect to the established regularity). To study meter processing using MMN, participants would listen to a repeating rhythmic pattern and occasional changes would be introduced at different positions within the pattern (Geiser et al., 2009; Ladinig et al., 2009; Winkler et al., 2009; Geiser et al., 2010; Bouwer et al., 2014; Bouwer and Honing, 2015; Bouwer et al., 2016; Haumann et al., 2018). Assuming that (i) the violation is processed differently depending on its alignment with the perceived metric pulses, (ii) the period and phase of the perceived pulses are known, and (iii) the acoustic context across violations is equal, any difference in brain response to violation on vs. off the perceived metric pulse can be interpreted as evidence of pulse representation in the brain. Due to this list of assumptions, the method has several limitations. Firstly, it has been difficult to exclude confounds with low-level acoustic context (Vuust et al., 2005, 2009; Bouwer et al., 2014). Utilizing control sequences with repeated deviant stimuli typical for MMN studies (see e.g. Ladinig et al., 2009; Winkler et al., 2009) is also problematic as parameters of the induced meter can significantly differ for such stimuli (as already discussed in Honing et al., 2014). Another limitation arises because period and phase of the perceived meter may not be known, especially for stimuli providing less clear acoustic cues to the listener (Parncutt, 1994; McKinney and Moelants, 2006; Chemin et al., 2014; Cameron et al., 2015). Finally, many types of violation can be possibly introduced in the sound input (e.g. sound omission or addition, change in timbre or timing, increase or decrease of intensity), however, their relevance for meter perception is not clear. If meter is linked with expectations (predictions) as reviewed in section 1.1.2, then a change that increases acoustic salience (e.g. intensity increment) would elicit greater response when misaligned from the perceived metric pulses (and the reverse would be true for violations

that decrease acoustic salience, e.g. tone omission) (Bouwer et al., 2014; Vuust et al., 2018). On the other hand, DAT would predict greater response to any type of change when coinciding with the perceived metric pulses. Studies aiming to dissociate between these hypotheses did not provide clear results (Bouwer and Honing, 2015), which is not surprising given that the links between meter and expectations (or attention) are not straightforward (see section 1.1.2.1).

1.2.2 Direct measures

To develop a direct measure of meter as a psychological phenomenon, I start from the definition of meter as *a nested set of perceived pulses that are time-locked onto the temporal structure of the auditory input*. Essentially, a pulse can be defined as a **systematic contrast** in time. In other words, a pulse emerges *when (1) something consistently happens at periodically spaced time points, and (2) it is different from what happens otherwise*. Time-locking requires these periodically spaced time points to relate to the temporal intervals defined by the rhythmic pattern (typically to form integer ratios, but the issue is more complex, see e.g. Polak et al., 2016). I argue that once any measure of brain activity or behavior meets these criteria, it can be considered directly relevant to meter perception (even though it may not be a direct measure of meter perception in a phenomenological sense).

The definition of a pulse as a systematic contrast in time is directly related to the mathematical concept of periodicity (a tendency of a function to recur at regular intervals). In other words, to represent a pulse with a particular period, a signal must repeat itself at intervals corresponding to the pulse period. Nevertheless, explicitly addressing the different signal properties that contribute to its periodicity (i.e. decomposing the term “periodicity” as I do in the following sections) is important when assessing the validity of different methods that aim to quantify how well a particular set of pulses (i.e. meter) is represented in a signal.

1.2.2.1 Meter as transformation

Another important point for the direct measures is the conceptualization of meter perception as a *transformation*. That is, meter perception can be thought of as a mapping of the rhythmic sensory input onto an internal representation of pulses, which can drive behavior (Agmon, 1990; Parncutt, 1994; Nozaradan et al., 2017a). Crucially, various rhythmic inputs can give rise to internal representation of a meter with the same parameters, thus indicating a form of *perceptual categorization*. In other words, a set of physically different stimuli can elicit categorically equivalent types of behaviour (including perceptual experience). This form of many-to-one mapping can be illustrated with an example of ballroom dancing. In such scenarios, participants do not know in advance which dance will follow and need to determine this from the acoustic information in the musical stimulus. Particular dance styles are closely linked to meters with particular parameters (of course, the absolute tempo of the metric pulses, and certain timbres and rhythmic patterns specific to a particular style also play a role). For instance, waltz requires a meter where at least one pair of pulse periods has a {1,3} relationship. On the other hand, e.g. cha-cha requires a {1,2,2} relationship. Not only are the dancers capable of consistently mapping the acoustic stimulus onto the representation of meter so that they can all spontaneously start dancing the correct style. They are also able to map a range of different acoustic stimuli onto the same meter, e.g. dance waltz to Strauss' The Blue Danube, but also to Metallica's Nothing Else Matters. Another example of many-to-one mapping is when the listener perceives the same meter across different rhythmic patterns played by a musician at different time points during a solo performance (see e.g. Vulf, 2017 for a live performance where Joe Dart delivers a variety of rhythmic patterns that are mapped onto the same pulse by the audience, as expressed by their clapping).

At the same time, the transformation from a rhythmic input to the perceived meter is flexible, often leading to one-to-many mapping. In other words, acoustically identical rhythmic stimuli can induce perception of different meters within an individual due to context and intention (Repp, 2007; Repp et al., 2008; Iversen et al., 2009; Nozaradan et al., 2011), and across individuals due to prior experience (Parncutt, 1994; Phillips-Silver and Trainor, 2008; Chemin et al., 2014; Polak et al., 2018). For instance, the phase of the meter spontaneously induced by a particular rhythmic pattern may change after the listener has

been exposed to an auditory input where this specific pattern is paired with additional sound layers that prominently cue a metric phase different to the phase that would be induced by this pattern prior exposure (a phenomenon called “turning the beat around” in electronic dance music, see Butler, 2006).

To summarize, the examples above emphasize three main points: (1) the perceived meter is induced by rhythmic patterns without one-to-one correspondence between acoustic modulations and perceived metric pulses (transformation), (2) a range of acoustically different inputs can elicit perception of the same meter (categorization, many-to-one mapping), and (3) identical acoustic stimulus can elicit perception of different meters (one-to-many mapping). To gain understanding of the transformation fundamental to meter perception, one needs to use direct measures to estimate meter representation in (i) the acoustic input, (ii) the elicited behavior, typically movement, and (iii) the brain processes that take place in between the sensory input and behavior. In other words, when comparing sound input, brain activity, and behavior, a *selective enhancement of contrast defining the perceived metric pulses* is expected.

It should be noted that the terms “transformation” and “mapping” are often used as synonyms in the current thesis. Yet, the term “mapping” implies a complete process, where a particular acoustic input induces internal representation of particular metric pulses. On the other hand, the term “transformation” emphasizes (i) the specific nature of the mapping, i.e. selective enhancement of metric periodicities, and (ii) that the process of mapping may be gradual, particularly when studying brain responses elicited by acoustic inputs.

1.2.2.2 The need for a better method to capture periodic contrast

The definition of a pulse as a systematic temporal contrast can be understood in terms of **generalization** (some feature of a signal takes on similar values at time points separated by a stable period), **differentiation** (these values are different from values of this feature at other time points), and **time locking** (the period with which the feature change repeats is precisely linked to the rhythmic input). A number of methods have been proposed to

measure the properties relevant for a contrast at meter periodicities within signals representing sound input, brain activity, and behavior. However, many of these methods solely focus on one signal property and abstract from others (e.g. focusing on time locking but neglecting differentiation, see section 1.2.2.6). Furthermore, many methods are too specific to be generalized across signals (e.g. syncopation scores developed for abstract representation of sound, see section 1.2.2.5).

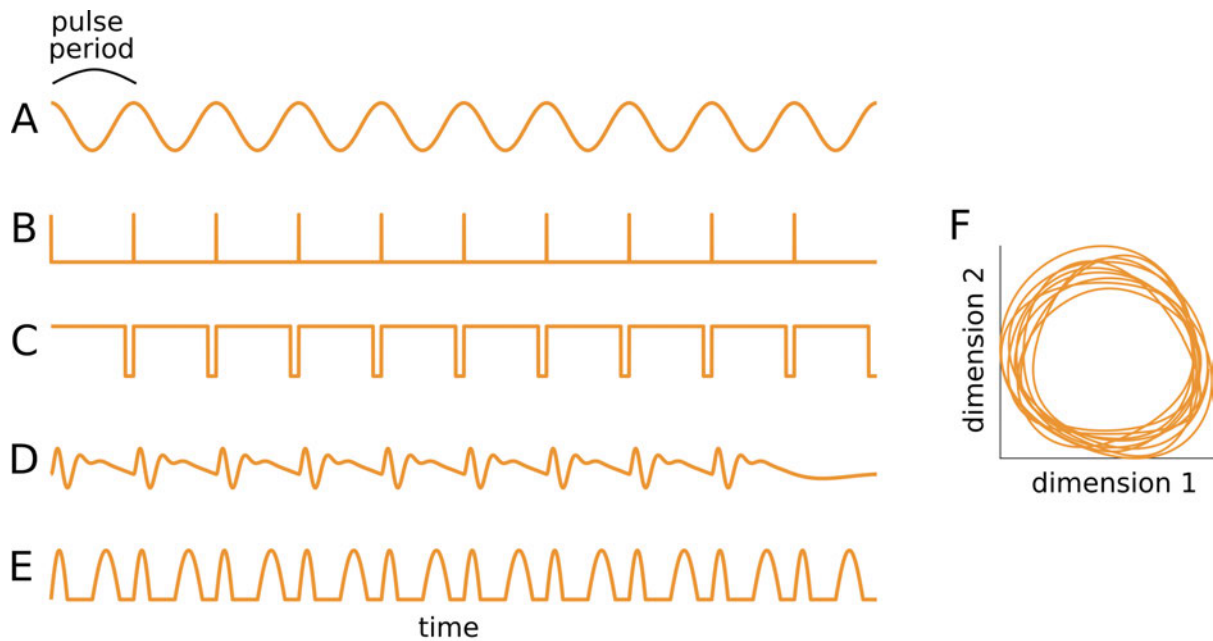


Figure 1.3. Multiple realizability of pulse representation. The example signals shown in the different panels all represent a pulse with the same period according to the criteria of generalization, differentiation, and time-locking. **(A)** The cosine wave completes exactly one cycle every pulse period. **(B)** The impulse happens exactly separated by the pulse period and not otherwise. **(C)** The square wave switches from low to high state precisely at times separated by the pulse period and not otherwise. **(D)** The complex waveform repeats itself precisely every pulse period, but not otherwise. **(E)** A narrower bump in the waveform repeats exactly separated by pulse period and not otherwise. **(F)** An example system repeats its trajectory in the state space precisely every pulse period and not otherwise.

A large family of methods that have been used to directly measure meter processing is based on time-domain representation of signals (Longuet-Higgins and Lee, 1984; Povel and Essens, 1985; Snyder and Large, 2005; Iversen et al., 2009; Fujioka et al., 2010, 2012; Schaefer et al., 2011; Witek et al., 2014b; Fitzroy and Sanders, 2015; Rajendran et al., 2017, 2020). Here, I argue that these methods may entail important limitations for directly measuring meter processing in continuous signals such as brain or behavioral signals. To

illustrate these limitations, I take a recent influential study where the authors analyzed spiking rates of neurons in the subcortical auditory nuclei of ferrets as the animals were stimulated with rhythmic patterns (Rajendran et al., 2017). The patterns were constructed by arranging a set of sound events (short noise snippets) on a fast isochronous grid of time points. They employed an intuitive method to measure contrast in the neural recordings by measuring mean spiking rate within short windows aligned to each grid point in the rhythmic pattern. The idea was that larger spiking rate consistently appearing every N th window (relative to the rest of the windows) would indicate greater contrast at a pulse with period N . While intuitive, this method illustrates several shortcomings of time-domain approaches (see Figure 1.4 for some examples using EEG signals). Firstly, it assumes that larger value of a signal feature (here firing rate) must occur at some specific latency with respect to the time points defined by the assumed pulse positions. This is not in line with the definition of a contrast that only requires a systematic *change* in a feature (it may as well be a consistently occurring “dip” in otherwise high firing rate). Similar assumptions have been taken for noninvasive studies of brain activity (Snyder and Large, 2005; Fitzroy and Sanders, 2015), even when “higher” or “lower” value of a feature has little direct physiological meaning, as can be the case with negative or positive values measured from negative or positive ERP deflections (Luck, 2014). Even when contrast is correctly measured as a relative difference (see e.g. Fujioka et al., 2010, 2012), there is an inherent tradeoff when setting the duration of the analysis windows. On one hand, the windows must be short enough such that the analysis is sensitive to transient changes and allows measuring stability of fine temporal locking onto the pulse period over time. On the other hand, the windows should be long enough to capture slower dynamics (note this is also the case for sounds, where slow modulations may be present). Furthermore, to measure the feature value within a window, one needs to establish a baseline, which is not straightforward for many types of relevant signals (particularly noninvasive brain recordings, see Figure 1.4).

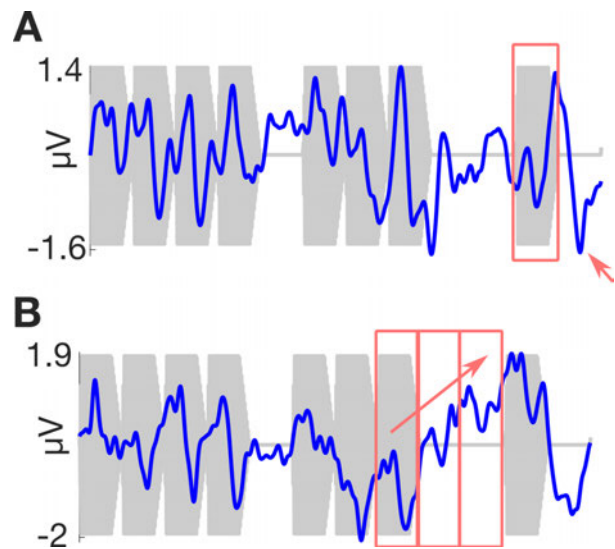


Figure 1.4. Difficulties of time-domain approaches to measure contrast. EEG response of two example participants taken from a cluster of fronto-central channels (shown in blue). The response was captured as participants listened to a repeating rhythmic pattern (acoustic waveform of one pattern repetition is shown in grey). The EEG was averaged over 21 pattern repetitions and 10 trials. The signal in panel **(A)** illustrates multiple disadvantages of window-based analysis. An example analysis window aligned with one 0.2-s long sound event in the pattern is shown in red. Because the signal variations go above and below zero and do not return to baseline within the window, it is difficult to obtain a meaningful feature that could be used to define the contrast across different windows. A different possibility would be to measure the peak-to-peak variation within the window. However, the value of this feature would be much higher if the following negative peak (marked with the red arrow) was captured within the same window. Similar arguments stemming from the arbitrariness of window duration hold for measuring root-mean-square amplitude etc. **(B)** A slow signal variation (marked by the red arrow) is not captured when the signal is analyzed with small successive windows if only relative signal change within each window is considered.

To overcome the limitations of time-domain methods, a new approach has been recently developed based on the frequency-domain representation of signals. I propose that this frequency-tagging approach offers a robust tool to capture all signal properties relevant for measuring pulse as a systematic contrast in time (i.e. generalization, differentiation, time-locking). Moreover, because this approach can be used on a range of signals, including acoustic input, brain, and movement, it has a potential to provide important insights into the transformations fundamental for meter perception. In the next section, I introduce the frequency-tagging approach applied to a general form of signal. Along the way, I touch upon

specific applications of the approach on signals representing sound, brain activity, and movement. I highlight similarities and differences between frequency-tagging and other methods that have been widely used in the field to measure contrast at meter periodicities.

1.2.2.3 Frequency-tagging approach to measure contrast at meter periodicities

Frequency-tagging is a method originally developed to measure periodic contrast in brain activity (Regan, 1989; Norcia et al., 2015; Gao et al., 2018; Rossion et al., 2020). Here, the method is introduced with a general type of signal representing changes in a variable over time. This variable can represent sound input, brain activity, or movement data, and the specifics of each will be discussed where appropriate. The assumptions of the approach as well as its sensitivity to signal properties relevant for meter will be explored using simulations.

Frequency-tagging is based on the fact that a contrast repeated systematically with a particular period can be identified in the frequency domain as narrow-band peaks of energy centred at the frequencies directly related to the contrast period (Nozaradan, 2014; Nozaradan et al., 2017a).

1.2.2.4 Contrast affected by differentiation of events

Let's assume that the shape of a signal is defined by a unitary waveform (hereafter “kernel”) that is periodically repeated every x seconds. Because the kernel represents a change of signal, and consistently appears at time points separated exactly by x seconds, and not otherwise, the resulting signal is a basic form of contrast defining a pulse with period x . Analyzing this signal with a Discrete Fourier Transform (DFT), or equivalently a Fast Fourier Transform implementation (FFT) yields a magnitude spectrum with distinct peaks centered at the frequency $1/x$ and harmonics. This contrast will be referred to as the “base”, and its prominence can be quantified by summing the FFT magnitudes across the harmonics. If every N th kernel is systematically different in some way, another periodic contrast is created in the signal (hereafter “target”). An example difference used in Figure 1.5 is a simple multiplicative gain of the kernel. The magnitude of this target contrast can be measured at frequencies corresponding to the period at which the target contrast happens

$1/(N \cdot x)$ and harmonics (again by summing the FFT magnitudes). Increasing the systematic difference between the base kernel and target leads to a greater contrast at the target period, as shown in Figure 1.5E. There are different ways beyond changing kernel amplitude to create the target contrast. For instance, the contrast can be driven by systematically different shape of the kernel, as shown in Figure 1.6.

Before summing the harmonics, it is important to apply a baseline correction, otherwise the measure will be biased. This is because magnitudes only take positive values, and therefore, even the smallest amount of noise (unavoidable in real-world signals) will lead to increased magnitudes across a broad range of frequencies due to spectral leakage and aliasing (even if the analyzed segment has a duration equal to an integer number of cycles for sine waves at the frequencies of interest). Figure 1.7 shows a simple example, where white noise with very small amplitude was added to the signal. Without applying the baseline correction, the summed magnitudes at target frequency and harmonics were positively biased, leading to a wrong conclusion that there was a significant response to the target contrast. The same situation could arise even when testing whether magnitude at a single frequency (i.e. without summing across harmonics) is significantly above zero across participants in an experiment. One commonly used way to prevent this is to subtract the mean magnitude at the neighboring frequency bins on both sides from each frequency bin in the spectrum (e.g. Xu et al., 2017). As shown in Figure 1.7, this procedure effectively centers the distribution around 0 in case there is no contrast at the target periodicity in the signal (i.e. no peak of energy at the frequencies of interest). Note that there are different alternatives to the subtraction method, for instance taking a ratio (e.g. Xiang et al., 2010; Riecke et al., 2014) or z-score (e.g. Lochy et al., 2018) of each frequency bin with respect to the neighboring bins. All these different methods are often used in a complementary way on the same set of data, as they offer distinct advantages (Jacques et al., 2016; Jonas et al., 2016). For instance z-scoring is convenient to assess whether the response at a particular frequency is significantly above the noise level. On the other hand, using subtraction yields values in the original units (e.g. μV in case of EEG signals), which can be easily combined across harmonics to obtain an estimate of the overall raw magnitude of the periodic contrast (by summing across harmonics) or to quantify the relative prominence of particular harmonics relative to others in the spectra. Because the overall magnitude and relative prominence are the main estimates of interest in the current thesis, the subtraction method will be primarily

used for baseline correction. Two important things to keep in mind when performing the baseline correction are: (1) the bins right next to the frequency of interest should not be included in the baseline in case some spectral leakage occurs, and (2) the furthest bin included in the baseline should be closer than the closest neighboring frequency of interest (i.e. harmonic where a magnitude peak is expected).

Importantly, the magnitudes at harmonics of the base contrast that overlap with the target are driven by both base and target contrasts, therefore cannot contribute to their differentiation (and must be excluded from the sum). However, if one is interested in the proportional contrast at the target period, the base magnitude can be used as a normalization factor (see Figure 1.5F). This is highly relevant, as sensory contrast is typically perceived proportionally (Halpern and Darwin, 1982; Gescheider, 1997), and normalization is a fundamental principle in the brain (Dean et al., 2005; Robinson and McAlpine, 2009; Rabinowitz et al., 2011; Lohse et al., 2020).

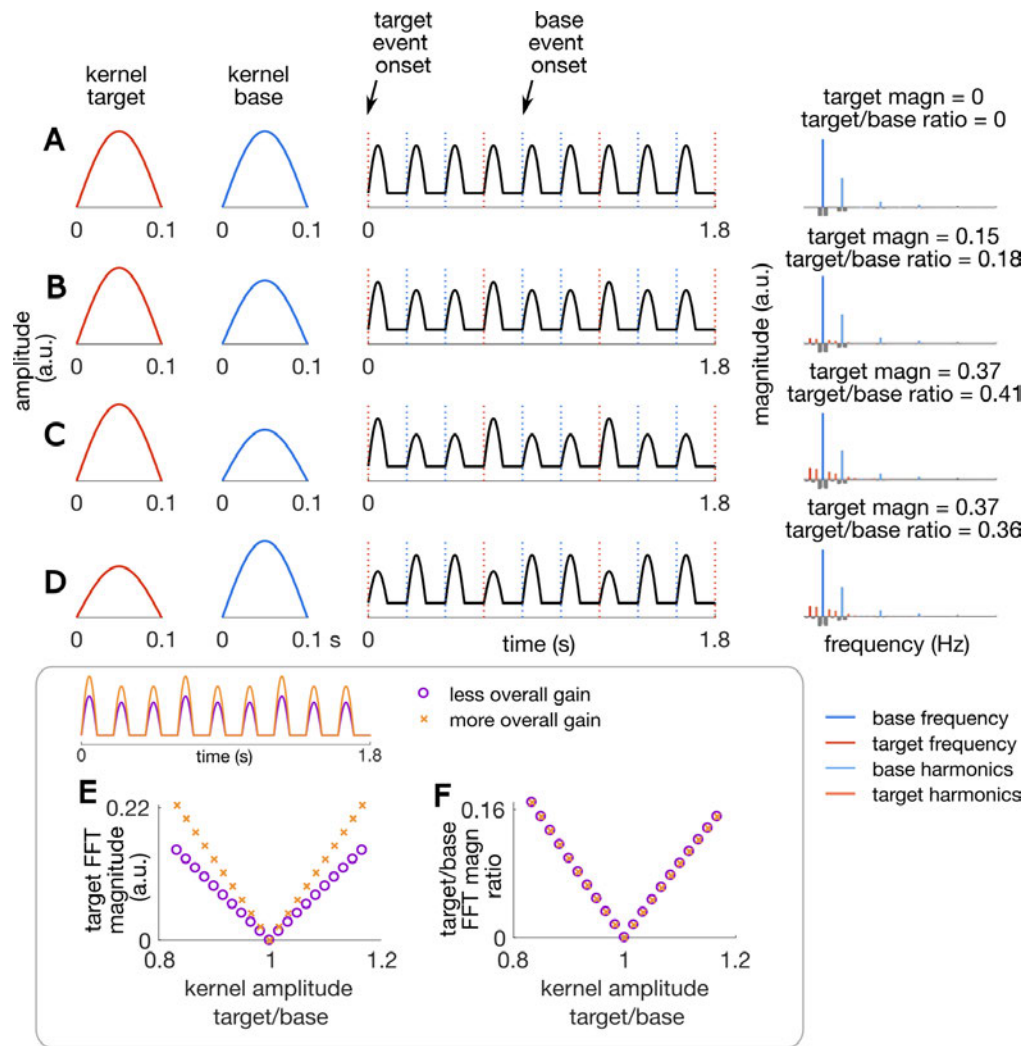


Figure 1.5. Using frequency-tagging to measure contrast in an isochronous sequence of events. The sequence (in black) is constructed as a series of isochronous events where every 3rd event is defined by the target kernel waveform (in red), and the rest of the events by the base kernel waveform (in blue) as shown on the left. Example target and base event onsets are marked by arrows on the top. FFT of the sequence is plotted on the right, with frequencies related to the base contrast shown in blue, and frequencies related to the target contrast shown in red. The summed magnitude at target frequency and harmonics (“target magn”) is plotted on the top of each spectrum, as well as the ratio of this value and the summed magnitudes at base-related frequencies (“target/base ratio”). **(A)** If there is no difference between target and base kernel, the signal only contains contrast at the base period. **(B)** If the amplitude of the base kernel is decreased, a contrast at the target period emerges. **(C)** If the difference between the target and base kernel increases so does the target contrast measured with frequency-tagging. **(D)** Contrast is always relative, thus decreasing the amplitude of the target kernel instead of the base kernel leads to the same result. **(E)** Systematically increasing the ratio between target and base kernel amplitude leads to proportional increase in the contrast measured as summed magnitudes at target-related frequencies. However, this measure is sensitive to the overall gain of the sequence (multiplying the whole signal with a constant, see example in the top part of the panel). **(F)** Dividing

the summed magnitudes at target-related frequencies by the summed magnitudes at base-related frequencies gives a relative measure invariant to the overall gain.

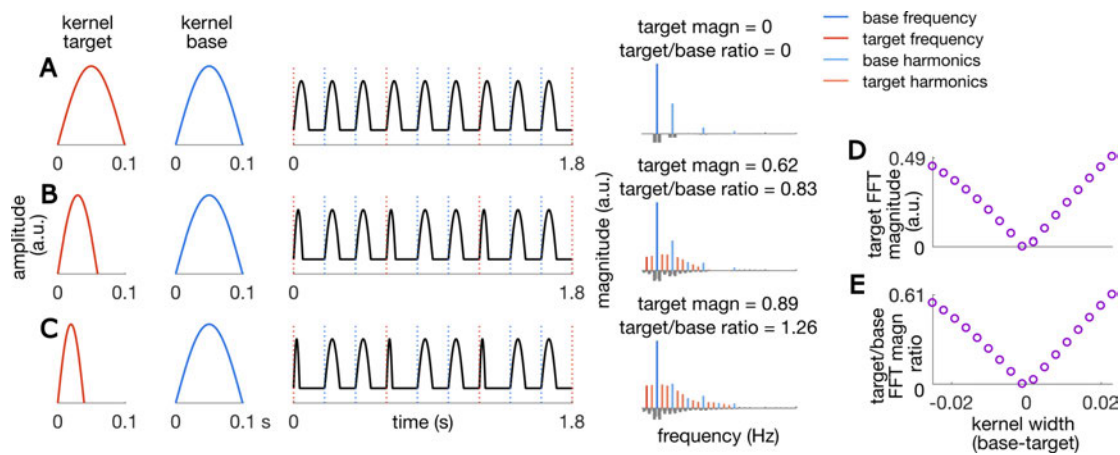


Figure 1.6. Contrast created by differences in kernel shape. Same as Figure 1.5, but instead of using kernel amplitude to create the contrast, kernel shape (width) is manipulated (A, B, C). The method is sensitive to the relative difference between the target and base kernel, yielding a monotonically increasing value as the difference in width increases (D, E).

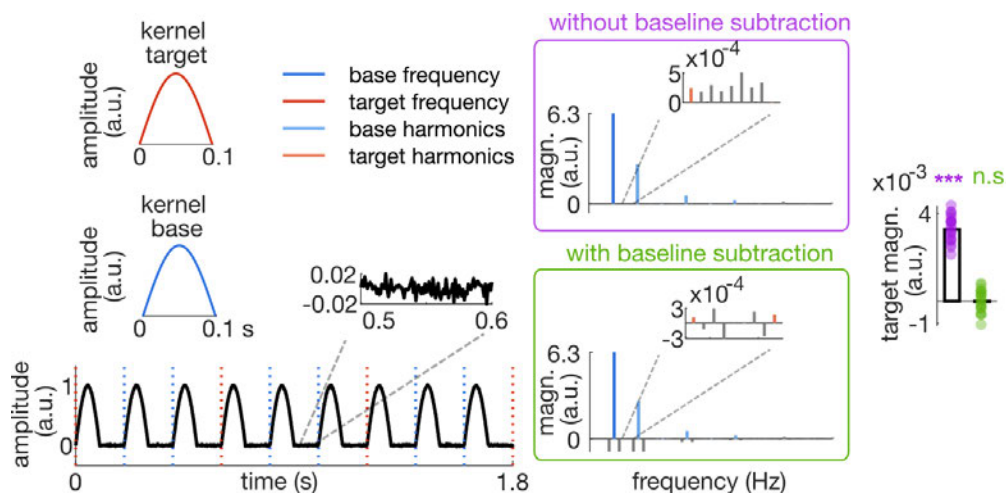


Figure 1.7. Illustration of the baseline subtraction method. Signal simulated in the same way as for Figure 1.5, except white noise was added in the time domain (see inset zooming onto the time domain signal, bottom left). Spectra of the signal are plotted with (green) or without (purple) baseline correction. The correction

involved subtracting the average magnitude at bins 2 to 5 taken from both sides of each frequency bin. Note that this subtraction results in negative values at bins near prominent peaks in the spectra. The signal was generated 20 times and the magnitudes at target frequencies were summed. The distribution of the summed magnitudes is shown on the right, indicating that the values are positively biased when baseline correction is not applied (***) significantly above 0), whereas applying the baseline correction centers the summed magnitudes around zero (n.s. not significant).

However, musical rhythms go beyond isochronous series of onsets. To demonstrate the utility of the method, one can use a repeating non-isochronous rhythmic pattern with events constructed on a fast isochronous grid, as shown in Figure 1.8. The simulation assumes a metric pulse plausible for this pattern and shows that the contrast between pulse positions and the other positions in the pattern can be directly measured by frequency-tagging. Moreover, normalizing by magnitude at frequencies unrelated to the target contrast makes the measure robust to overall multiplicative gain of the whole signal (see Figure 1.8D). Here, these “base” frequencies are chosen differently to a situation with isochronous series of base and target events as described above. Instead, one can use the fact that the spectrum of a repeated non-isochronous rhythmic pattern contains peaks at frequencies corresponding to the pattern repetition rate and harmonics. From this set of frequencies, a subset will correspond to the pulse rate and harmonics (target frequencies) and the rest is unrelated to the pulse (base frequencies). The relatively larger magnitude of target frequencies compared to base frequencies (calculated e.g. using z-scores as described in Figure 1.8) provides a robust relative measure of contrast at the pulse periodicity in the signal.

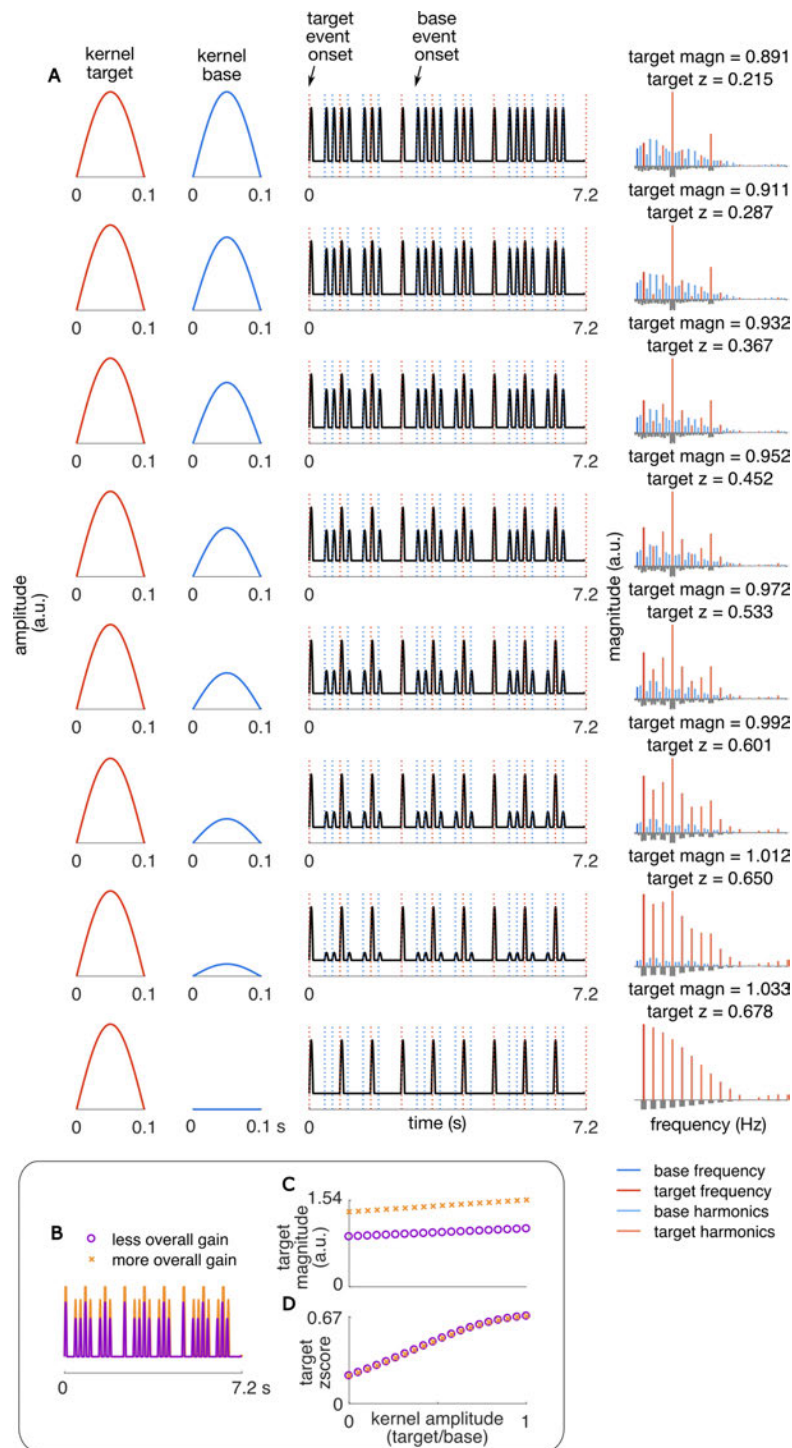


Figure 1.8. Contrast at one pulse periodicity in a rhythmic sequence consisting of a repeating rhythmic pattern. **(A)** Each row represents a sequence with gradually increasing relative contrast at pulse periodicity (defined by target events). The contrast is enhanced by decreasing the amplitude of the base kernel. The contrast can be measured by summing the magnitudes at target frequency and harmonics (“target magn”), however, this measure is sensitive to the overall multiplicative gain of the sequence **(B, C)**. A measure robust to the overall gain is achieved by first converting magnitudes to z-scores across all frequencies (target- and base-related)

using the formula $[(\text{magnitude at the frequency}) - (\text{mean magnitude across all frequencies})] / [\text{SD of magnitudes across all frequencies}]$, and calculating the mean z-score across target-related frequencies ("target z") (D).

Figure 1.8 provides a straightforward illustration of how meter contrast can be created by changing the amplitude of events aligned vs. misaligned with the pulse positions. However, in the context of musical rhythms, a rhythmic pattern does not have to contain a sound event at all perceived pulse positions (see section 1.1). Moreover, periodic contrast can be created not only by amplitude differences between events, but also by the arrangement of identical events in time alone. In addition, there is often more than one perceived pulse, and instead a meter of multiple nested pulses is perceived.

1.2.2.5 Meter contrast affected by arrangement of events

A number of theoretical measures have been developed to quantify how temporal arrangement of events generates contrast at metric periodicities. These measures typically work with discrete representations that define rhythmic patterns by assigning a fast isochronous grid with sound events (see section 1.1 for a definition of sound event and grid representation). The grid positions without sound events are often referred to as silent events. While originally developed for sound, these methods could be used for any type of signal that can be represented as a series of identical discrete events arranged on an isochronous grid.

A widely used method developed by Povel and Essens (PE, 1985) is summarized in Figure 1.9B. This method assumes a single pulse with particular period and phase, and calculates a score based on how many silent and unaccented sound events coincide with the pulse positions. Because a sound event is considered unaccented when surrounded by other sound events (or first within a group of two sound event, see Povel and Okkerman, 1981), the score will be smallest when prominent contrast is present at pulse positions (i.e. when "accented" sound events are present on-the-pulse and not otherwise). Note that variants of this method were proposed where the algorithm searches for a contrast even more explicitly (e.g. the mixed model in McAuley and Semple 1999).

A similarly popular algorithm from Longuet-Higgins and Lee (LHL, 1984) takes into account multiple metric pulses with particular periods and phase (Figure 1.9C). Positions where the pulses coincide determine weights for each grid point. Higher weights occurring when multiple pulses coincide are based on the assumption that contrast at these points is more important, as it contributes to multiple metric periodicities simultaneously. If the pattern contains a contrast that does not align with these weights (i.e. with the assumed set of metric pulses), the resulting score is high (similar to PE). Both, PE and LHL require assumptions about periods and phases of the perceived meter. However, if one aims to find the amount of contrast present in a repeating rhythmic pattern for a particular pulse period, the phase assumption can be removed by taking the score for each circular rotation of the pattern, and taking the minimum across all calculated values (see Figure 1.9B and C). Calculating the minimum score across several phase shifts may also work for nonrepeating longer rhythmic sequences if a small number of shifts is considered (Lenc et al., 2020).

Another way to assess contrast at multiple periodicities could be to define a vector with larger values corresponding to positions where multiple metric pulses coincide and cross-correlate this vector with the rhythmic sequence represented as a vector of ones (sound events) and zeros (silent events). Taking the highest correlation across lags quantifies the contrast (defined by the metric pulses) in the rhythmic sequence. This approach is shown in Figure 1.9D.

Finally, approaches based on the frequency-domain representation have been developed by multiple authors to analyze rhythmic patterns in musical contexts (Amiot, 2016; Chiu, 2018; Milne and Herff, 2020). Typically, a rhythmic pattern is represented as a vector of ones and zeros (same as for the cross-correlation method above), and transformed into the frequency domain using DFT. The DFT can be understood as taking a dot product of the sequence with a set of basis vectors that represent complex sinusoids with periods such that they complete $\{0, 1, 2, \dots, N-1\}$ cycles within the span of the sequence (where N is the number of events in the sequence). Each dot product yields a complex coefficient, and its absolute value can be interpreted as spectral magnitude at the frequency defined by the sinusoid. Thus, magnitude represents the fit between the complex sinusoid and the rhythmic sequence. To search for contrast at particular pulse periods, one must measure the magnitudes for sinusoids with the number of cycles corresponding to the pulse period and its integer

multiples (i.e. harmonics). Focusing only on a single sinusoid would bias the measure, as the contrast can take different shape than a pure sine wave.

Figure 1.9A shows an example where two metric pulses are considered, with periods corresponding to 2 and 4 events. The frequencies and harmonics corresponding to these two periods can be labeled as meter-related frequencies and in this case comprise the two frequency bins highlighted in red. Note that the second harmonic of the slower pulse overlaps with the first harmonic of the faster pulse. The rest of the valid frequencies in the DFT can be considered meter-unrelated. It is important to point out that the magnitudes at frequencies higher than half of the individual event rate cannot be measured in the current representation of the rhythmic signal due to aliasing. This is because the resolution of the vector representation is limited to one sample per event. To obtain a normalized measure, the magnitude at each frequency is first converted into a z-score using the formula: $[(\text{magnitude at the frequency}) - (\text{mean magnitude across all frequencies})] / [\text{SD of magnitudes across all frequencies}]$. Mean z-score at meter-related frequencies quantifies a relative contrast in the sequence (already discussed in section 1.2.2.4, and further advantages of this measure are demonstrated in Figure 1.17, section 1.2.2.6).

To assess whether the DFT method provides similar measurement of the contrast at metric periodicities compared to time domain approaches such as PE, LHL and cross-correlation with a template, a simple simulation was performed in which a set of 160 rhythmic patterns was created by assigning a 12-point isochronous grid with all possible permutations of 6, 7, and 8 sound events (and allowing maximum 4 successive silent events). Minimum LHL and PE syncopation scores across 12 phase shifts of each pattern were calculated to estimate contrast stemming solely from the arrangement of events in time. The goal was to measure the prominence of contrast locked onto metric pulses with periods 2 and 4 grid points. Therefore, metric pulses with periods 2, 4, and 12 grid points were assumed for LHL and a pulse with period 4 grid points was assumed for PE (recall that PE can only measure one pulse at a time). When tested across this pool of rhythmic patterns, all three time-domain approaches (LHL, PE, cross-correlation) give highly similar results to the DFT method, as shown in Figure 1.9. This indicates that the frequency-domain approach is sensitive to contrasts in rhythmic patterns that emerge solely from the arrangement of identical events in time. This DFT-based approach can therefore be considered a generalization of frequency-tagging to measure contrast in rhythmic sequences that are abstractly

represented on a fast isochronous grid. Because the original frequency-tagging terminology of target vs. base events is not quite appropriate for non-isochronous rhythmic signals in the context of meter perception, the terms “meter-related” and “meter-unrelated” frequencies are used hereafter.

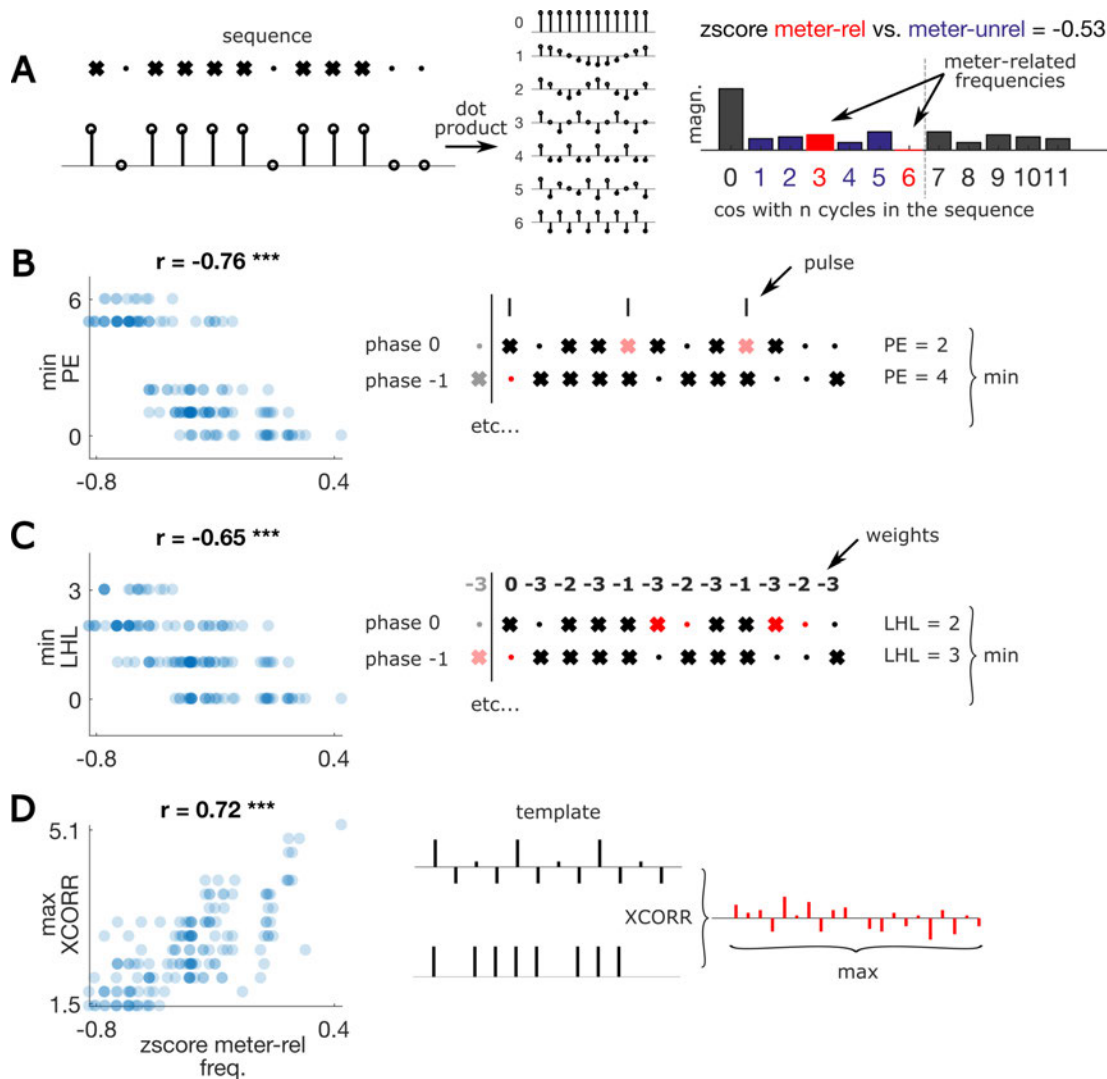


Figure 1.9. Different methods to assess contrast at metric periodicities using abstract representation of rhythmic sequences. **(A)** DFT approach used in music analysis. On the left, a rhythmic sequence is represented on a grid (see also Figure 1.1) as sound events (“x”) and silent events (“.”). Below is the same sequence represented as discrete series of ones (sound event) and zeros (silent event). Obtaining magnitude spectra using DFT involves taking the dot product of this series with complex sinusoids that complete 0 to $N-1$ periods within the span of the sequence (N is the number of events in the sequence). The absolute value of each dot product (spectral magnitude) is plotted on the right. For the complex sine wave with period 0, magnitude only

depends on the total number of sound events. Complex sine waves that complete 7 and more periods are aliased (mirror image of the first half of the spectrum) and can be ignored. Assuming meter with pulse periods 2 and 4 events, one can select frequencies with corresponding periods (and harmonics) in the spectra as meter-related. Z-scoring the magnitude across frequencies and taking the mean z-score at meter-related frequencies estimates how prominently meter-related frequencies stand out in the spectra, and thus the strength of contrast at these periods in the sequence. **(B)** Syncopation scores proposed by Povel and Essens (1985). The algorithm is shown on the right. Positions of the assumed pulse with period of 4 grid points are depicted as black vertical lines on the top. The score increases when silent events (dark red) or unaccented sound events (bright red) overlap with the pulse. The score is taken for each circular rotation of the sequence, thus testing each possible pulse phase and taking the minimum. As shown on the left, z-scored magnitude at meter-related frequencies calculated with DFT strongly correlates with PE score (Spearman's ρ , *** $p < 0.001$). The set of 160 rhythmic sequences to test the correlation was generated by assigning a 12-point grid with all possible permutations of 6, 7, and 8 sound events (while removing phase-shifted versions and only allowing maximum 4 successive silent events). **(C)** Syncopation scores proposed by Longuet-Higgins and Lee (1984). The algorithm is shown on the right, assuming pulse periods 2, 4, and 12 grid points (i.e. {2,2,3} meter). Each grid point receives a weight based on the number of coinciding pulses. If the sequence contains contrast in an opposite way as defined by the pulses (i.e. silent event preceded by sound event with lower weight, marked in red), the score increases. The minimum score is taken across all circular rotations of the sequence, thus testing each possible meter phase. **(D)** A template (top) that expresses the hypothesis about contrast at coinciding metric pulses with periods 2 and 4 is cross-correlated with the sequence represented as discrete series of ones (sound event) and zeros (silent event). Cross-correlation as a function of lag is shown in red, and the largest value is taken to represent how much the sequence fits the template, i.e. as a measure of contrast at meter periodicities. Again, the measure strongly correlates with meter-related z-scores obtained with DFT (shown on the left).

Even though the frequency-domain method may seem less straightforward than the other three approaches introduced above (i.e. PE, LHL and cross-correlation), its strength becomes apparent when one moves beyond rhythmic sequences represented on an abstract grid towards continuous rhythmic signals, where PE, LHL or cross-correlation methods cannot be easily applied (without suffering from the limitations inherent to time-domain approaches, see section 1.2.2.2). The main weakness of these methods is that they treat rhythms as discrete sequences of identical events (see section 1.1 for differentiation between continuous and discrete view of rhythm). Thus they abstract from the actual modulation waveforms that necessarily deliver the rhythmic information. However, real-

world rhythmic signals are more complex than a discrete sequence of identical unit-amplitude spikes.

To illustrate the advantages of the frequency-tagging approach, I will build the connection between arrangement of discrete events and continuous rhythmic signals using examples of linear systems. Throughout this discussion, it is important to keep in mind that by no means do I claim that the generative mechanism of any system analyzed in the empirical part of this thesis is linear. Yet, linear systems may provide important insights into the phenomenon of interest and the methods that measure it (Zhou et al., 2016; Broderick et al., 2018, 2019; O’Sullivan et al., 2019; Di Liberto et al., 2020a). For similar reasons, I will restrict the discussion to rhythmic signals that can be thought of as transformations of patterns generated on a fast isochronous grid. This way, one can explore how contrast at metric periodicities can emerge in continuously modulated rhythmic signals from (i) arrangement of events in time, and (ii) shape of signal modulation aligned to these individual events.

To transform a rhythmic pattern from an abstract representation as a sequence of ones and zeros (see Figure 1.9) to a real-world continuously modulated signal, one first needs to stretch it by inserting zeros (Figure 1.10 A vs. B). This operation results in exact repetition of the original spectra towards higher frequencies (“Stretch theorem” of the DFT, Smith, 2007). The resulting signal can be further convolved with a kernel that characterizes a particular system, and this can enhance or attenuate certain frequencies, as shown in Figure 1.10C, D and E. Indeed, characterizing the response of a system across the frequency gradient has important implications for the way it responds to signals with different temporal properties, and has been an important part of neuroscientific investigations (Nozaradan et al., 2012; Alonso-Prieto et al., 2013; Zhou et al., 2016). The shape of the kernel can have profound influence on contrast at particular periodicities. For instance, the kernel in Figure 1.10E completely zeros out magnitude at all harmonics of 5 Hz, effectively suppressing contrast at period $1/5 \text{ Hz} = 0.2 \text{ s}$. This can be directly seen in the time domain where the fast modulation with period 0.2 s disappears. The example in Figure 1.10E is informative, as the kernel is very similar to the one used in Figure 1.10C where one can still intuitively observe contrast with period 0.2 s in the time domain. The crucial difference is the frequency-domain representation of the kernel, as convolution in the time domain can be understood as multiplication in the frequency domain (Oppenheim and Schaffer, 2009). The kernel in

Figure 1.10C has periodically appearing zeros in the frequency domain, but they do not repeat with a period of exactly 5 Hz, thus still preserving magnitudes at harmonics of 5 Hz (i.e. 10, 15, etc. Hz) in the “input” signal that represents the arrangement of events in time (shown in Figure 1.10B). On the other hand, the zeros of the kernel in Figure 1.10E are spaced by exactly 5 Hz, thus completely cancelling all the harmonics. This example emphasizes the *importance of considering magnitudes at higher harmonics* when assessing contrast at a particular periodicity in signals. In other words, contrast representing a pulse with a particular period may be driven by energy at higher harmonics, which correspond to integer multiples of the fundamental frequency (inverse of pulse period). Hence ignoring these higher harmonics may yield a biased measure of pulse representation in the signal.

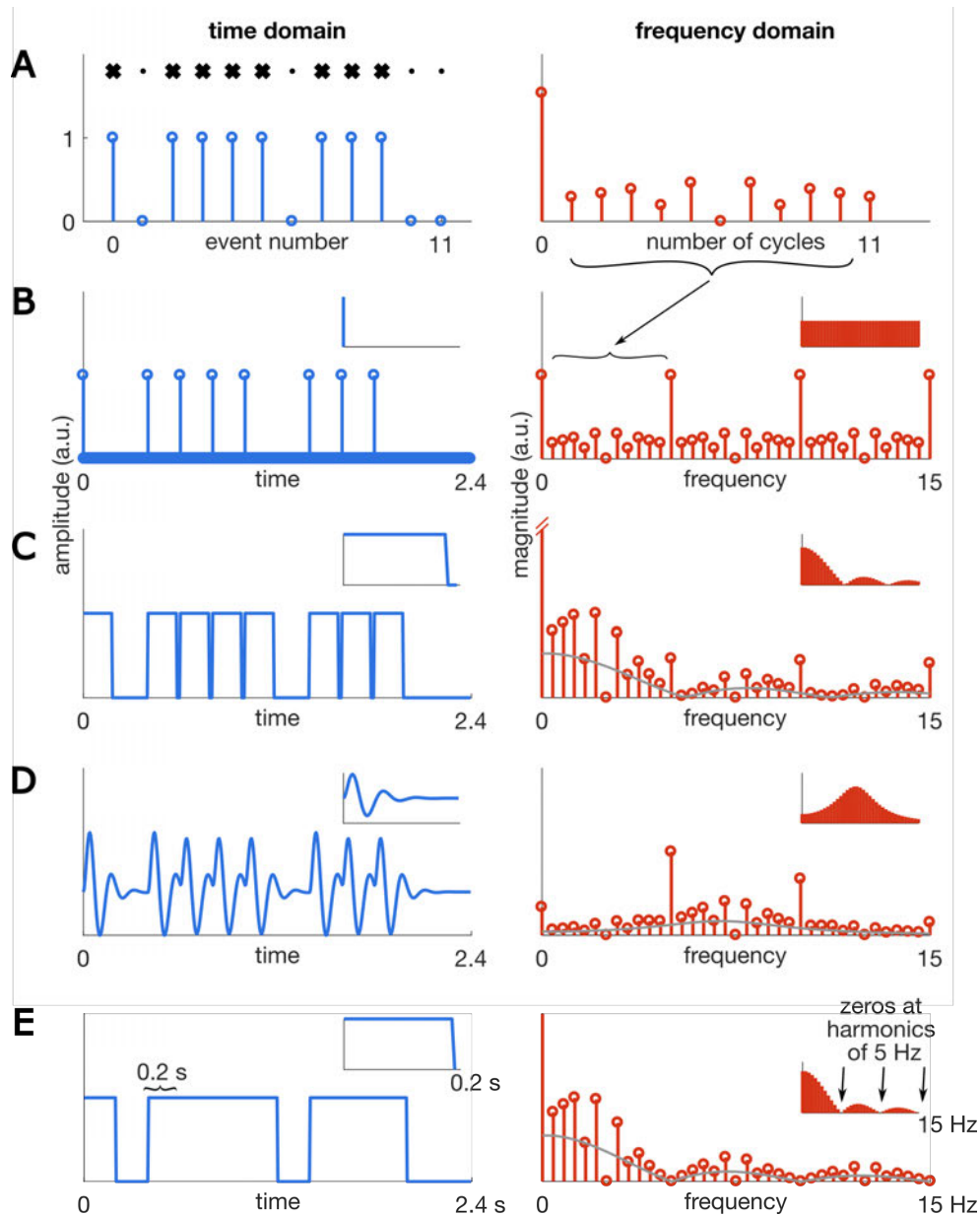


Figure 1.10. From grid representation of a rhythmic sequence to continuous modulated signal. Time-domain representation of the signal (i.e. variations in amplitude over time) is shown on the left (in blue) and frequency-domain representation (i.e. variations in magnitude over frequency) is shown on the right (in red). Insets depict time-domain and frequency-domain representation of a kernel used to generate the signal by convolving with signal shown in panel B. Grey curves overlaid over spectra in B, C, D, and E represent magnitude spectra of the respective kernels. This illustrates how a kernel can attenuate particular frequencies. **(A)** Abstract representation of a rhythmic sequence similar to Figure 1.9. The sequence can be represented on a grid (top) as sound events (“x”) and silent events (“.”), or as a discrete series of ones and zeros (sound and silent events respectively, bottom). **(B)** Transforming the sequence in panel A to a physical time dimension requires “stretching”, resulting in repetition of magnitude spectra as shown on the right. **(C)** Convolution of signal in panel B with a boxcar kernel shown in the inset changes the signal in both time and frequency domain. **(D)**

Example of a different kernel, reminiscent of an event-related potential (van Diepen and Mazaheri, 2018). **(E)** Same boxcar kernel as in panel C, but its duration is set to exactly 0.200 s. This way the kernel attenuates all harmonics of 5 Hz, effectively removing contrast at 0.200 s from the signal.

This draws attention to an important point, i.e. how to select meter-related and meter-unrelated frequencies in the spectra of continuously modulated rhythmic signals (such as the example signals in Figure 1.10), in order to obtain a robust and reliable measure of contrast at meter periodicities? Across previous studies using the frequency-tagging approach, the selection has often been heuristic. For instance, only frequencies with periods directly corresponding to the assumed metric pulses were considered as meter-related, while higher harmonics were not taken into account (Nozaradan et al., 2012, 2016b, 2016a, 2017b, 2018; Bouwer et al., 2020a). This can be justified for EEG signals by the fact that the EEG (as well as MEG) response typically has a low-pass characteristic, i.e. higher frequencies are attenuated (Wang et al., 2012). Combined with prominent sources of noise related to alpha activity and tonic muscle artifacts, this may lead to biased estimates of magnitudes in the higher portions of the spectrum, therefore justifying the focus on lower harmonics, which elicit responses with high signal-to-noise ratio. Moreover, some studies interpreted magnitudes at individual frequencies separately, without considering harmonically related frequencies as a whole (Stupacher et al., 2017; Tal et al., 2017; Li et al., 2019; Hickey et al., 2020). While the prominence of individual frequencies across the spectrum can provide important information about the system, a lower magnitude at a single frequency does not directly imply little contrast at the periodicity corresponding to that frequency. These heuristics have provoked criticism of the frequency-tagging method (Rajendran et al., 2017). To explore whether conclusions drawn from frequency-tagging depend on the selection of meter-related frequencies, a simple simulation was run based on the same pool of 160 rhythmic patterns created above (see Figure 1.9, testing the correlation between DFT-based method and PE, LHL and cross-correlation methods to measure contrast at metric periodicities in rhythmic patterns). As before, minimum LHL and PE syncopation scores across 12 phase shifts of each pattern were calculated to estimate contrast stemming solely from the arrangement of events in time. Here, the goal was to measure the prominence of contrast locked onto metric pulses with periods 1, 2, 4, and 12 grid points. Therefore, the

assumed metric pulses had periods 2, 4, and 12 grid points for LHL (note that LHL cannot measure period of 1) and period 4 grid points for PE (recall that PE can only measure one pulse at a time). Next, the abstract pattern representations were stretched as described in Figure 1.10, such that the time between successive grid points became 0.2 s. This signal was convolved with three types of kernels with very different effects on the spectrum: a unit impulse, boxcar, or decaying sine wave (see Figure 1.11A for time- and frequency-domain representations of the kernels). For each type of kernel, the resulting signal was transformed into the frequency-domain using FFT and magnitudes at meter-related and meter-unrelated frequencies were extracted after noise subtraction (see section 1.2.2.4), z-scored, and mean z-score at meter-related frequencies was calculated. Four different ways to select meter-related frequencies were used:

(1) Taking frequencies corresponding exactly to the periods of pulses in the assumed meter, i.e. $1/(0.2 \times [1, 2, 4, 12 \text{ grid points}]) = [5, 2.5, 1.25, 0.416 \text{ Hz}]$. This selection has been used across a large number of studies (e.g. Nozaradan et al., 2012, 2018; Lenc et al., 2018). The upper limit of 5 Hz has been used in EEG studies, with the justification that response magnitudes are smaller at higher frequencies due to the low-pass characteristic of the EEG signal.

(2) The second method included all harmonics of 1.25 Hz up to 5 Hz (i.e. 3.75 Hz was added to the set of meter-related frequencies). This selection therefore captured all harmonics (up to 5 Hz) related to pulses with periods of 1, 2, and 4 grid points (corresponding to frequencies 5, 2.5, and 1.25 Hz). The lowest frequency corresponding to the pulse period of 12 grid points (i.e. 0.416 Hz) was not considered as meter-related, as a contrast at this period cannot be easily measured in this setting. This is because the spectrum of a stretched rhythmic pattern only contains energy at the harmonics of this frequency. Thus, all frequencies in the spectrum would have to be considered meter-related, removing the possibility to normalize magnitudes using relative measures (such as z-scoring).

(3) The third selection method excluded 5 Hz, while keeping the rest of meter-related and meter-unrelated frequencies identical to method 1. This method has been used in EEG studies to make sure that changes in simple low-pass filter characteristics cannot explain putative contrast differences measured with method 1 (Nozaradan et al., 2018; Lenc et al., 2020).

(4) Finally, all harmonics of 1.25 Hz (similar to method 2) up to 30 Hz were considered. Note that these frequencies capture harmonics related to the pulse with period 4 and all the faster pulses in the assumed meter.

For each method, meter-unrelated frequencies were selected as all harmonics of 0.416 Hz (i.e. frequencies with non-zero magnitudes) that did not overlap with the meter-related frequencies and were in the frequency range of interest specified for each selection method.

The results are shown in Figure 1.11. When analyzed separately for each kernel type, all selection methods yielded strong correlations with LHL and PE scores, indicating that contrast resulting from the distribution of events in time was captured irrespective of the way meter-related and meter-unrelated frequencies were selected. All methods were also sensitive to the effect of kernel on the signal. This can be observed on the right side of each panel in Figure 1.11, where the distribution of meter-related z-scores across the 160 patterns shifts depending on the kernel used. Importantly, these differences between kernels were not consistent across the selection methods. For example in Figure 1.11D, generally larger meter-related z-scores for unit impulse kernel are only evident when high harmonics are included. These inconsistencies therefore raise question on the selection method to best capture the exact effect of kernel shape on the contrast at meter periodicities.

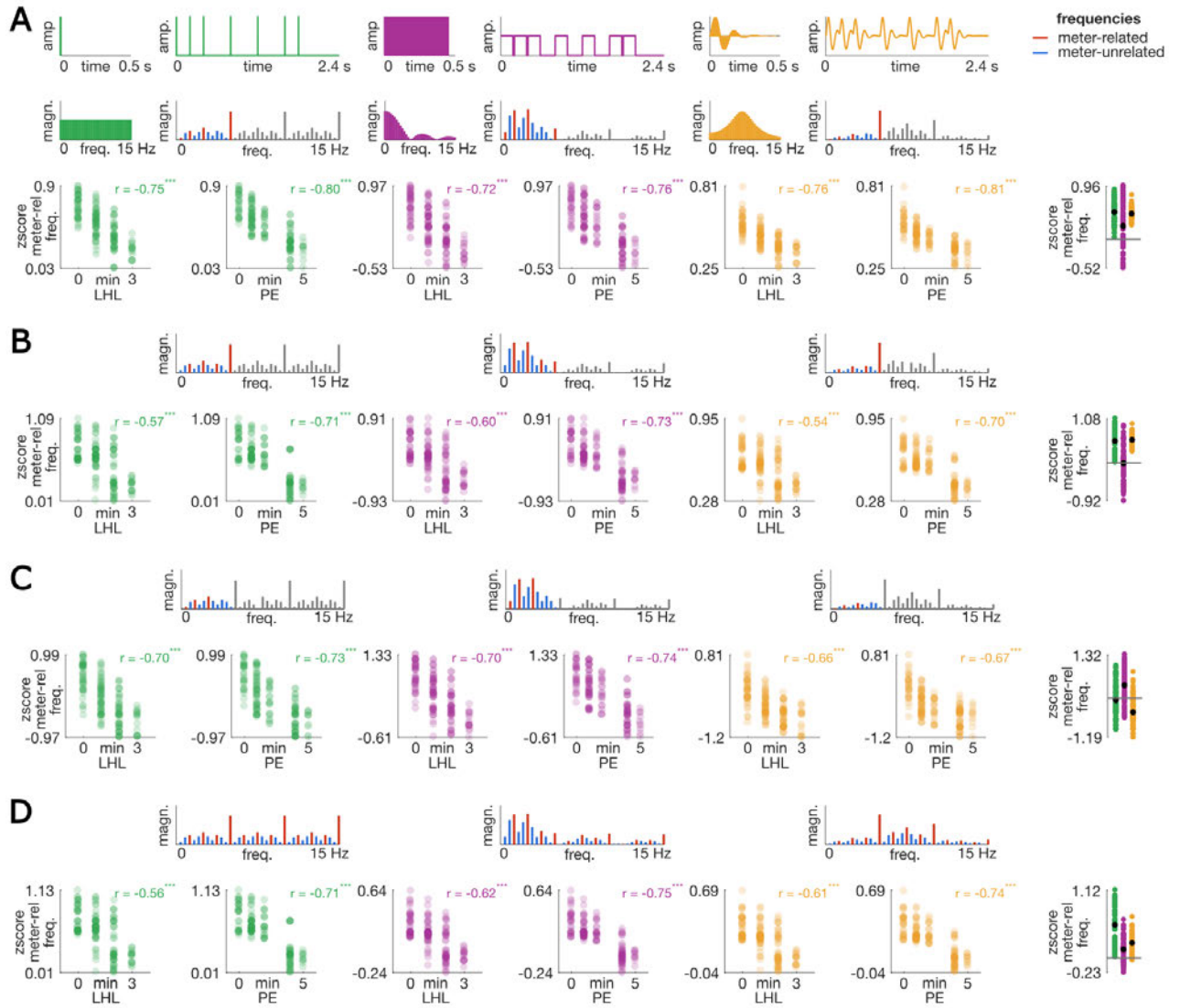


Figure 1.11. Response shape can attenuate or enhance contrast at specific periodicities. A set of 160 rhythmic patterns was used to synthesize rhythmic signals by convolving with different kernels displayed on the top of panel A (unit impulse shown in green, boxcar shown in magenta, decaying sine wave shown in yellow). Different methods to select meter-related frequencies are displayed across panels, using spectra of an example pattern. Meter-related frequencies are shown in red and meter-unrelated frequencies in blue. For each kernel and method, mean z-score at meter-related frequencies was correlated with LHL and PE syncopation score across the 160 patterns (Spearman's rho, *** $p < 0.001$), as shown on the bottom of each panel. The distribution of meter-related z-scores across the three kernels is shown on the right side of each panel (i.e. separately for each method, grey horizontal line represents zero). **(A)** Taking 0.416, 1.25, 2.5, 5 Hz as meter-related. **(B)** Taking 1.25, 2.5, 3.75, 5 Hz as meter-related. **(C)** Taking 0.416, 1.25, 2.5 Hz as meter-related and excluding 5 Hz altogether. **(D)** Taking harmonics of 1.25 Hz up to 30 Hz as meter-related.

As discussed above, kernel shape (representing frequency response of a system) can genuinely affect a periodic contrast (see Figure 1.10). Yet, there could also be situations where the response of a system affects the spectrum in a way that is not specific to the contrast at meter periodicities. An example of such a situation could be a change in the gain of the kernel, which results in a proportional magnitude increase across the whole spectrum (see Figure 1.5 and 1.8). Within the frequency-tagging approach, this is accounted for by normalizing the magnitudes (e.g. by using a z-score), and this should work irrespective of which frequencies are selected as meter-related (as long as there are any meter-unrelated frequencies to use for the normalization). Yet, the particular selection of meter-related frequencies may affect sensitivity to changes in the broad shape of the spectrum, such as low-pass filtering, that do not necessarily change contrast at meter periodicities. Therefore, it is important to determine which way of selecting meter-related frequencies is most sensitive to genuine changes in contrast, and most robust to biases from non-specific changes. To this end, another simulation was run, where contrast was directly manipulated in a controlled way using a similar method as in Figure 1.8. To make the simulation more relevant to studies measuring contrast at meter periodicities in EEG activity, the constructed signals had certain features reminiscent of EEG responses to auditory rhythmic stimuli (Nozaradan et al., 2012, 2018; Tal et al., 2017). The aim of the simulation was to explore how selection of meter-related and meter-unrelated frequencies affects sensitivity of frequency-tagging to genuine changes of contrast at meter periodicities, but also to noise and broad spectrum changes that are unrelated to the contrast at meter periodicities.

A single rhythmic pattern was selected, seamlessly repeated 18 times to create a longer sequence, and stretched into a real-world time-domain signal in the same way as described for the simulation above. The main difference from the previous simulation was that another version of the signal was created before convolving with a kernel. This contrast-enhanced version was characterized by additively increasing amplitude of an impulse at time points coinciding with metric pulses (see Figure 1.12A, and notice similarity to Figure 1.8). The meter consisted of two metric pulses with periods of 2 and 4 grid points, thus equal to 0.4 and 0.8 s after stretching. Ten trials were simulated separately for two conditions, one with smaller and the other with greater contrast enhancement. For each condition, “EEG-like” responses from 15 participants were simulated by convolving with a

kernel consisting in a sum of two decaying sinusoidal components using the equation below (see also van Diepen and Mazaheri, 2018).

$$kernel(t) = \left[A_1 \frac{t}{\tau_1} e^{1-t/\tau_1} \sin 2\pi f_1 t \right] + \left[A_2 \frac{t}{\tau_2} e^{1-t/\tau_2} \sin 2\pi f_2 t \right]$$

In the equation, t represents time from 0 to 0.5 s. The parameters were randomly selected for each participant (but fixed across conditions), as follows: A_1 was uniformly distributed between 0.4 and 0.8, and A_2 between 0.5 and 1. Next, τ_1 was normally distributed around 0.2 (SD = 0.05) and τ_2 around 0.05 (SD = 0.01). Finally, f_1 was normally distributed around 1 (SD = 0.2), and f_2 around 7 (SD = 1). This choice of parameters yielded a response somewhat similar to an ERP waveform observed from an actual EEG recording that seems to include sharp peaks (akin to P1, N1 components) but also a slow integrative component (see Figure 1.4 and inset in Figure 1.12B). Moreover, such a kernel attenuates higher frequencies, thus simulating the low-pass characteristic of EEG responses (Wang et al., 2012). Because the kernel was longer than the shortest time between successive events in the pattern representation, the successive responses elicited by each event did not go back to baseline but interacted, yielding a complex waveform. Therefore this simulation showcases the advantages of using frequency-tagging approach instead of time-domain approaches that would have limited success when analyzing such signal. Pink noise (i.e. 1/f) was added to the resulting signal, either with very low power (yielding high SNR = 2), or with high power (yielding low SNR = 0.1). The simulated trials were averaged in the time domain and transformed into the frequency domain using FFT, separately for each participant, and baseline was subtracted from each frequency bin (mean magnitude at bins 2 to 5 on both sides). Subsequently, meter-related and meter-unrelated frequencies were selected using the four methods described above, and magnitudes at these frequencies were converted to z-scores as for the previous simulations. The mean z-score at meter-related frequencies was compared across the two conditions (as in a repeated-measures design), i.e. testing whether the value was larger for the contrast-enhanced condition. 100 experiments were simulated separately for the low and high SNR, and for small and large contrast enhancement. The resulting effect sizes (Hedge's g) were calculated using the function *cohen.d* from package *effsize* (Torchiano, 2020), and aggregated over experiments by fitting a mixed model using the function *rma* from package *metafor* for R (Viechtbauer, 2010).

The results are shown in Figure 1.12D, suggesting greatest sensitivity when higher harmonics related to the metric pulse periods are taken into account. While other selection methods generally yielded effect sizes significantly different from 0, this was not the case with low SNR and small difference in contrast between conditions. Importantly, the direction of effect was the same across all selection methods. In other words, with genuine contrast differences between conditions the methods should give converging results. Figure 1.13A shows that when no contrast enhancement was present between the conditions but the power of background $1/f$ noise differed (thus decreasing signal-to-noise ratio in one condition), the methods yielded diverging effects depending on whether the lowest frequency (0.416 Hz) was included as meter-related or meter-unrelated. This is because $1/f$ noise interferes most with low-frequency components (thus could be considered a form of high-pass filter).

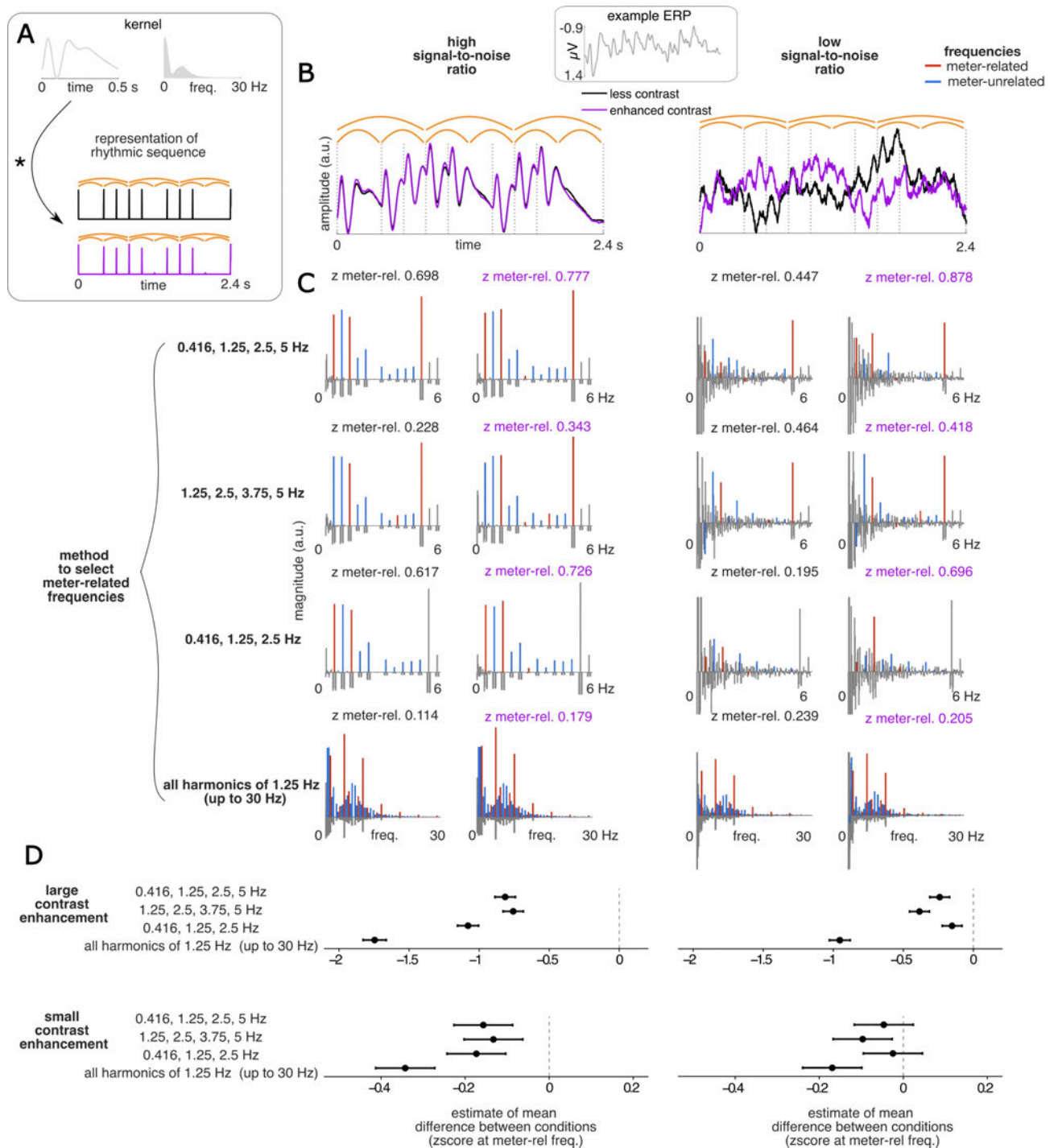


Figure 1.12. Simulation of contrast enhancement in a continuously modulated signal. Condition with high SNR is shown on the left, and low SNR on the right side. **(A)** EEG-like signals were simulated by convolving the representation of a rhythmic pattern with a kernel shown in grey. Signal with enhanced contrast at meter periodicities was generated by increasing the amplitude in the pattern representation at time-points defined by the metric pulses and not otherwise (see the difference between signals shown in black vs. magenta). Metric pulses are shown as yellow arches. **(B)** Example time-domain response simulated for one participant. The simulated signal was chunked into successive windows of 2.4-s duration (length of one pattern repetition) and averaged across all chunks and trials. Event positions in the rhythmic pattern are marked with grey vertical

dashed lines. Note the similarity of the simulated signal with an empirically observed ERP response to the same rhythmic pattern (grand average taken from fronto-central channels from high-tone syncopated condition in Study 2, see section 3) **(C)** Example magnitude spectra obtained from one simulated participant showing different methods to select meter-related frequencies (each row is one method, meter-related frequencies shown in red, meter-unrelated frequencies in blue). Mean z-scored magnitude at meter-related frequencies is shown on top of each spectrum. There may be slight variability in the spectra across rows due to noise added to the signals to simulate experiments. **(D)** Results of meta-analysis across 100 simulated experiments. Effect sizes with 95% CIs estimated across experiments using a mixed model are shown on a horizontal axis. Different methods to select meter-related frequencies are shown on the vertical axis. Results for simulations with large contrast enhancement in the second condition are shown on the top, and for smaller contrast enhancement on the bottom. Overall, all considered selection methods are sensitive to the difference between conditions. The method that uses all harmonics of 1.25 Hz up to 30 Hz consistently yields larger effect sizes, even when enhancement is small and SNR is low (bottom right).

Finally, the simulation was re-run, but instead of convolving with an “EEG-like” kernel, two different low-pass filters were applied (1st order Butterworth with cutoff at 10 vs. 20 Hz) in separate conditions without changing the contrast at meter periodicities. Note that this can be viewed as convolving the stretched representation of the rhythmic pattern (see Figure 1.12A) with two slightly different kernels, each representing one low-pass filter. The results shown in Figure 1.13B reveal that low-pass filtering affected the z-score at meter-related frequencies most when a large number of harmonics was used. The other selection methods (only considering frequencies up to 5 Hz) were biased to a lesser extent. Importantly, the bias occurred in an opposite direction when the highest frequency (5 Hz) was not included.

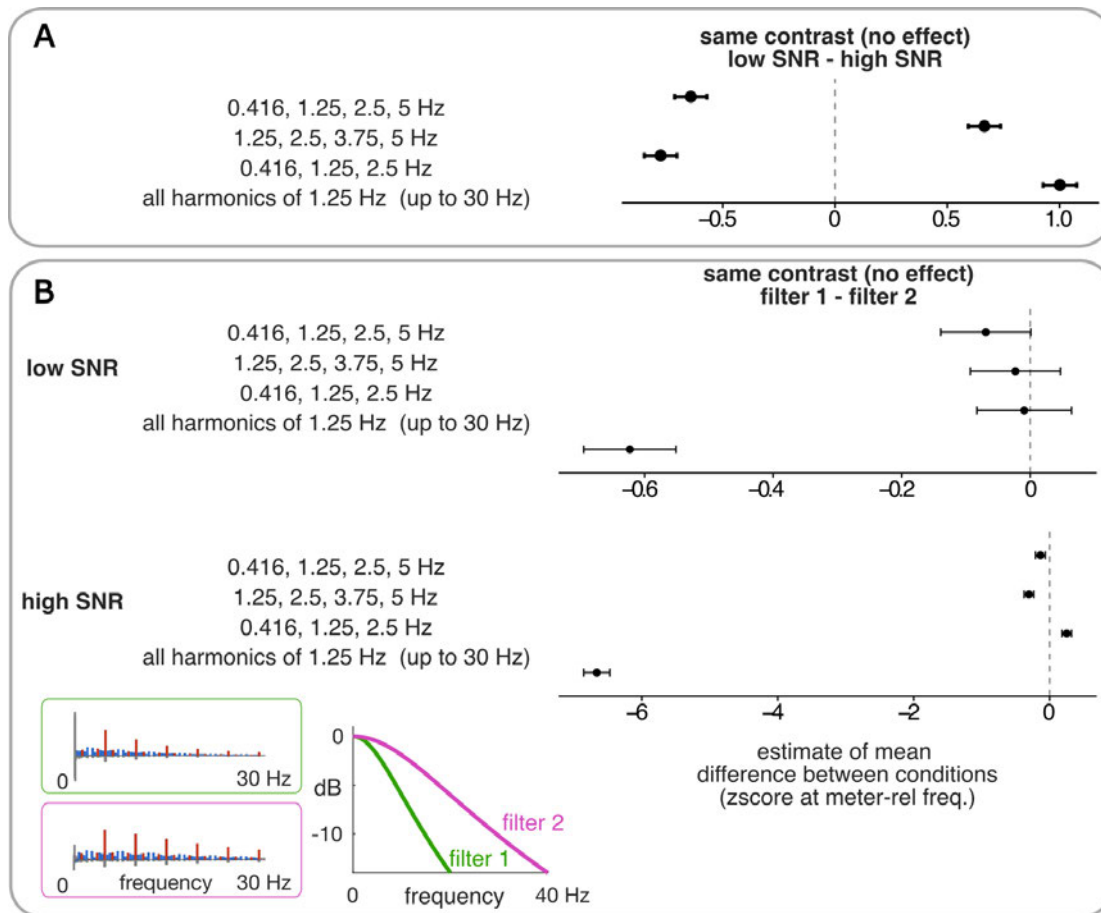


Figure 1.13. Results of meta-analysis across 100 simulated experiments, testing for spurious effects. **(A)** Using the same dataset and method as in Figure 1.12, two conditions with different signal-to-noise ratio, but with identical contrast at meter periodicities were compared. Low SNR (i.e. greater power of the 1/f noise) attenuated predominantly low frequencies, thus leading to spurious enhancement of the mean z-score at meter-related frequencies when the selected set contained 0.416 Hz (the lowest analyzed frequency). For selections that did not include this frequency as meter-related, low SNR led to spurious effect in the opposite direction, i.e. a decrease of the mean z-score at meter-related frequencies. **(B)** Applying low-pass filter (i.e. attenuating predominantly higher frequencies) with different cutoff frequencies across conditions with otherwise identical contrast at meter periodicities. The frequency response of each filter is shown on the bottom, along with the effect of each filter demonstrated on an example spectrum. Filter 1 (green) had lower cutoff frequency compared to filter 2 (magenta).

Together, these simulations can be summarized in the following way. When thinking of a rhythmic signal as a non-isochronous pattern of events arranged in time, frequency-tagging can capture systematic contrast locked onto metric pulses that emerges from this temporal structure (Figure 1.9). When generalizing towards continuously modulated signals, frequency-tagging is still sensitive to the temporal arrangement of events (Figure 1.11), as

well as contrast created by relative differences in the shape of signal modulation aligned to the individual events (Figure 1.8 and 1.12). However, when analyzing such signals, attention must be paid to the categorization of frequencies as related vs. unrelated to the periodic contrast in question. The different ways to select meter-related and meter-unrelated frequencies used in previous experiments (Nozaradan et al., 2012, 2016a, 2017b, 2018) seem all sensitive to changes in contrast at relevant periodicities (Figure 1.12). That is, if a change of contrast is present, the different methods should give converging results. However, the different methods are also prone to biases from non-specific signal changes such as low-pass filtering (Figure 1.13). Yet, the *direction* of these biases can differ when different subsets of frequencies are selected as meter-related. Thus, if a contrast change is absent but there is a broad change in signal's spectrum, the different selection methods can give diverging results. Consequently, multiple selection methods could be used to analyze the same dataset, and in case of a genuine difference in periodic contrast between the tested signals, the methods should all converge towards the same conclusion. This practice has been used in most studies conducted in our lab, i.e. checking the robustness of the results with different methods to select meter-related and -unrelated frequencies.

While thinking about any signal as a series of discrete events is good as an exercise, applying such a view to continuous signals would be an oversimplification. In fact, even a musical audio signal cannot be simply reduced into a series of discrete events without losing important information (yet, this is a common approach, see e.g. Lerdahl and Jackendoff, 1983; Longuet-Higgins and Lee, 1984; Povel and Essens, 1985; Amiot, 2016; Milne et al., 2017). For instance, slow gradual modulations of features (e.g. loudness, filter cut-off) are commonly used to create rhythms, yet there is no easy way to reduce these into discrete time points. Indeed, if a continuous signal is reduced into a series of discrete onsets, one may lose important information about the signal that is relevant for contrast. Even in a simple case when the signal could be explained by convolution of discrete onset series with a kernel representing the system's impulse response (which is not possible for most non-linear biological systems, see e.g. Pikovsky et al., 2003; Kuchibhotla and Bathellier, 2018), the shape of the kernel could lead to resonances and therefore must be taken into account when assessing periodic contrast generated by the system (Galambos et al., 1981; Ross et al., 2000; Alonso-Prieto et al., 2013; Teng et al., 2017; Arnal et al., 2019; Lozano-Soldevilla and VanRullen, 2019). To better investigate the contribution of such resonances to the

transformation relevant for meter processing, future empirical studies should systematically test rhythmic inputs containing contrasts at a range of tempi (thus extending the previous work, see e.g. van Noorden and Moelants, 1999; Nozaradan et al., 2012, 2017b, 2018). Specifically, slightly changing input tempo would shift and stretch its spectrum but preserve spectral magnitudes across the individual (shifted) frequency components. This way, in one condition a particular frequency (in Hz) may be considered meter-related and in another condition, the very same frequency (in Hz) may be considered meter-unrelated. If the system passively enhances contrast at meter periodicities due to resonances related to its impulse response, the same absolute frequencies (in Hz) would be systematically enhanced/suppressed according to the frequency-content of system's kernel (see Figure 1.10 and 1.11), irrespective of input tempo. If the system goes beyond passive resonances, it should be capable of selectively enhancing the shifted meter-related frequencies across a range of tempi.

1.2.2.6 Contrast affected by fine time locking

So far I have discussed arrangement of events in time only from the viewpoint of a discrete, isochronous grid. While this conceptualization is insightful, real-world signals contain much more subtle variations in the temporal arrangement of events. Importantly, small timing inconsistencies can lead to less systematic contrast in time (due to less precise “generalization”, i.e., values at time points separated by a stable period are less similar), and thus need to be accounted for when using direct measures of metric pulses in sound, brain, and movement. Different methods used to measure fine temporal consistency of contrast in signals are typically based on the expectation of coherent phase at time-spans defined by the period of a metric pulse. I first discuss these methods from the perspective of movement, and show that frequency-tagging can yield comparable results.

While syncopation scores (e.g. LHL or PE introduced in the previous section) can be thought of as measuring representation of meter in the sound input, analyzing movement represents the other end of the perception-action processing loop: the behavioral output. A common way to assess meter perception is to ask the participant to tap (e.g. with their hand) along with a pulse they perceive in the rhythmic input (Handel and Oshinsky, 1981; van Noorden and Moelants, 1999; Drake et al., 2000; Snyder and Krumhansl, 2001;

Toiviainen and Snyder, 2003; Nozaradan et al., 2012; Large et al., 2015). This is an ecologically valid task, as humans spontaneously move along with a pulse when listening to music, e.g. by tapping their foot or bobbing their head. Importantly, most methods used to analyze spontaneous tapping data can be conceptualized as searching for a contrast at periodicities defined by metric pulses. In other words, to claim that the participant has an internal representation of a metric pulse, taps must (i) consistently occur at regularly spaced time points (ii) locked onto the temporal structure of the rhythm, and (iii) not otherwise.

Traditionally, tapping data is analyzed as a discrete series of tap-onset times, i.e. identical events arranged in time (Repp, 2005; Repp and Su, 2013). Regular spacing of taps can be typically quantified as the variability of inter-tap intervals (ITIs) normalized by the mean ITI (i.e. a coefficient of variation). An equivalent circular measure represents each tap as a vector with length 1 on a unit circle where 2π represents the mean inter-tap interval and 0 degrees is typically set to the beginning of the stimulus (or trial). The mean vector length is taken as a measure of regularity (Nozaradan et al., 2016b). However, these measures are not plausible as full measures of meter processing. While they are sensitive to the consistency of a periodic contrast in time, they do not implement the constraint of time-locking from the definition of meter. In other words, tapping with a stable period would yield low variability even if no time-locking onto the rhythmic stimulus occurred.

A better measure emerges from classic sensory-motor synchronization paradigms where subjects tap 1:1 with an isochronous metronome. These quantify time-locking as the variability of asynchronies between tap times and sound-event times (Repp, 2005). This is a measure of uni-directional entrainment, or synchronization between two systems, which can generally have more complex forms, e.g. 2:1, 3:1 etc. (Pikovsky et al., 2003; Zelic et al., 2018). However, when synchronizing with the perceived pulse induced by a rhythmic sequence, the pacing acoustic signal is not isochronous, thus the synchronization is not easily expressed with one ratio (Repp and Su, 2013). Instead, it is assumed that the tapping is synchronized with an internal representation of one pulse from the perceived meter (Parncutt, 1994; van Noorden and Moelants, 1999; Nozaradan et al., 2012) (but see Study 3, section 4, and Figure 1.18 for a discussion of non-stationary tapping). Thus the first step in the analysis is typically to determine the period of this metric pulse.

For highly controlled rhythmic patterns constructed on an isochronous grid without expressive timing and tempo changes, a common heuristic to determine the plausible

metric pulses is to take integer multiples of the grid interval. This constraint yields pulses with integer ratios to any interval in the rhythm. However, it is important to note that such mathematical derivation becomes increasingly difficult when naturalistic stimuli from different musical traditions are considered (Polak et al., 2016, 2018).

After defining plausible pulses, the analysis must determine one of them that was likely tapped by the participant. This can be done heuristically by visually inspecting the distribution of ITIs, or selecting the plausible period with the smallest distance from the mean ITI (Kung et al., 2013; Cameron and Grahn, 2014; Nozaradan et al., 2016b). This period is used to define a pulse, with a starting point typically set to the beginning of the stimulus. The variability of signed time differences from each tap to the closest pulse position (i.e. asynchronies) can be taken as a measure of regular spacing of the taps, while controlling for time-locking with the metric pulse (Repp and Su, 2013). A circular equivalent of this measure helps to avoid spuriously high variability when the taps consistently occur halfway between successive pulse positions, as shown in Figure 1.15. The individual taps are mapped as unit vectors onto a circle where 2π equals the selected pulse period and phase 0 is typically set to the beginning of the stimulus, and the mean vector length is taken to quantify regularity. Indeed, it is not the actual distance, but rather *consistency* of the distance between the taps and positions of the pulse that matters. This is in line with the definition of a periodic contrast as a *relative difference* repeating at a particular rate (see section 1.2.2.2 and Figure 1.5).

Based on this definition of a periodic contrast, it may be somewhat misleading that many researchers have been putting strong emphasis on the “predictive” nature of sensory-motor synchronization to the beat (i.e. a metric pulse) in music (Patel and Iversen, 2014; Ross et al., 2018b). This stems from classic sensory-motor synchronization studies where negative mean asynchrony with respect to the pacing stimulus is commonly observed (for reviews, see Repp, 2005; Repp and Su, 2013). Accordingly, considerable effort has been invested into training animals to synchronize with asynchronies shorter than typical reaction times (Yc et al., 2018). Indeed, it is important to ensure that consistent asynchronies cannot be explained by “passive” reactions to each pacing sensory event. However, in the case of synchronizing to a metric pulse in music, such explanation is impossible, as perceived pulses are not one-to-one with the acoustic events in the rhythmic stimulus. Therefore, even if individual taps were “reactions” to the perceived pulse, this would imply that the pulse is

perceived, which is the main point of the analysis in the first place. Consequently, an appropriate measure of a systematic contrast should be sensitive to jitter and period locking, but should not be affected by systematic misalignment.

A related issue is that the phase of the perceived metric pulse that serves as an internal pacing signal is difficult to estimate from tapping data itself (see Figure 1.14). This is especially the case for patterns constructed on a fast isochronous grid (i.e. with tight spacing of the subsequent points). Even if the participant is targeting a particular time position, the executed taps can be systematically misaligned from this position (Repp and Su, 2013). The least subjective way to determine the “target times” from the data is to quantize each tap to the closest grid position and determine the tapped phase based on grid positions with the largest number of quantized taps. However, because of negative mean asynchrony common in human tapping, quantizing to the closest grid point irrespective of direction may provide a biased estimate. Thus, caution should be taken when making conclusions based on the phase of the perceived meter estimated from tapping data (Rajendran et al., 2017). However, the issues related to estimating meter phase do not apply to the measures of contrast at meter periodicities based on the variability of asynchronies, mean vector length, but also frequency-tagging.

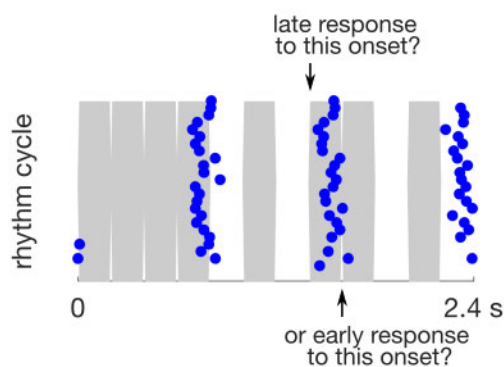


Figure 1.14. Example of uncertainty when assigning taps to individual events in a rhythmic pattern constructed on an isochronous grid. Data from one participant spontaneously tapping the pulse perceived in a rhythmic pattern cycled for 60 seconds. The audio waveform of one pattern cycle is shown in grey. Overlaid as blue points are the individual tap times, wrapped at cycle boundaries. The taps cluster around stable positions across pattern cycles, indicating time-locking. However, it is difficult to determine which exact event onset the taps are locked to.

Importantly for our discussion, frequency-tagging is sensitive to the fine-temporal consistency in signals. This is illustrated in Figure 1.15 and 1.16, which highlight the similarities between traditional robust measures of synchronization (such as circular mean vector length) and frequency-tagging. The reason FFT magnitudes decrease when the tapping is not stable is the stationarity assumption of the Fourier Transform. In other words, analyzing longer signals assumes stable phase of the signal, which is not the case if temporal variability is increased (particularly for higher harmonics). Similarly, if the contrast in tapping occurs at a period that is unrelated to the assumed meter, magnitude at meter-related frequencies will be near zero.

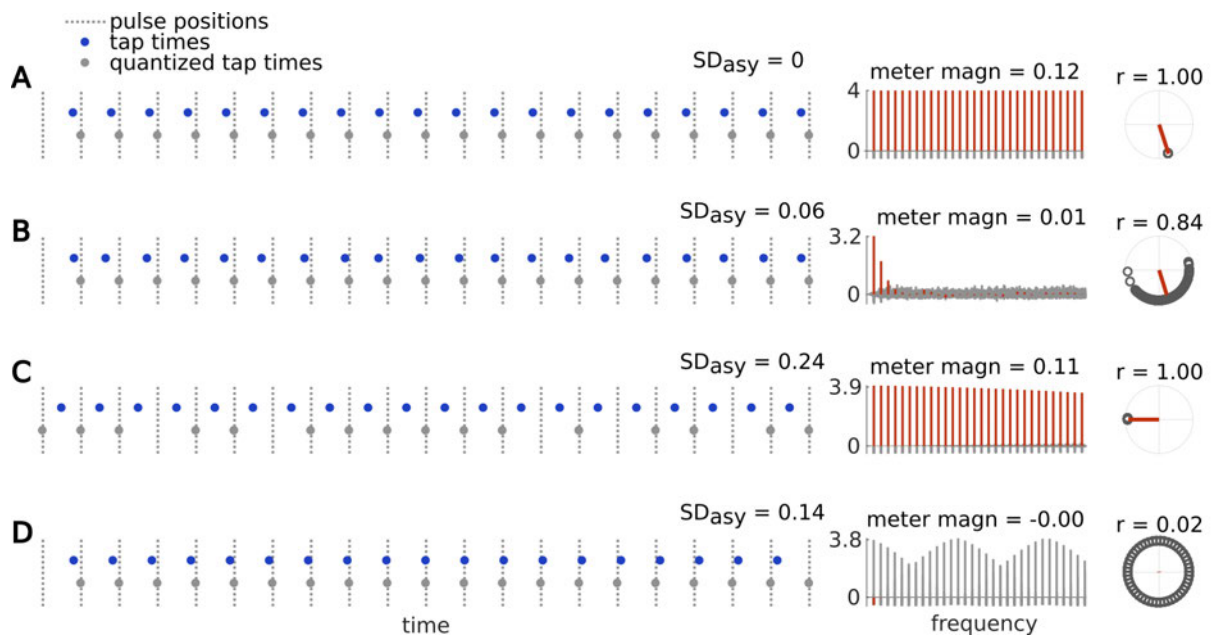


Figure 1.15. Examples of tapping data. Tap times are shown as blue points. Time points defined by the assumed metric pulse with period 0.5 s are shown as grey vertical dashed lines. Three measures of time locking onto the pulse are shown on the right: standard deviation of asynchronies, FFT magnitudes summed across meter-related frequencies, and mean vector length. Standard deviation of asynchronies (SD_{asy}) was calculated with respect to target positions, determined by quantizing each tap to the closest pulse position (grey circles). Magnitude spectra were calculated by applying FFT to time series where each tap is represented as an impulse with amplitude 1 (zero otherwise). Meter-related frequencies (in red) were selected as harmonics of the pulse period (i.e. 2, 4, 6, ..., Hz) and magnitudes at these frequencies were summed (value displayed above each spectrum). For the circular method, each tap is displayed as a point on a unit circle. The mean vector is shown in red and its length r is displayed above each unit circle. **(A)** Inter-tap interval is exactly 0.5 s (i.e. identical to the pulse period) and taps consistently occur 0.1 s before each pulse position. This yields low SD_{asy} , high magnitude at meter-related frequencies, and high r . **(B)** Same as above, but jitter with

standard deviation 0.05 s added. This leads to increased SDasy, lower FFT magnitude at meter-related frequencies, and lower r . **(C)** No jitter, but taps systematically occur 0.25 s after each pulse (i.e. in the midpoint between successive pulse positions). This leads to high FFT magnitude at meter-related frequencies, high mean vector length, but artificially high SDasy because the quantization randomly assigns the target position to the previous or following pulse position. **(D)** No jitter, but inter-tap interval is 0.51 s, thus not locked to the assumed pulse period. This leads to high SDasy, and essentially zero FFT magnitude at meter-related frequencies and r .

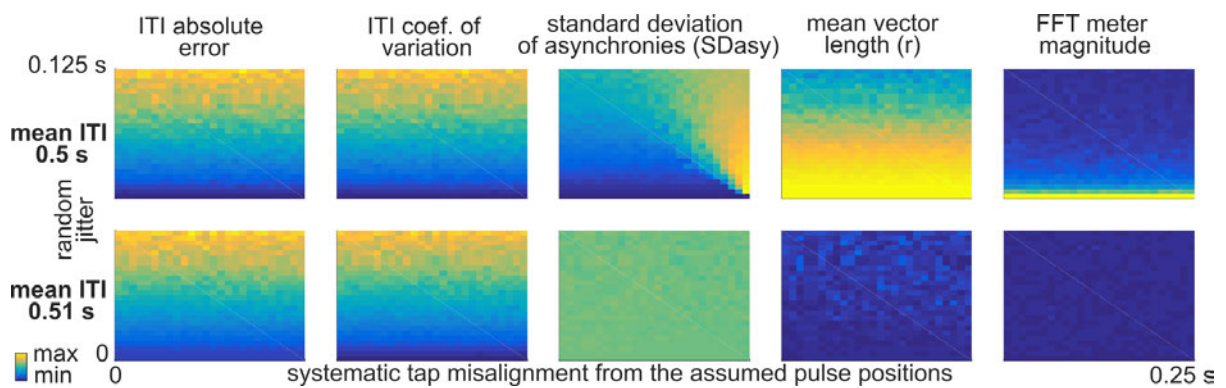


Figure 1.16. Sensitivity of different methods to systematic contrast locked onto a metric pulse. Same setup as for Figure 1.15, but two parameters of tapping were continuously manipulated: (1) the systematic misalignment from the assumed pulse shown on the x-axis, and (2) random jitter shown on the y-axis. Moreover, the top row shows tapping with period identical to the assumed metric pulse (0.5 s), whereas in the bottom row the tapping period 0.51 s is unrelated to the metric pulse. While a good measure of a systematic contrast should be sensitive to jitter, and period locking, it should not be affected by systematic misalignment. In addition to three analysis methods introduced in Figure 1.15, ITI absolute error (mean difference between inter-tap intervals and pulse period) and coefficient of variation of ITIs (standard deviation / mean) are included. These measures based on ITIs are sensitive to jitter and not systematic misalignment, but they give similar results even when tapping period is not locked onto the pulse, i.e. does not define contrast at the correct periodicity. Note that this is not exactly true for ITI absolute error, as this measure necessarily increases as the tapped intervals get further away from the target interval. However, the small change in the current simulation cannot be easily seen in the figure, despite the fact that the color scales are fixed across the top and bottom row for each measure. Standard deviation of asynchronies (SDasy), mean vector length (r), and frequency-tagging are all highly sensitive to period locking. Yet, SDasy is inflated when the systematic misalignment gets close to the midpoint between successive pulses. Frequency-tagging method seems more strict than r when jitter increases, yet both methods respond in qualitatively similar ways.

Because tapping data is typically abstracted into a series of tap times, the tapping force or other parameters of the movement are usually not taken into account (in other words, what matters is the phase, not amplitude, Pikovsky et al., 2003). However, when thinking of meter as a transformation towards a periodic contrast, tapping force is an important source of information (Keller, 2012). Examples of this are shown in Figure 1.17 (notice the principle is similar to Figure 1.8). This showcases the utility of the frequency-tagging method that is capable of capturing the difference between panel B and C, whereas these two signals would be treated identically by measures only considering tap timing and not force. These examples also show another advantage of using z-score normalization of the extracted magnitudes instead of other relative measures (such as proportion between mean magnitude at meter-related vs. -unrelated frequencies). As shown in Figure 1.17D and E, the z-score at meter-related frequencies is not affected when meter-unrelated frequencies have zero magnitudes and the magnitudes at meter-related frequencies change. This is a commonly encountered situation, where participants tap the pulse with extraordinary precision (yielding data similar to Figure 1.17D and E), yet, some of these participants may tap with a greater overall force. Thus, using z-scores ensures that the contrast at meter periodicities within the tapping of these participants is not overestimated when using frequency-tagging.

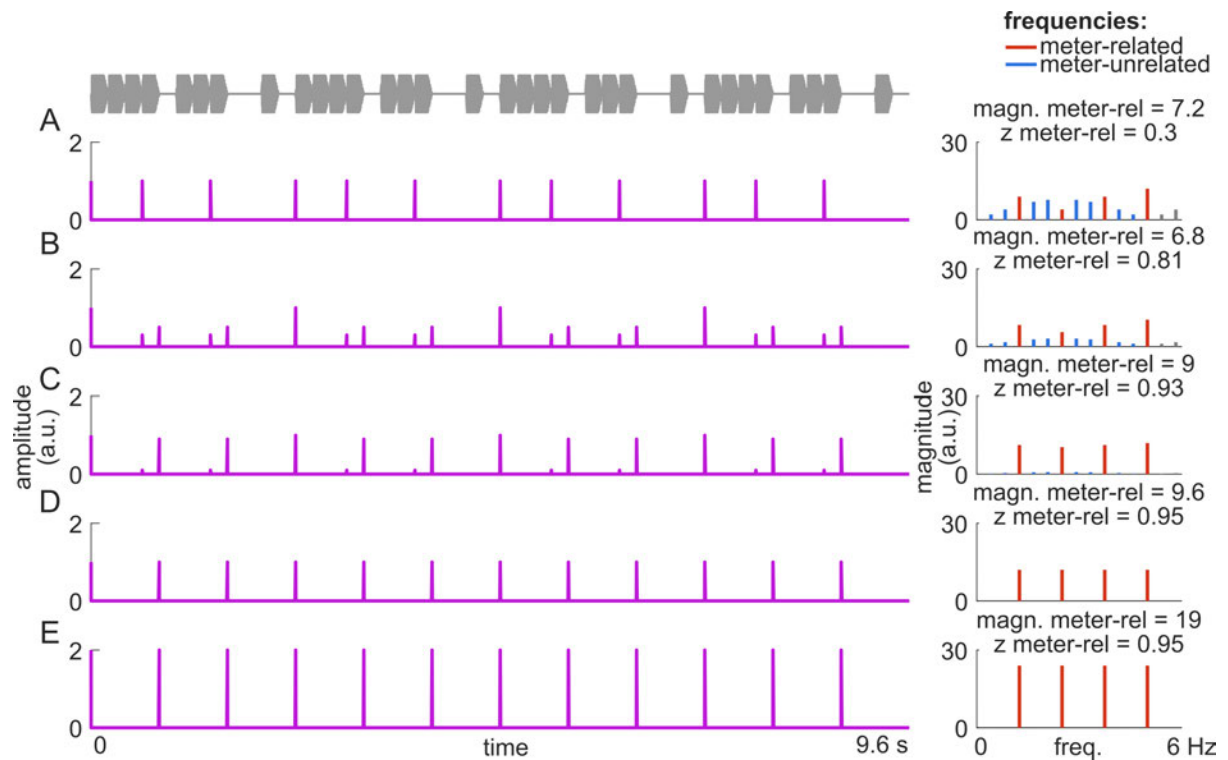


Figure 1.17. Importance of taking tapping force (i.e. amplitude) into account when assessing contrast at meter periodicities. The rhythmic stimulus sequence is plotted on the top in grey. Simulated tapping data are shown in magenta. The height of the peaks represents tapping force. FFT of the tapping signal is shown on the right. The contrast is quantified with respect to a metric pulse with period 0.8 s, thus harmonics of this frequency up to 5 Hz are considered meter-related (in red). Harmonics of the frequency corresponding to the pattern repetition rate that do not overlap with the meter-related frequencies are considered as meter-unrelated (in blue). Two measures are reported above each spectrum: (1) summed magnitude at meter-related frequencies, and (2) mean z-score at meter-related frequencies. **(A-D)** Gradually emerging contrast with period related to the metric pulse. While in panel A, the simulated participant taps with particular sound events in the rhythmic stimulus, the tapping does not define a periodic pulse. This pulse gradually emerges in panels B, C, and D as the amplitude of taps that disrupt the periodic contrast decreases. Both, summed magnitudes and z-scores are sensitive to these changes. **(E)** This signal has identical contrast as the signal in panel D, however the participant increases the overall tapping force. While summed magnitude at meter-related frequencies is sensitive to this change, the z-score remains constant.

While tapping can provide rich information about the perceived meter, there are certain caveats to keep in mind. Even if we were able to determine which events within a rhythm represent target time points for the tapped pulse (which is quite problematic on its own as discussed above), the tapped pulse can differ from the perceived metric pulse that is

internally used to organize the movement in time (Repp, 2007; Repp et al., 2008). A typical example is swing, where the phase of the tapped pulse is systematically misaligned from the perceived metric pulses (for a demonstration, see Lewis, 2018). Another issue, already mentioned in previous sections, is that only a single pulse can be tapped at once. Thus the whole set of pulses simultaneously perceived within the meter can remain concealed. Relatedly, if the participant frequently changes the tapped pulse from the perceived meter, specific issues arise during data analysis using FFT (see Figure 1.18 and Study 3, section 4). Moreover, while convenient to use within an experimental setup, tapping with the hand may not be the most natural form of movement when listening to musical rhythm (Janata et al., 2012; Burger et al., 2014). Finally, active movement can significantly change meter perception (Su and Pöppel, 2012).

Nevertheless, movement is closely linked to rhythm perception in most ecological settings (Phillips-Silver et al., 2010; Maes et al., 2014), and it may not be plausible to artificially separate the two (see also Patel and Iversen, 2014). In addition, despite the limitation to tap a single pulse at a time, the whole nested set of perceived metric pulses can be estimated when taking into account variability across and within individuals (e.g. across participants and trials within an experiment). Indeed, participants often differ in terms of which pulse from the perceived meter they choose to tap (McKinney and Moelants, 2006; Martens, 2011). Thus, if a particular meter is consistently perceived by participants when listening to a specific input, the distribution of tapped periods (individually estimated from the inter-tap intervals, see above) should converge towards a single plausible meter (i.e. satisfying the nesting constraint, see section 1.1). A similar approach has been used in the empirical studies presented in the current thesis.

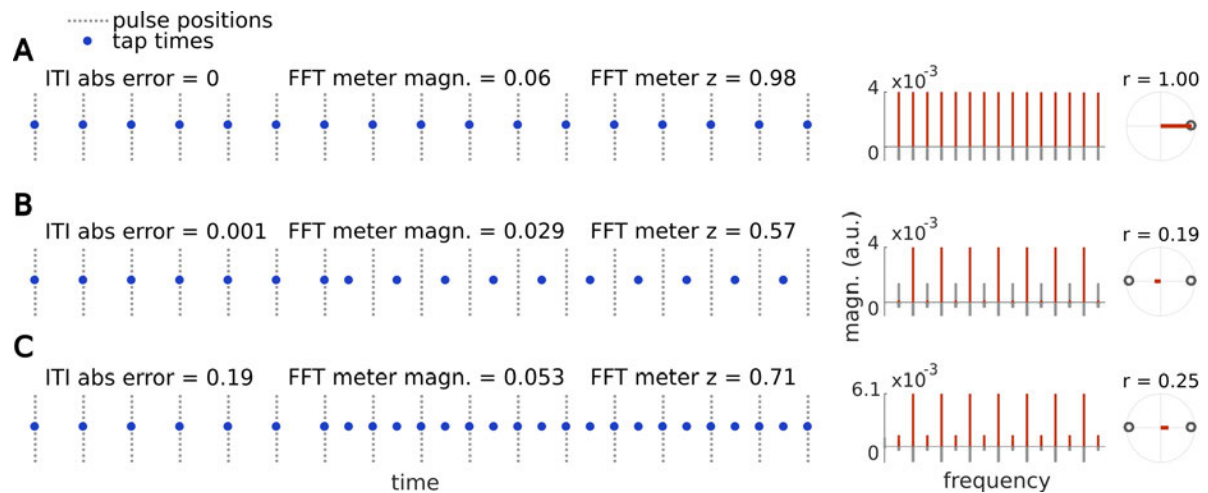


Figure 1.18. Sensitivity of different methods to changes in systematic alignment and tapping period. Data were simulated and analyzed in the same way as for Figure 1.15, i.e. assuming metric pulse with period 0.5 s. In addition to Figure 1.15, ITI absolute error was measured as a mean difference between inter-tap intervals and assumed pulse period. Mean z-score at meter-related frequencies (in red) was calculated as well, using harmonics of 1 Hz that did not overlap with harmonics related to the pulse period as meter-unrelated frequencies (see also Figure 1.17). No jitter was applied in any condition. **(A)** Example of perfect tapping performance with inter-tap interval exactly 0.5 s (i.e. identical to the pulse period) and taps consistently occurring at assumed pulse positions. This yields zero ITI absolute error, high magnitude and z-score at meter-related frequencies, and high r . **(B)** In approximately one third of the trial, the systematic alignment with respect to the assumed pulse positions changes from 0 to -0.25 s, and is kept constant for the rest of the trial. This does not strongly affect ITI absolute error, but magnitude and z-score at meter-related frequencies are decreased, as well as r . **(C)** In approximately one third of the trial, the tapped period changes from 0.5 to 0.25 s, and is kept constant for the rest of the trial. That is, the simulated participant changes to a different pulse from the same perceived meter. This strongly increases ITI absolute error and decreases r , but magnitude and z-score at meter-related frequencies are more robust (i.e. more similar to the values in panel A).

While this section mainly focused on signals originating from tapping, the conclusions about the frequency-tagging approach hold for any type of signals, including brain activity measured with EEG (Lenc et al., 2019). One important note concerns a measure of phase stability across trials (ITPC, inter-trial phase coherence). While this method is useful in certain paradigms, it has little potential to capture periodic contrast. Indeed, contrast is defined over time and not over trials, therefore stable phase across trials could arise even from a signal that has little contrast at meter periodicities (note that this may depend on how trial duration is defined). Therefore, the increasingly widespread use of this method to

measure signal properties relevant for meter processing is surprising (e.g. Cameron et al., 2019; Hickey et al., 2020).

To summarize, real-world rhythmic signals can create contrast at meter periodicities. The prominence of this contrast depends on generalization, differentiation and time locking. These factors can be affected by (i) arrangement of events in time, (ii) systematicity in time, in other words, temporal variance or jitter, (iii) shape of the modulations aligned to individual events (i.e. frequency response of the system), (iii) relative changes in the shape of modulations aligned to individual events. It is important to note that in real-world signals and systems these factors are related, and one cannot fully attribute contrast to a single one. For instance, the arrangement of identical events in time (Figure 1.9), and attenuation/enhancement of certain events (Figure 1.8) can be understood as the same kind of signal manipulation. The only difference is that for the former, attenuation of certain events is always all the way down to zero. Moreover, real-world rhythmic signals do not objectively consist of discrete “events”, and real-world systems are often nonlinear, i.e. cannot be characterized by a stable kernel (unlike the linear systems simulated in this section) (see e.g. Keshishian et al., 2020). Yet, such abstractions can be useful to gain deeper insights into the nature of contrast in rhythmic signals, and the methods used to measure it. It is important to emphasize that the method I have been developing here is a general method to measure contrast in a signal. It can be used for sound, brain, or movement signals, but it *should not be interpreted as a perceptual measure of cues to meter periodicities in the sound* (even though it correlates with measures such as PE or LHL that have been developed to directly model perceptual experience) (see e.g. Henry et al., 2017 for such misinterpretation). The fact that when applied to rhythmic audio signals, the contrast measured with frequency-tagging does not map directly onto perceptual experience is not surprising. One could say it is almost trivial, considering the overwhelming evidence across perceptual domains that distinct sensory inputs can be mapped onto similar perceptual experiences (many-to-one mapping, see section 1.2.2.1). In fact, this is exactly the purpose of the method: providing a tool to describe mappings (i.e. transformation) between acoustic input, brain activity, and behavioral output, instead of directly offering a model of mapping between sensory input and perception.

In this section, I have demonstrated that frequency-tagging is a powerful approach to measure contrast at meter periodicities. The method is sensitive to all fundamental signal properties that are crucial to measure periodic contrast that defines metric pulses. In the following section, I briefly summarize previous studies that have addressed the transformation between sound input, brain response and behavioral output with regards to meter perception. Based on this review, I propose a set of fundamental questions that I aimed to address in a set of experiments constituting the empirical part of this thesis.

1.3 Nature of the phenomenon

The central hypothesis of this thesis is that meter perception fundamentally involves a transformation of a rhythmic sensory input towards metric pulses. This transformation involves selective enhancement of contrast defining the perceived metric pulses and can be captured when comparing brain activity and movement to the sound input. Thus, one way to provide evidence for this hypothesis is to show that brain activity or movement with prominent contrast at periodicities corresponding to a particular meter can be elicited by a sound input with little contrast at those particular periodicities. Such observations have been reported across a large number of studies over the last decade.

1.3.1 Transformation from sound to behavior

One source of evidence comes from analyses of musical corpora, showing that meter typically perceived for a musical piece cannot be directly explained by the periodic contrasts emerging from its acoustic structure (London et al., 2017; van der Weij et al., 2017). However, these analyses do not directly measure contrast in the behavioral output, and instead rely on more or less anecdotal evidence of how the rhythmic stimuli are typically perceived by listeners (and dancers) familiar with the genre. Other studies provided clear evidence of transformation while directly measuring contrast at meter periodicities in both sound input and behavioral output in the form of movement (Chapin et al., 2010; Nozaradan et al., 2012, 2018; Large et al., 2015; Tal et al., 2017). This view may seem incompatible with a number of widely cited studies that claim to show decreased contrast at meter periodicities in tapping when this contrast decreases in the rhythmic stimulus (i.e. an argument against transformation). Yet, closer inspection of these latter studies reveals serious limitations. Firstly, a number of them measured how well participants were able to replicate the rhythmic pattern itself and therefore do not offer a direct measure of transformation (Povel and Essens, 1985; Grahn and Brett, 2007; Chen et al., 2008b; Cameron and Grahn, 2014). A different study instructed participants to tap a metric pulse they perceived in a rhythmic stimulus, however the stimuli were presented for very short durations (only ~5 s of rhythm followed by ~2-s long pause) (Kung et al., 2013). Such a protocol is not ecologically plausible, as meter perception may need time to develop,

especially when the input has little contrast at metric periodicities (i.e. when prominent transformation is required). In fact, studies using rhythmic stimuli with much longer durations report participants taking at least 5 s to start tapping the pulse (Chapin et al., 2010; Tal et al., 2017), whereas 5 s was the approximate duration of entire stimuli used by Kung et al. (2013). Another study used longer rhythmic sequences (Fitch and Rosenfeld, 2007), however, their results are difficult to interpret for two reasons. Firstly, they did not use a valid measure of relative contrast, and instead evaluated whether participants tapped with a specific alignment relative to a target pulse (without quantifying the consistency of this alignment, see section 1.2.2.6). Moreover, they averaged data over participants before assessing the relationship between the tapping and the stimulus. Therefore, the variability between participants was not taken into account, and the results could have been driven by a small number of outliers.

Together, these considerations speak in favor of studies concluding that meter perception indeed involves a transformation from sound to behavior, such that the contrast at meter periodicities is selectively enhanced.

1.3.2 Transformation in the brain

Using frequency tagging combined with EEG, it has been shown that when humans listen to different rhythmic inputs that give rise to the perception of metric pulses, the elicited neural response is selectively enhanced at the frequencies related to the periods of these metric pulses (Nozaradan et al., 2012, 2016a, 2017b, 2018; Tal et al., 2017; Lenc et al., 2018). Importantly, this enhancement is observed even when meter-related frequencies are not prominent in the input, i.e. when the sound input has little acoustic contrast at meter periodicities. In other words, the rhythmic input is transformed within the brain towards the *behaviourally-relevant category* (Windsor, 1993). This transformation is sensitive to the temporal limits of pulse perception. When the rhythmic input is speeded up, frequencies related to slower metric pulses are selectively enhanced in the neural response, and this is linked to spontaneous behaviour as measured with tapping (Nozaradan et al., 2012, 2018). The enhancement of contrast at metric periodicities seems to emerge gradually along the auditory pathway, particularly for inputs that lack prominent contrast at meter periodicities in their acoustic structure. For such inputs, little contrast enhancement is observed in

responses elicited predominantly in early processing stages (including subcortical nuclei) (Nozaradan et al., 2018), however, a prominent transformation is evident already in the auditory cortex, as observed in human intracerebral recordings (Nozaradan et al., 2016a). While intracerebral recordings provide excellent spatial resolution, the response recorded from a particular location can often be considered a result of widespread interactions between regions connected within a functional network. Indeed, data from patients with focal brain damage indicate that subcortical regions such as basal ganglia and cerebellum specifically contribute to the transformation measured with surface EEG (Nozaradan et al., 2017b).

Besides investigating the functional anatomy of the transformation, another important aspect is its flexibility. One extreme view is that the mapping of rhythmic inputs onto an internally represented set of metric pulses is a result of biophysical constraints and more or less fixed neurophysiological mechanisms. This view is closely related to modeling efforts that aim to predict the mapping of an arbitrary acoustic input onto behavior by solely considering its acoustic features (Longuet-Higgins and Lee, 1984; Povel and Essens, 1985; Palmer and Krumhansl, 1990; Parncutt, 1994; van Noorden and Moelants, 1999; Large, 2000a; Eck, 2003; Toiviainen and Snyder, 2003; Hannon et al., 2004; Todd and Lee, 2015a; Bouwer et al., 2018). Similar assumptions have been present in neuroscientific literature, aiming to describe the transformation in the neural response in terms of interaction between stimulus features and low-level physiological mechanisms (Large and Snyder, 2009; Large et al., 2015; Rajendran et al., 2017, 2020). This approach assumes that once the relevant mechanisms have been described, they can be generalized across rhythmic inputs, contexts, individuals, and cultures. In other words, the approach assumes universals. Accordingly, all we need to know is the physical features of the input, and we should be able to predict its mapping onto brain activity and behavior.

On the other hand, the transformation involved in meter processing may be flexible, shaped by a number of factors including culture, context, exposure, training, and current internal state of the listener such as behavioral goals and attentional focus. There are multiple lines of evidence in favor of this view.

Firstly, there seems to be flexibility due to long-term exposure and learning. This is showcased by observing that the same acoustic input can be transformed towards different

metric pulses across and within individuals depending on their long-term experience (Hannon and Trehub, 2005a; Hannon et al., 2011, 2012b; Cameron et al., 2015; London et al., 2017; van der Weij et al., 2017). This prior experience may be inherently multimodal, involving simultaneous auditory, visual, and vestibular stimulation (Phillips-Silver and Trainor, 2005, 2008). In particular, mappings between specific rhythmic inputs and internal representation of pulses may be learned via observing and actively participating in music making and dance, which is a common practice across cultures (Savage et al., 2015). Indeed, it has been shown using frequency-tagging that prior movement to a rhythmic input can significantly bias which periodic contrasts are enhanced in the neural response to the same input during subsequent listening (Chemin et al., 2014). These long-term exposure effects may be closely linked to associations between a musical idiom and meter with particular parameters. For instance, recognizing timbres, rhythmic patterns, or melodic motifs idiomatic for a particular genre may bias transformation of the input towards a specific metrical category (London et al., 2017; van der Weij et al., 2017). In fact, similar phenomena may be ubiquitous in music perception beyond rhythm-meter mapping (Honing and Ladinig, 2009).

At the same time humans seem to show flexibility on much shorter timescales. For instance, periods and phase parameters of the meter perceived in a rhythmic input can be changed voluntarily, and in fact, musicians often intentionally practice this flexibility through exercises (Greb, 2017; Guilianna, 2018). When a metric pulse with a specific period is intentionally perceived, this leads to increased contrast at the pulse periodicity in the neural response, and this observation has been replicated in a large number of studies using frequency-tagging (Nozaradan et al., 2011; Celma-Miralles et al., 2016; Okawa et al., 2017; Li et al., 2019), as well as time-domain approaches to measure contrast in brain responses (Iversen et al., 2009; Fujioka et al., 2010, 2015; Schaefer et al., 2011).

Not only top-down intention, but also directly preceding acoustic context could influence the transformation of the rhythmic input. Theoretical writings commonly mention the idea that meter perception has a tendency to persist over time once induced by a particular rhythmic input with prominent contrast at meter periodicities (Lerdahl and Jackendoff, 1983; Large and Palmer, 2002; London, 2004; Large and Snyder, 2009; Honing and Bouwer, 2018). Yet, empirical evidence for this remains largely anecdotal, as most previous work has

focused on the short-term context effect in interval timing (Desain and Honing, 2003; McAuley and Jones, 2003).

To summarize, there is mounting evidence that meter perception involves transformation of sound inputs towards periodic pulses in the brain and behavior. However, the nature of this transformation remains largely unknown. While there is evidence suggesting flexibility of the mapping between sound input and internal representation of meter, the role of specific endogenous or exogenous factors remains unclear. In the empirical part of this thesis, I used the frequency-tagging approach to shed light on the nature of this transformation. To this end, I conducted a series of experiments where I manipulated the attentional state of the listener, spectral content of the sound input, and recent acoustic context, to investigate whether these factors can bias the selective enhancement of contrast at meter periodicities in neural responses to rhythm. While the main goal was to explore the sound-brain mapping, the analyses did not simply abstract from behavior. Instead the transformation in the brain was always analyzed in the context of the transformation in behavior (captured in the form of finger tapping). Thus the current results contribute to the larger aim of comprehensively describing the nature of the transformation between sound, brain, and behavior, which takes place when humans listen and move along with complex rhythmic sounds. It is important to note that the main goal of the current thesis was not to build a comprehensive theory of meter perception, nor a formal or informal model of the phenomenon. Instead, I took a rudimentary position of only focusing on explanatory processes based on empirical evidence. Accordingly, the aim was to elucidate the internal processes underlying meter perception by addressing the nature of the phenomenon in a series of empirical studies.

2 Study 1: Attention affects overall gain but not selective contrast at meter frequencies in the neural processing of rhythm

In the first study, I investigated whether changing the internal state of the listener affects the selective contrast at meter periodicities in the neural response to rhythm. To this end, the attentional focus of participants was manipulated using demanding tasks, while they were presented with auditory rhythms. The results showed that meter periodicities are robustly enhanced even for rhythmic inputs that lack prominent contrast at these periodicities in their acoustic structure. Importantly, it was observed that this enhancement was present even when overall neural responsiveness to the sound was attenuated due to a distracting task. These results suggest that the neural mechanisms involved in transforming the rhythmic input towards a metric category might be, to a certain extent, engaged quite automatically in different behavioural contexts.

2.1 Abstract

When listening to music, humans spontaneously perceive and synchronize movement to periodic pulses of meter. A growing body of evidence suggests that this widespread ability is related to neural processes that selectively enhance meter periodicities. However, to what extent these neural processes are affected by the attentional state of the listener remains largely unknown. Here, we recorded EEG while participants listened to auditory rhythms and detected small changes in tempo or pitch of the stimulus, or performed a visual task. The overall neural response to the auditory input decreased when participants attended the visual modality, indicating generally lower sensitivity to acoustic information. However, the selective contrast at meter periodicities did not differ across the three tasks. Moreover, this selective contrast could be trivially accounted for by biologically-plausible models of subcortical auditory processing, but only when meter periodicities were already prominent in the acoustic input. However, when meter periodicities were not prominent in the auditory input, the EEG responses could not be explained by low-level processing. This was also confirmed by early auditory responses that originate predominantly in early auditory areas and were recorded in the same EEG. The contrast at meter periodicities in these early responses was consistently smaller than in the EEG responses originating mainly from higher-level processing stages. Together, these results demonstrate that selective contrast at meter periodicities involves higher-level neural processes that may be engaged automatically, irrespective of behavioral context. This robust shaping of the neural

representation of rhythm might thus contribute to spontaneous and effortless synchronization to musical meter in humans across cultures.

2.2 Introduction

Perception of rhythmic sound sequences involves much more than just a precise representation of constituent time intervals. Already perception of single intervals is not one-to-one with respect to the sensory input, but reflects a representation constructed with respect to prior individual experience (Desain and Honing, 2003; Jazayeri and Shadlen, 2010; Jacoby and McDermott, 2017). An even higher level of perceptual organization is arguably at stake when the rhythmic input induces perception of musical meter, i.e., a nested set of periodic pulses to which people tend to move or dance (Cohn, 2020). That is, the internal representation of meter guides perceptual organization of the incoming rhythmic sequence in time (Povel and Essens, 1985; McAuley and Jones, 2003) and drives body movement such as head bobbing or foot tapping (Toiviainen et al., 2010; Janata et al., 2012). Perception and sensory-motor synchronization to meter is a spontaneous human ability that has been widely observed across cultures and musical traditions (Nettl, 2000; Savage et al., 2015).

In some cases, meter perception can be largely driven by the acoustic features of the sensory input, particularly when clear periodicities are present in the temporal structure of the stimulus (although even in such cases the alignment of the perceived pulses with the input is not trivial, see e.g. off-beat rhythm in reggae). However, meter perception is often induced by stimuli that lack unambiguous acoustic cues to meter periodicities (Chapin et al., 2010; Nozaradan et al., 2012; Witek et al., 2014b; Large et al., 2015; London et al., 2017; Vuust et al., 2018; Matthews et al., 2020), and the same rhythmic sequence can be perceptually organized in different ways depending on prior experience at multiple timescales (Phillips-Silver and Trainor, 2005; Hannon et al., 2012a; Chemin et al., 2014; van der Weij et al., 2017). This shows that meter perception goes beyond the mere tracking of periodicities in the sensory input, and additionally involves higher-level processes that transform the input towards a particular metric category with a great degree of robustness and flexibility with respect to the input (Nozaradan et al., 2017a).

This is in line with a number of recent neurophysiological studies based on the assumption that meter perception is related to neural processes that emphasize the contrast between

time points marked by the perceived metric pulses and other time points not marked by pulses. This contrast in the neural response can be driven already by the physical features of the sensory input along with a set of low-level nonlinear transformations throughout early auditory processing stages (Rajendran et al., 2017, 2020). Importantly there is also increasing evidence for higher-level neural processes that transform the input by selectively enhancing this contrast beyond physical features and low-level nonlinearities (Lenc et al., 2018, 2020). These higher-level neural processes may thus play a key role in building internal representation of meter dissociated from the physical features of the sensory input (Nozaradan et al., 2011, 2012, 2017a, 2017b; Tal et al., 2017) .

However, to what extent these processes are engaged automatically, and whether they depend on the behavioral goals of the listener remains largely unknown. Previous neurophysiological and neuroimaging studies of meter processing in humans have employed a wide range of behavioral tasks, some instructing participants to attend directly to the pulse-like metric structure of the stimuli (Grahn and Rowe, 2009, 2013; Lenc et al., 2020; Matthews et al., 2020) or the temporal properties of the stimulus (Nozaradan et al., 2017b; Lenc et al., 2018), while other studies used an orthogonal task such as attending to a non-temporal sound feature (e.g. pitch; Haumann et al., 2018) or attending to a different modality (e.g. visual; Chapin et al., 2010) or no task at all (Bengtsson et al., 2009). However, how neural processing of a rhythmic input changes across these different tasks has not been systematically explored using a consistent set of stimuli and analysis methods.

Additionally, in a series of studies investigating putative “pre-attentive beat perception” using event-related brain response to regularity violations, participants were typically asked to perform a passive task, such as watching a silent movie, while listening to the rhythmic stimuli (Vuust et al., 2005; Ladinig et al., 2009; Geiser et al., 2010; Bouwer et al., 2014, 2016). The lack of strict control of participant’s attentional focus combined with the low load of the task make the results of these studies difficult to interpret (Lavie and Dalton, 2014; Sussman et al., 2014; Murphy et al., 2017). In addition, these studies mostly used stimuli with clear acoustic cues to meter periodicities. Therefore, it remains unknown whether these results would generalize to rhythmic inputs that lack such prominent sensory cues and may thus require higher-level processes to induce meter perception (Chapin et al., 2010; Nozaradan et al., 2011).

In the current study, we aimed to address these issues by recording human brain electroencephalographic (EEG) activity in response to (i) a consistent set of rhythmic stimuli with varying amounts of sensory cues to meter periodicities, along with (ii) a set of three demanding behavioral tasks in the same sample of participants. We presented participants with two rhythmic sequences. One sequence contained prominent acoustic cues to meter periodicities, while the other sequence lacked such prominent periodic cues. This latter sequence enabled us to control for a low-level confound which could trivially explain enhanced neural response at meter periodicities. That is, if selective contrast at meter periodicities is observed in the EEG in response to a sequence lacking such prominent periodic cues, this selective contrast at meter periodicities cannot be explained easily by the stimulus structure or low-level processing of the stimulus. Importantly, the decision as to what frequencies would correspond to meter periodicities was informed by previous studies, which used tapping tasks to carefully test the metric pulses most consistently induced by these two rhythmic patterns across listeners (Nozaradan et al., 2012, 2018; Lenc et al., 2018). This ensured that these specific frequencies were relevant for meter perception, in contrast to other frequencies that are also elicited by the rhythms but are irrelevant to the perceived meter.

Participants listened to the rhythms while performing three different demanding tasks. In the first task, participants were required to detect small changes in the speed of the rhythmic sequence. Because the sequence was non-isochronous, this task cannot be carried out by simply comparing successive inter-tone intervals and therefore encourages participants to build an internal representation of meter that aids tracking of the overall speed of the rhythm (Schulze, 1978; Grube and Griffiths, 2009; Grube et al., 2010). In the second task, participants were required to detect small changes in the pitch of a single tone among the rhythmic sequences, thus still focusing on the sound but not necessarily on its timing. Finally, in the third task, participants were required to mentally sum numbers sequentially presented on the screen while ignoring the sounds altogether.

The EEG was recorded while participants were presented with the auditory sequences and carried out the behavioral tasks without any movement. From the EEG, we measured the difference in amplitude of the neural activity at meter-related frequencies vs. meter-unrelated frequencies elicited by the rhythms, i.e., the contrast at perceptually-relevant timescales, using frequency tagging. This approach has proven to be a powerful tool for

capturing the contrast in brain responses between periodically spaced time points with high signal-to-noise ratio and without assumptions about the latency or the shape of the response (Nozaradan, 2014; Rossion, 2014; Norcia et al., 2015; Nozaradan et al., 2017a; Rossion et al., 2020). Moreover, the approach also allows the overall gain of the response (i.e. the general sensitivity to auditory stimulation) to be disentangled from the selective contrast at meter-relevant periodicities.

We also examined whether the contrast at meter frequencies in the EEG activity elicited across behavioral tasks could be trivially accounted for by fixed nonlinear transformations along the early auditory pathway. To this end, we used biologically plausible models to simulate responses to the rhythmic stimuli in the auditory nerve, as well as inferior colliculus. To complement these simulations, we also directly captured responses presumed to be predominantly driven by brainstem auditory nuclei and primary auditory areas using the same frequency-tagging method as Nozaradan et al. (2016c, 2018). These early responses were observed at a faster timescale (> 150 Hz) due to neural tracking of the amplitude-modulated fine structure of the sound input. By comparing these early auditory responses to the higher-level responses observed at slower timescales (< 5 Hz, corresponding to the amplitude envelope of the input), which mainly capture activity in higher-level cortical networks, we aimed to estimate the contribution of different processing stages to the selective contrast at meter frequencies across different attentional contexts.

2.3 Materials and methods

2.3.1 Participants

Seventeen healthy volunteers (mean age = 23.3, SD = 6.7, 15 females) with various levels of formal musical training (mean = 2.5, SD = 4.6, range = 0-16 years) participated in the study after providing written informed consent. All participants reported normal hearing and no history of neurological or psychiatric disorder. The study was approved by the Research Ethics Committee of Western Sydney University.

2.3.2 Auditory stimuli

The auditory stimuli were created in Matlab R2016b (The MathWorks, Natick, MA) and presented binaurally through insert earphones (ER-2; Etymotic Research, Elk Grove Village, IL) at a comfortable listening level (~ 75 dB SPL) using PsychToolbox, version 3.0.14 (Brainard, 1997) running on a MacBook Pro laptop (mid-2015, OSX 10.12). Triggers were sent to the EEG system using LabJack U3 interface. The stimuli consisted of a 2.4-s long rhythmic pattern (made up of twelve 200-ms long events) continuously looped 14 times to create a 33.6-s long sequence. The rhythmic structure of the pattern was based on a specific arrangement of 8 sound events and 4 silent events (amplitude at 0). Each sound event corresponded to a complex tone consisting of three partials ($f_1 = 209$ Hz, $f_2 = 398$ Hz, $f_3 = 566$ Hz) with linear onset and offset ramp lasting 10% of the event duration (i.e. 20 ms).

We used two different rhythmic patterns (depicted in Figure 2.1). These two patterns were selected based on previous evidence that they both induce a perception of musical meter, consistent across individuals, based on nested grouping of the individual event rate (200 ms) by 2 (2×200 ms = 400 ms), 2 (2×400 ms = 800 ms) and 3 (3×800 ms = 2400 ms) (Nozaradan et al., 2012, 2018; Lenc et al., 2018).

Importantly, although the two rhythmic patterns induce perception of musical meter at consistent periods across individuals, they provide the listener with different amounts of direct sensory cues to this perceived metric structure. One way to quantify this is to examine the degree of mismatch between the perceived meter and the arrangement of sound events in the rhythm using syncopation scores. Even though different ways to calculate syncopation scores have been proposed, the main principle they share is quantifying to what extent the rhythmic stimulus creates a contrast between time points that coincide with the putative metric pulses, and the rest of the time points, i.e. a contrast at meter periodicities (Longuet-Higgins and Lee, 1984; Povel and Essens, 1985; Parncutt, 1994; Eck, 2003). We calculated syncopation scores for the two rhythmic patterns using an algorithm originally proposed by Longuet-Higgins and Lee, which simultaneously takes into account the whole nested hierarchy of metric pulses (Longuet-Higgins and Lee, 1984; Witek et al., 2014b). Additionally, a C score (counterevidence) was calculated using the method and parameters proposed by Povel and Essens (1985). While C score calculates syncopation using only one pulse in the metric structure, it accounts for variable perceptual salience of

tones making up the pattern based on their relative temporal proximity (Povel and Okkerman, 1981).

Even though the periods of the perceived metric pulses for the two rhythms are generally consistent across participants, the alignment of these pulses with respect to the rhythmic stimulus can vary (Nozaradan et al., 2012, 2018; Lenc et al., 2018). To avoid assumptions regarding particular pulse alignment, the minimum syncopation and C score across all 12 possible positions of the slowest metric pulse with respect to the rhythm was taken (see also Lenc et al., 2020). This yielded smaller scores for one rhythm (syncopation = 1, C = 1), in comparison to the other rhythm (syncopation = 2, C = 2). In other words, both measures revealed a greater mismatch between the perceived meter and the arrangement of sound events for the second rhythm.

This reflects the fact that the physical structure of the first rhythm provides clear and unambiguous information about the perceived meter. On the other hand, the second rhythm provides less sensory information about the metric periodicities (there is no plausible alignment of the perceived pulses that would lead to systematic match with the distribution of sound onsets in the pattern). For these reasons, the first and the second rhythm are further referred to as "high meter contrast" and "low meter contrast" rhythm, respectively (note that various terms have been previously used to describe these same rhythms, e.g. unsyncopated and syncopated, Nozaradan et al., 2016b, 2017b, 2018). Despite these differences, both rhythms consistently induce meter perception across listeners, as revealed by previous studies (Nozaradan et al., 2012, 2018; Lenc et al., 2018).

2.3.3 Frequency-tagging analysis

Another way to measure the amount contrast at meter periodicities is to directly analyze the modulation spectrum of the acoustic stimulus using Fourier transform. This allows quantification of the extent to which the continuous modulation of acoustic features of the input (here amplitude envelope) emphasizes particular periodicities.

Because the stimulus sequence consisted of seamless repetitions of the same rhythmic pattern, the modulation spectra were expected to contain energy at frequencies corresponding to the repetition of the pattern ($1/2.4\text{ s} = 0.416\text{ Hz}$) and harmonics. The relative distribution of energy across these different harmonics reveals how much contrast

was present in the signal modulations at the corresponding frequencies. From the set of first 12 harmonics (up to 5 Hz, the frequency of individual event rate in the rhythms), four frequencies were considered meter-related (0.416, 1.25, 2.5, 5 Hz), as they corresponded to the frequencies of the perceived metric pulses (1/2.4 s, 1/0.8 s, 1/0.4 s, 1/0.2 s respectively). The remaining 8 frequencies in the set were considered meter-unrelated.

To measure the relative prominence of meter frequencies, amplitudes at the 12 frequencies corresponding to the stimulus modulation spectrum were converted to z-scores as follows: $([x] - [\text{mean across the 12 frequencies}]) / [\text{SD across the 12 frequencies}]$. A higher z-score at a specific frequency indicates that the response at that frequency stands out prominently relative to the whole set of frequencies in the modulation spectrum. The z-scores for meter-related frequencies were averaged to obtain an index of their relative prominence in the modulation spectra.

The main advantage of using FFT is that it can be applied to a variety of signals representing (i) modulations in the acoustic input, (ii) simulated responses of neurons in the subcortical auditory nuclei, (iii) surface EEG, and (iv) movement. Importantly, using the z-scoring standardization yields a measure invariant to differences in unit and scale, thus allowing for objective measurement of the relative distance between these different signals. In sum, this method represents a powerful tool to track the transformation of the input, i.e. the changes in contrast at meter periodicities across different processing stages from input to output.

2.3.4 Models of subcortical auditory processing

To estimate to what extent the neural transformation of a rhythmic acoustic stimulus could be driven by early stages of the auditory pathway, we simulated responses to the rhythmic stimuli using multiple biologically-plausible models of subcortical auditory processing, as described below. Comparing the EEG responses to these early representations thus helps to disentangle the contribution of higher-level transformations that cannot be trivially explained by early sound processing stages.

(i) *Broadband envelope*. A number of previous EEG studies used broadband envelopes to represent modulations in the acoustic input (Aiken and Picton, 2008; Nozaradan et al., 2012, 2018; Chemin et al., 2014; Cirelli et al., 2016; Tal et al., 2017; Broderick et al., 2019; Di Liberto et al., 2020b). To provide a point of comparison with these studies, the broadband

amplitude envelope of the 33.6-s auditory sequences (high and low meter contrast rhythm) was extracted using the Hilbert transform (as implemented in Matlab) and then transformed into the frequency domain using a fast Fourier transform (FFT, yielding a spectral resolution of $1/33.6$ s, i.e. approximately 0.03 Hz).

(*iii*) *UR-EAR-AN*. The model of the auditory nerve developed by Bruce et al. (2018) as implemented in UR_EAR toolbox (version 2020a) was used to simulate responses from 128 cochlear channels with characteristic frequencies logarithmically spaced between 130 and 8000 Hz. The parameters used for cochlear tuning matched data available from human subjects (Shera et al., 2002). For each channel, 51 auditory nerve fibers were simulated with biologically plausible distribution of high, mid, and low-spontaneous-rate fibers (Liberman, 1978). The model provides faithful simulation of physiological processes associated with cochlear nonlinearities, inner hair cell transduction process, the synapse between the hair cell and the auditory nerve, and the associated firing rate adaptation.

(*ii*) *UR-EAR-IC*. The simulated auditory nerve firing rates were fed into the same-frequency inhibition and excitation model (SFIE) used to simulate enhanced onset synchrony and the decreased upper limit for phase-locking to stimulus envelope in the ventral cochlear nucleus (Nelson and Carney, 2004). The default parameters in the UR_EAR toolbox were used, which were based on Carney et al. (2015). A second SFIE model was then used to simulate band-pass modulation filtering and enhanced onset responses of neurons in the inferior colliculus (IC). The parameters were set to simulate IC units with the best modulation frequencies separately at 2, 4, 8, 16, 32, and 64 Hz.

For both the AN and IC stage of the UR_EAR model, the simulated instantaneous firing rates were summed across cochlear channels (Zuk et al., 2018; Rajendran et al., 2020), and transformed into the frequency domain using FFT. While averaging firing rates across channels might yield different results than averaging FFT magnitudes for spectrally complex inputs, the two methods should give very similar results for the stimuli in the current study, as the modulation waveform was identical across the whole spectrum. Subsequently, amplitudes at the 12 frequencies of interest were extracted from the obtained spectra and normalized by z-scoring separately for each model output (see section 2.3.3).

2.3.5 Early auditory responses

The frequencies of the partials of the complex tones delivering the rhythm ($f_1 = 209$ Hz, $f_2 = 398$ Hz, $f_3 = 566$ Hz) were selected because sustained frequency-following responses at these frequencies are expected to originate predominantly from sub-cortical auditory nuclei due to low-pass characteristics of the ascending auditory pathway (Chandrasekaran and Kraus, 2010; Skoe and Kraus, 2010; but see Coffey et al., 2016, 2019, who show that a portion of this response could also be explained by activity from early cortical stages). Non-harmonic spacing of the partials was used in the current study as it is expected to elicit responses at frequencies that are not physically present in the stimulus spectrum. These frequencies corresponded to distortion-product otoacoustic emissions generated by nonlinear processes at the cochlear level and transmitted along the ascending auditory pathway (Lee et al., 2009). Hence, any EEG response at these frequencies could not be explained by an electromagnetic artifact from the sound-delivery system. These responses were expected at frequencies corresponding to quadratic distortion products across the three partials, i.e. $f_2 - f_1$ (168 Hz), $f_3 - f_2$ (189 Hz), and $f_3 - f_1$ (357 Hz). Due to the frequency-shifting theorem, each of the distortion-product frequencies was expected to be symmetrically flanked by sidebands representing the amplitude modulation spectrum of the response (Oppenheim and Schaffer, 2009). This allowed the contrast at meter frequencies to be quantified at earlier auditory processing stages with the same method as described above for the sound input (see section 2.3.3). Furthermore, this contrast at meter frequencies obtained from earlier auditory stages was also compared to the contrast at meter frequencies obtained from EEG responses measured in a much lower frequency range (here at 5 Hz and below) and assumed to predominantly originate from higher-level processing stages (further referred to as "higher-level" responses) (Nozaradan et al., 2018). Importantly, because the index of contrast at meter frequencies consists in a relative measure of the amplitude at meter frequencies vs. meter-unrelated frequencies obtained after z-scoring standardization, this measure is invariant to differences in unit and scale, thus providing valid estimation of the relative distance between signals as different as the early auditory responses and higher-level responses, irrespective of differences in overall gain.

2.3.6 Experimental design and procedure

Participants were presented with the rhythmic auditory stimuli in separate blocks of 10 self-paced trials. The polarity of the acoustic waveform was inverted on every other trial to prevent potential electromagnetic artifact at the frequencies of the sound input (Skoie and Kraus, 2010). In each block, participants were asked to perform a specific behavioral task.

Tempo task. The block contained two additional randomly-placed trials where one rhythm cycle (at a random position after the first 3 cycles) contained a decrease in tempo. This was implemented by gradually increasing (and then decreasing) the inter-onset intervals of the individual constituent events within one rhythm cycle according to a cosine window from the standard inter-onset interval (200 ms) to the maximum interval determined individually for each participant. Participants were asked to focus on the tempo of the stimuli, while ignoring all other parameters, as well as any visually presented stimuli. They reported whether the change was present after the end of each trial.

Pitch task. The block contained two additional trials with increased pitch of a single constituent tone (implemented as a proportional increase in the frequency of each partial). Participants were asked to report the presence of the pitch change at the end of each trial, while ignoring other sound parameters and visual stimuli.

Visual stimuli and task. Throughout all trials and blocks, participants also viewed sequentially presented numbers in the center of the screen positioned in front of them (approximately one meter distance). The numbers were randomly sampled such that the first number for each trial was between 100 and 200, and all subsequent numbers were between 10 and 30. The time interval between the onset of each sequential number was individually determined for each participant, and a jitter of 10% of this time interval was then applied to avoid any strict periodicity in the visual presentation of the numbers, which could result in a narrow frequency peak elicited in the EEG spectrum at the frequency of the visual presentation. Each number stayed on the screen for 80% of the inter-onset interval (with 10% random jitter applied to this value). Participants were asked to fixate their eyes on the numbers in every trial across all blocks in order to prevent eye movements. During the Visual task, they were asked to mentally add these numbers and report the sum at the end of each trial, while ignoring the sound stimuli. Participants were instructed to keep adding the incoming numbers even in case they missed any. This was to make sure

participants did not “give up” in the middle of the trial, but kept continuously engaged with the visual task.

Each task and rhythm were presented as a separate block, yielding $3 \times 2 = 6$ blocks in the whole EEG session (block order was counterbalanced across participants). Participants were seated in a comfortable chair and asked to avoid any unnecessary movement or muscle tension. For each block, the two trials containing tempo or pitch changes were excluded from the EEG analyses, thus leaving 10 trials per task and rhythm for subsequent analysis.

Before the EEG session, the parameters for the three tasks were individually adjusted for each participant using a two-down, one-up staircase method, targeting 70.7% accuracy in all tasks (Leek, 2001), separately for the high meter contrast and low meter contrast rhythm. This individual adjustment aimed to make each block equally demanding for the EEG session. These additional trials performed before the EEG session to determine individual parameters also allowed participants to familiarize themselves with the nature of the tasks. For the Pitch and Tempo tasks, the staircase procedure contained a single run where the rhythmic pattern was seamlessly cycled and deviants appeared randomly, separated by at least one intact pattern cycle. Participants were instructed to press a button as soon as they detected a deviant. Button presses within 1 second were considered hits, otherwise the response was considered a miss. The procedure finished after 6 reversals. The threshold was determined as the average deviant magnitude at the last 4 reversals. For the Visual task, participants were asked to mentally sum 5 sequentially presented numbers with the same parameters as in the EEG session. This was done in discrete trials, and the mean inter-stimulus interval for each trial was adjusted according to the correctness of participant’s response on the previous trial. The procedure finished after 6 reversals (threshold estimated as the average deviant magnitude at the last 4 reversals), or after 20 trials (threshold taken as the mean of any available reversals, or the value from the last trial).

After the EEG session, participants rated the subjective difficulty of each task on a discrete scale from 1 (easy) to 7 (difficult).

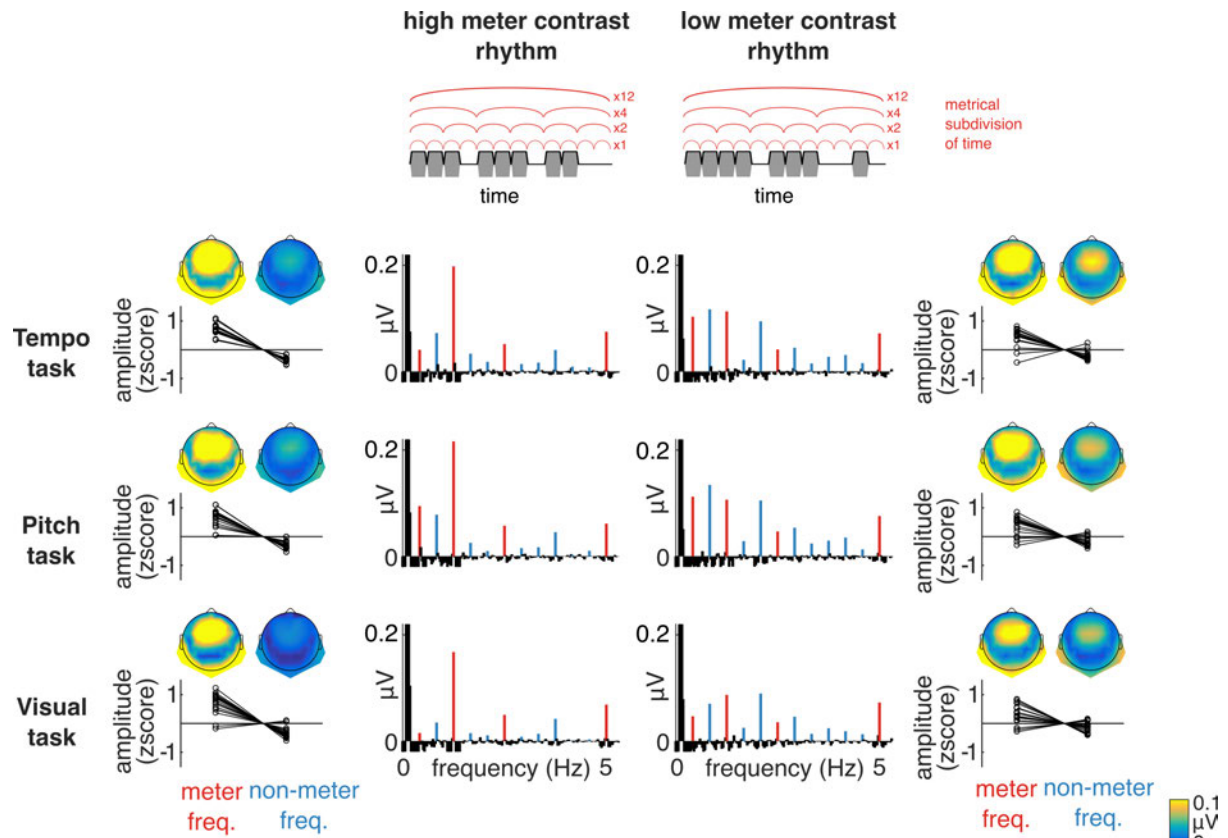


Figure 2.1. Stimulus design and higher-level EEG responses. (Top) The sound waveform representing one cycle of the high meter contrast (Left) and low meter contrast (Right) rhythmic pattern is depicted in grey. The broadband envelope is overlaid as a black line. Above each pattern, the meter typically induced by these patterns is shown as red arches representing individual pulses in the metric structure. (Bottom) Spectra of higher-level EEG responses elicited for each rhythm and task (average across all participants and EEG channels). Mean z-scored amplitude elicited at meter-related (red) and meter-unrelated (blue) frequencies is shown next to the corresponding spectra (data points represent individual participants), along with the topographical distribution of mean EEG amplitude at these two subsets of frequencies (average across all participants).

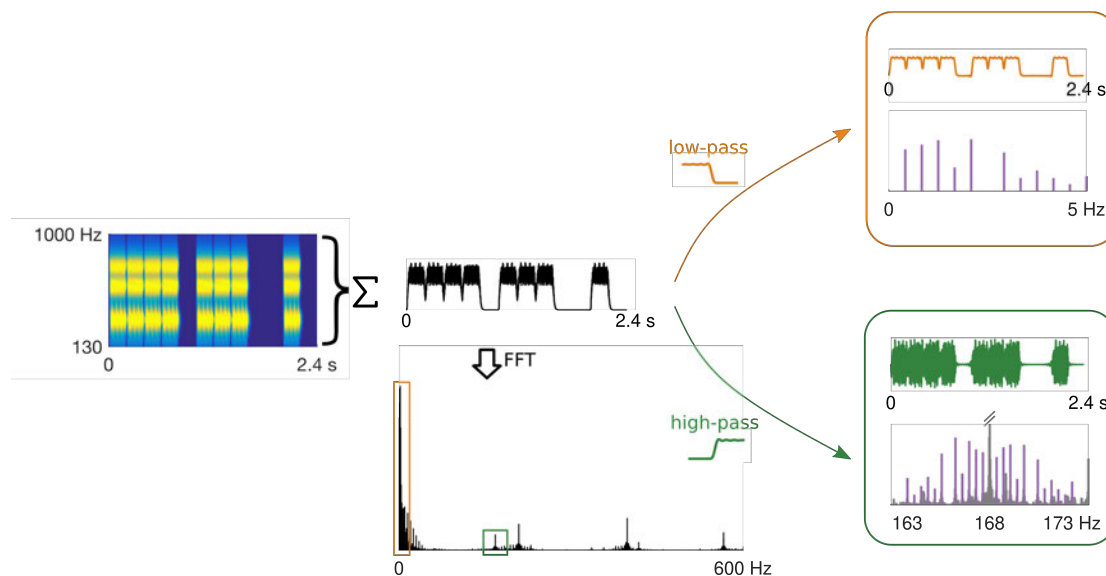


Figure 2.2. Diagram showing dissociation between higher-level and early auditory EEG responses. Cochleogram on the left shows a response to one cycle of the low meter contrast rhythm across a population of cochlear channels tuned to different frequencies (obtained using the model of Bruce et al., 2018). Summing the peristimulus time histogram across all cochlear channels yields a composite response (shown in black in the middle). The FFT of this composite response (shown on the bottom) reveals how the whole population tracks amplitude envelope modulations (concentrated in the low frequency portion of the spectrum, i.e. at the exact amplitude modulation frequencies), but also phase-locks to the fine structure of the sound input (higher frequency range in the spectrum, at the actual frequencies of the partials and distortion products). Because the fine structure is itself amplitude modulated, the spectrum of the modulator (i.e. amplitude envelope) is reflected in symmetrical sidebands surrounding each partial and distortion product frequency, in line with the shifting theorem of the Fourier Transform (Oppenheim and Schaffer, 2009). Thus, the two responses can be separated in the frequency domain by zooming onto the relevant portions of the spectrum, as depicted by the orange rectangle (for the higher-level response) and the green rectangle (for the early auditory response at 168-Hz distortion product). To isolate the higher-level response in the time domain (orange waveform, top right), low-pass filter can be applied to the peristimulus time histogram, which is assumed to take place along the auditory pathway (Chandrasekaran and Kraus, 2010). The early auditory response (green waveform, bottom right) can be isolated in the time domain by high-pass filtering the peristimulus time histogram.

2.3.7 EEG recording and preprocessing

The EEG was recorded using a Biosemi Active-Two system (Biosemi, Amsterdam, Netherlands) with 64 Ag-AgCl electrodes placed on the scalp according to the international 10/20 system, and two additional electrodes attached to the mastoids. Head movements

were monitored using an accelerometer with two axes (front-back and left-right) attached to the EEG cap and recorded as 2 additional channels. The signals were digitized at 8192-Hz sampling rate, which was high enough to capture distortion-product frequencies relevant for the early auditory responses (Skoe and Kraus, 2010).

Analysis of higher-level EEG responses. Higher-level EEG responses refer to EEG activity measured in a low-frequency range (here, at 5 Hz and below), thus corresponding to the frequency range of the actual envelope modulations in the rhythmic inputs (see Figure 2.2). These responses were analyzed by first downsampling the EEG signals offline to 512 Hz. The continuous EEG signals were then high-pass filtered at 0.1 Hz (4th-order Butterworth filter) to remove slow drifts from the signals. Artifacts related to eye blinks and horizontal eye movements were identified and removed using independent component analysis (Bell and Sejnowski, 1995; Jung et al., 2000) based on visual inspection of their typical waveform shape and topographic distribution. One component was removed for 7 participants, two components for 9 participants, and 9 components for one participant (the eye-movement related activity was clearly distributed across a larger number of components for this participant). Channels containing excessive artifacts or noise were manually selected and linearly interpolated across all trials and conditions, separately for each participant (1 channel for 3 participants, 2 channels for 1 participant). The data were then segmented into 33.6-s long epochs, starting from the onset of the sound sequence in each trial and re-referenced to the average of the 66 channels. The mastoid channels were included because they were expected to prominently capture the responses to auditory rhythms based on previous studies (Nozaradan et al., 2012, 2016b; Lenc et al., 2018, 2020, see also Figure 2.1). This was indeed the case, as revealed by the topographical distributions shown in Figure 2.1. Thus including the mastoid electrodes would enhance the overall signal-to-noise ratio (SNR) of the EEG spectra after averaging across all channels (see below).

Analysis of early auditory EEG responses. Early auditory EEG responses refer to EEG activity measured in a much higher frequency range than the higher-level responses (> 150 Hz), thus corresponding to the frequency range of the actual partials conveying the envelope modulations of the rhythmic inputs (see Figure 2.2). Here, preprocessing did not include independent component analysis and channel interpolation. The data at the original sampling rate (8192 Hz) were re-referenced to the average of mastoid electrodes, and only signals from three fronto-central channels (Fz, FCz, Cz) were kept for further analyses. Based

on previous studies, this standard montage was expected to most strongly capture the auditory frequency-following responses (Skoe and Kraus, 2010; Nozaradan et al., 2016c, 2018).

The preprocessed data were averaged in the time domain across the 10 trials separately for each participant and condition. Time-domain averaging was performed to increase the signal-to-noise ratio by cancelling signals that were not time-locked to the stimulus, while preserving evoked responses elicited by the stimulus, as well as any ongoing activity entrained by the stimulus, which were both assumed to be stationary across trials.

The averaged signals were transformed into the frequency domain using FFT. The obtained spectra were considered to consist of (i) activity elicited by the auditory stimulus, concentrated within narrow peaks and (ii) residual background noise smoothly distributed across a broad range of frequencies (Mouraux et al., 2011; Retter and Rossion, 2016). The contribution of broadband noise was therefore minimized by subtracting the average amplitude at neighboring bins on both sides relative to each frequency bin (bins 2-5 for the higher-level responses and 3-10 for the early auditory responses). A narrower range of bins used for the higher-level responses was to avoid bias in the noise estimate due to prominent $1/f$ in the lower part of the EEG spectrum. For the early auditory responses, two (instead of one) directly adjacent bins were excluded from the noise estimate due to potential FFT leakage of the response (as the tagged frequencies were not exactly centered on a single frequency bin).

The noise-subtracted spectra were averaged across all channels (66 channels for higher-level responses, 3 channels for early auditory responses) separately for each condition and participant. The magnitudes of higher-level responses were extracted from the bins centered at the 12 frequencies expected based on the stimulus modulation spectrum (see section 2.3.3). Magnitudes of the early auditory responses were estimated at the frequencies of the distortion products and their corresponding sideband frequencies (by taking the bin closest to the frequency of interest).

2.3.8 Overall EEG response magnitude

The overall magnitude of the higher-level responses was estimated as the summed amplitude across all 12 frequencies corresponding to the envelope modulation spectrum of

the stimulus (see Figure 2.3A). The same measure was taken for the early auditory responses by summing across all sideband frequencies, separately for the three distortion-product frequencies (see Figure 2.4A).

To make sure the differences in the overall response magnitude were not due to increased noise floor obscuring the sound-evoked responses, we carried out a control analysis using amplitudes from frequency-bins at positions offset by +7 (i.e. ~ 0.21 Hz) relative to the bins centered at the frequencies of interest. These were extracted from the EEG spectra obtained without any noise subtraction, and therefore provided an estimate of the broadband noise level across conditions. This control analysis was only performed for the higher-level responses, as no significant differences in overall response magnitude were found in the early auditory responses (see Results section).

Because the responses at sideband frequencies were generally small, particularly for the sidebands flanking higher distortion-product frequencies, the overall magnitude for the early auditory responses was first compared to zero separately for the sidebands flanking each distortion product to assess whether a significant response was elicited. The validity of this test relies on the fact that, because the spectra were noise-subtracted, an absence of response should result in magnitudes distributed around zero. Only responses at sidebands flanking the lowest distortion product at 168 Hz were consistently above zero (see Table 2.S1 and Figure 2.4A) across all rhythms and tasks. Therefore, all further analyses of the early auditory responses were carried out only on this distortion-product frequency and corresponding sidebands.

2.3.9 Relative EEG response at meter frequencies

To assess the relative prominence of specific frequencies in the higher-level responses, amplitudes at the 12 frequencies corresponding to the stimulus modulation spectrum were converted to z-scores, in the same way as for the models of subcortical auditory processing (see section 2.3.3). For the early auditory responses, amplitudes corresponding to the same modulation frequency were first averaged across the symmetrical positive and negative sidebands and the resulting 12 values were converted to z-scores (only for 168-Hz distortion-product frequency, see Table 2.S1 and Figure 2.4A). Higher z-score at a specific

frequency indicated that the response at that frequency stood out more prominently relative to the whole set of 12 frequencies elicited by the auditory stimulus.

The z-scores were averaged separately for the meter-related frequencies (frequencies where metric pulses are consistently perceived for these rhythms across listeners, i.e. 0.416, 1.25, 2.5, 5 Hz) and meter-unrelated frequencies (the remaining 8 frequencies in the stimulus modulation spectrum). The mean z-score at meter-related frequencies was taken as a measure of contrast at these frequencies in the neural response, and was compared across conditions.

Comparison of EEG with models of subcortical auditory processing. To assess whether the observed EEG responses could be explained by nonlinearities at the early stages of the auditory pathway, the elicited higher-level responses were directly compared to the sound representation estimated by the auditory models. First, the relative prominence of meter frequencies was calculated separately for each model of subcortical auditory processing by taking the mean z-score at meter-related frequencies. Then, the mean meter-related z-scores obtained from the higher-level EEG responses across participants were compared to the corresponding meter z-score from each auditory model, separately for each rhythm and task, using a one-sample t-test. The same analysis was performed for the early auditory responses to confirm that these responses could be largely explained by the auditory models.

Comparison of higher-level and early auditory responses. To compare the contrast at meter-related frequencies between early and later processing stages in the same participants, mean z-scored amplitude at these frequencies was compared between the higher-level and early auditory responses across tasks and rhythms.

Moreover, because the highest meter-related frequency (5 Hz) seemed selectively attenuated in the early auditory responses (see Figure 2.4B), we carried out a control analysis to make sure the differences between higher-level and early auditory responses were not solely driven by differences in the low-pass characteristics of the two responses. The z-scores were re-calculated using only 11 frequencies of interest, i.e. after excluding the amplitude at 5 Hz from the set. Subsequently, only z-scores at 0.416, 1.25, and 2.5 Hz were considered meter-related, and their average was taken to estimate how prominent these frequencies were relative to the whole set of 11 frequencies in the control analysis. This control measure was then used to repeat the comparison between higher-level and early

auditory responses, as well as their respective comparisons to the models of subcortical auditory processing.

2.3.10 Behavioral analyses

Responses to the Tempo and Pitch task from the EEG session were transformed into the sensitivity index d' using the equation $Z(\text{hit rate}) - Z(\text{false alarm rate})$, where Z is the inverse of the normal cumulative distribution function (Stanislaw and Todorov, 1999). To avoid infinite values, hit rates and false alarm rates with values 1 and 0 were converted to $1/(2N)$ and $1-1/(2N)$ respectively (where N is the number of trials on which the proportion is based, Macmillan and Creelman, 2005). The response accuracy on the Visual task was calculated by taking the root-mean-square deviation from the correct response (i.e. the sum of the sequentially-presented number) across trials and averaging across the two rhythm conditions.

2.3.11 Statistical analyses

The statistical analyses were performed in R (version 3.6.1). Comparisons of EEG measures across conditions were implemented using linear mixed models with lme4 package (version 1.1-21, Bates et al., 2015). The main effects of Rhythm (high meter contrast, low meter contrast) and Task (Tempo, Pitch, Visual), and their interaction were included as fixed effects. For the comparison between the higher-level and early auditory responses, the factor Response (higher-level, early auditory) was also included in the model. Each participant was included as a random-effect intercept. Diagnostic plots of the residuals from all models were inspected for violations of the assumptions of normality and homoscedasticity. No substantial violations were detected. The model comparison was carried out with the *Anova* function from package car (version 3.0-3), using F-tests. Post-hoc comparisons on the fitted models were conducted using emmeans package (version 1.4). Degrees of freedom were approximated using the Kenward-Roger approach and Bonferroni correction was used to adjust the post-hoc test for multiple comparisons. Nonparametric Wilcoxon signed rank tests were used to compare the behavioural responses between conditions, and to assess the significance of the early auditory overall response magnitude against zero (FDR correction for multiple comparisons was used in the latter case). One-

sample t-tests were used to compare the meter-related z-scores from the EEG data to the models of subcortical auditory processing (one-tailed, testing EEG > model; p-values were adjusted for multiple comparisons using FDR).

In addition to the null-hypothesis significance tests, we calculated Bayes factors (BF_{10}) to express the probability of data under alternative hypothesis (H_1) relative to null hypothesis (H_0), as implemented in BayesFactor package (version 0.9.12-4.2). We considered a $BF_{10} > 3$ as evidence in favour of the alternative hypothesis and $BF_{10} < 0.3$ as evidence in favour of the null hypothesis (Jeffreys, 1998; Lee and Wagenmakers, 2014).

2.4 Results

2.4.1 Behavioral results

Participants successfully detected the deviants during the EEG session for both the Pitch task (mean d' [SD] for the high meter contrast rhythm = 1.23 [0.92], for the low meter contrast rhythm = 1.28 [0.55]), and the Tempo task (mean d' [SD] for the high meter contrast rhythm = 1.28 [0.77], for the low meter contrast rhythm = 1.28 [1.10]). There was no difference in the sensitivity between the Pitch and Tempo task ($P_s > 0.25$, $BF_{s10} < 0.6$). In the Visual task, the root-mean-square error of the responses averaged across rhythms was 87.1 (SD = 71.6), suggesting that the task was difficult and participants might have often missed some numbers in the sequence. This was in line with participants rating the Visual task as much more difficult than the Pitch task (Wilcoxon signed rank test, two-sided $P = 0.0007$), or the Tempo task (Wilcoxon signed rank test, $P = 0.0003$). One participant's data were excluded from the calculation of the error, due to the misunderstanding of the instructions - instead of summing, the participant was concatenating the presented numbers and memorizing the sequence. However, the EEG data of this participant were not excluded, as she reported similar difficulty for the Visual task as the rest of the participants.

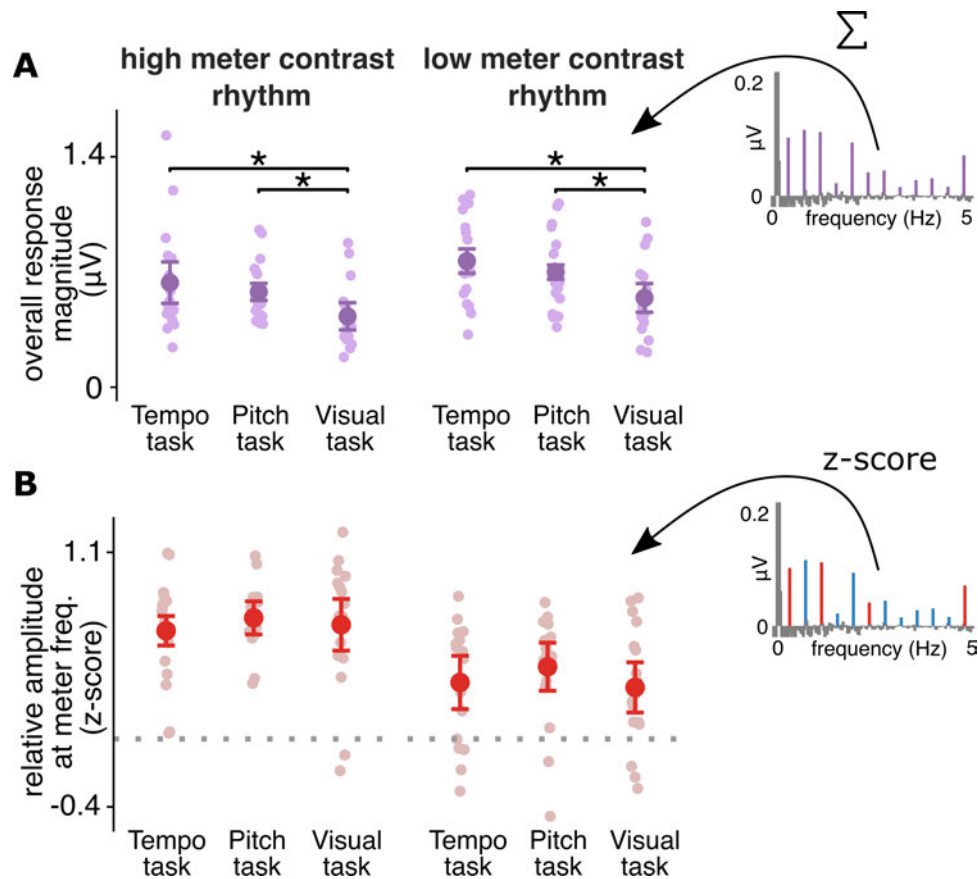


Figure 2.3. Characteristics of the higher-level EEG responses. The example magnitude spectra on the right visualize how each measure was quantified. Individual participants are shown as lightly shaded data points. Error bars represent 95% CIs (Morey, 2008). (A) Overall response magnitude for the higher-level EEG responses. The amplitudes of the response at all 12 frequencies corresponding to the modulation spectrum of the sound were summed and compared across conditions. For both rhythms, the response was significantly lower during Visual task compared to the two other tasks involving attention to the auditory stimulus (marked by asterisks). (B) Prominence of meter frequencies (mean z-scored amplitude at meter-related frequencies) in the higher-level EEG responses. There was no difference across the three tasks for either rhythm ($BF_{10} < 0.3$). Moreover, the z-scores show prominent meter frequencies even in response to the rhythm with low contrast at meter frequencies in the acoustic input. The horizontal dashed line represents zero.

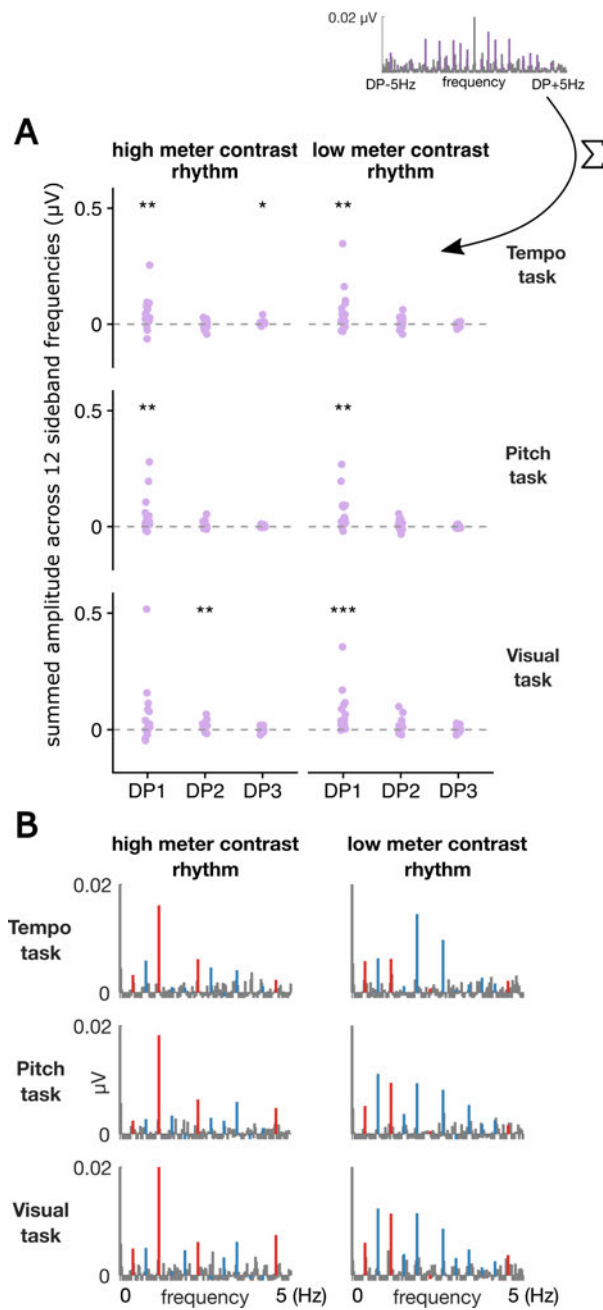


Figure 2.4. Characteristics of the early auditory EEG responses. (A) Summed early auditory response amplitude averaged across all sidebands, separately for each distortion product frequency (DP1 = 168 Hz, DP2 = 189 Hz, DP3 = 357 Hz). The example magnitude spectrum on the top illustrates how the measure was quantified. Purple data points represent individual participants. Asterisks indicate the statistical significance level of the response when tested against zero (grey dashed line) across participants. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (Wilcoxon signed rank test, FDR corrected). (B) Spectra of early auditory responses (average across all participants) plotted for the 168 Hz distortion product (DP1) after the corresponding symmetrical sidebands elicited at stimulus modulation frequencies were averaged. The frequency axis is normalized by subtracting the distortion-product frequency for better comparison with the higher-level EEG responses. Hence, the plots

represent spectra of demodulated EEG responses that followed the 168 Hz distortion product frequency. Meter-related frequencies are shown in red, meter-unrelated frequencies in blue.

2.4.2 Overall EEG response magnitude

Figure 2.1 shows the spectra of the higher-level EEG responses elicited by the rhythmic stimuli across the three tasks. As shown in Figure 2.3A, the summed amplitude of the higher-level responses across all twelve frequencies was significantly different across the three tasks ($F_{2,80} = 16.3$, $P < 0.0001$, $BF_{10} > 100$). This was due to smaller overall amplitude in the Visual task relative to the Tempo task ($\beta = -0.21$, $t_{82} = -5.61$, $P < 0.0001$, 95% CI = [-0.31, -0.12]), and the Pitch task ($\beta = -0.15$, $t_{82} = -4$, $P = 0.0004$, 95% CI = [-0.25, -0.06]). The corresponding analysis performed on the shifted frequency bins where no signal was expected indicated that broadband noise amplitude was comparable across conditions (no significant main effect of Task: $P = 0.62$, $BF_{10} = 0.13$, and no interaction Task x Rhythm: $P = 0.39$, $BF_{10} = 0.32$). Thus, increased noise alone could not account for the observed overall response decrease in the Visual task.

Figure 2.4B shows the spectra of the early auditory EEG responses. Unlike for the higher-level responses, the summed amplitude across all sidebands for the early auditory responses did not differ across conditions ($P_s > 0.22$, $BF_{s10} < 0.21$). This suggested that attentional focus did not affect the putatively earlier stage of sound processing, and only emerged at later stages.

2.4.3 Relative EEG response at meter frequencies

The relative prominence of meter frequencies in the elicited higher-level EEG responses was significantly larger for the high than low meter contrast rhythm ($F_{1,80} = 38.9$, $P < 0.0001$, $BF_{10} > 100$), as expected based on the physical structure of the rhythmic stimuli (see Methods section). However, as shown in Figure 2.3B, z-scores at meter frequencies were comparable across tasks (no significant main effect of Task: $P = 0.32$, $BF_{10} = 0.24$, and no interaction Task x Rhythm: $P = 0.79$, $BF_{10} = 0.18$). Similarly, meter-related frequencies were significantly more prominent in the early auditory responses to the high meter contrast rhythm ($F_{1,80} = 42.8$, $P < 0.0001$, $BF_{10} > 100$), but there was no effect of task ($P_s > 0.65$, $BF_{s10} < 0.13$).

Table 2.1. Comparison of mean z-scored amplitude at meter-related frequencies between higher-level EEG responses and models of auditory subcortical processing using one-sample t-tests.

EEG variable	rhythm	model	task	mean_model	mean_eeg	sd_eeg	t	df	p
higher-level response	high meter contrast rhythm	broadband	pitch	0.81	0.71	0.19	-1.99	16	0.988
			tempo	0.81	0.64	0.31	-2.28	16	0.988
			visual	0.81	0.67	0.38	-1.47	16	0.988
		UREAR_AN	pitch	0.52	0.71	0.19	4.15	16	0.0010 ***
			tempo	0.52	0.64	0.31	1.61	16	0.112
			visual	0.52	0.67	0.38	1.69	16	0.101
		UREAR_IC_BMF2	pitch	0.83	0.71	0.19	-2.36	16	0.988
			tempo	0.83	0.64	0.31	-2.51	16	0.988
			visual	0.83	0.67	0.38	-1.65	16	0.988
		UREAR_IC_BMF4	pitch	0.79	0.71	0.19	-1.58	16	0.988
			tempo	0.79	0.64	0.31	-2.01	16	0.988
			visual	0.79	0.67	0.38	-1.25	16	0.988
		UREAR_IC_BMF8	pitch	0.75	0.71	0.19	-0.787	16	0.988
			tempo	0.75	0.64	0.31	-1.51	16	0.988
			visual	0.75	0.67	0.38	-0.846	16	0.988
		UREAR_IC_BMF16	pitch	0.80	0.71	0.19	-1.76	16	0.988
			tempo	0.80	0.64	0.31	-2.13	16	0.988
			visual	0.80	0.67	0.38	-1.35	16	0.988
		UREAR_IC_BMF32	pitch	0.81	0.71	0.19	-2.09	16	0.988
			tempo	0.81	0.64	0.31	-2.34	16	0.988
			visual	0.81	0.67	0.38	-1.51	16	0.988
		UREAR_IC_BMF64	pitch	0.68	0.71	0.19	0.655	16	0.447
			tempo	0.68	0.64	0.31	-0.601	16	0.988
			visual	0.68	0.67	0.38	-0.105	16	0.896
	low meter contrast rhythm	broadband	pitch	0.06	0.43	0.33	4.66	16	0.0004 ***
			tempo	0.06	0.33	0.35	3.31	16	0.005 **
			visual	0.06	0.30	0.35	2.88	16	0.010 *
		UREAR_AN	pitch	-0.13	0.43	0.33	6.96	16	<0.0001 ***
			tempo	-0.13	0.33	0.35	5.48	16	0.0001 ***
			visual	-0.13	0.30	0.35	5	16	0.0003 ***
		UREAR_IC_BMF2	pitch	-0.04	0.43	0.33	5.93	16	<0.0001 ***
			tempo	-0.04	0.33	0.35	4.5	16	0.0005 ***
			visual	-0.04	0.30	0.35	4.04	16	0.001 **
		UREAR_IC_BMF4	pitch	-0.29	0.43	0.33	9.01	16	<0.0001 ***
			tempo	-0.29	0.33	0.35	7.42	16	<0.0001 ***
			visual	-0.29	0.30	0.35	6.89	16	<0.0001 ***
		UREAR_IC_BMF8	pitch	-0.27	0.43	0.33	8.76	16	<0.0001 ***
			tempo	-0.27	0.33	0.35	7.18	16	<0.0001 ***
			visual	-0.27	0.30	0.35	6.66	16	<0.0001 ***
		UREAR_IC_BMF16	pitch	-0.08	0.43	0.33	6.42	16	<0.0001 ***
			tempo	-0.08	0.33	0.35	4.97	16	0.0003 ***
			visual	-0.08	0.30	0.35	4.5	16	0.0005 ***
		UREAR_IC_BMF32	pitch	0.05	0.43	0.33	4.73	16	0.0004 ***
			tempo	0.05	0.33	0.35	3.37	16	0.004 **
			visual	0.05	0.30	0.35	2.94	16	0.010 **
		UREAR_IC_BMF64	pitch	0.01	0.43	0.33	5.28	16	0.0002 ***
			tempo	0.01	0.33	0.35	3.89	16	0.002 **
			visual	0.01	0.30	0.35	3.45	16	0.004 **

. P < 0.1, * P < 0.05, ** P < 0.01, *** P < 0.001, FDR corrected

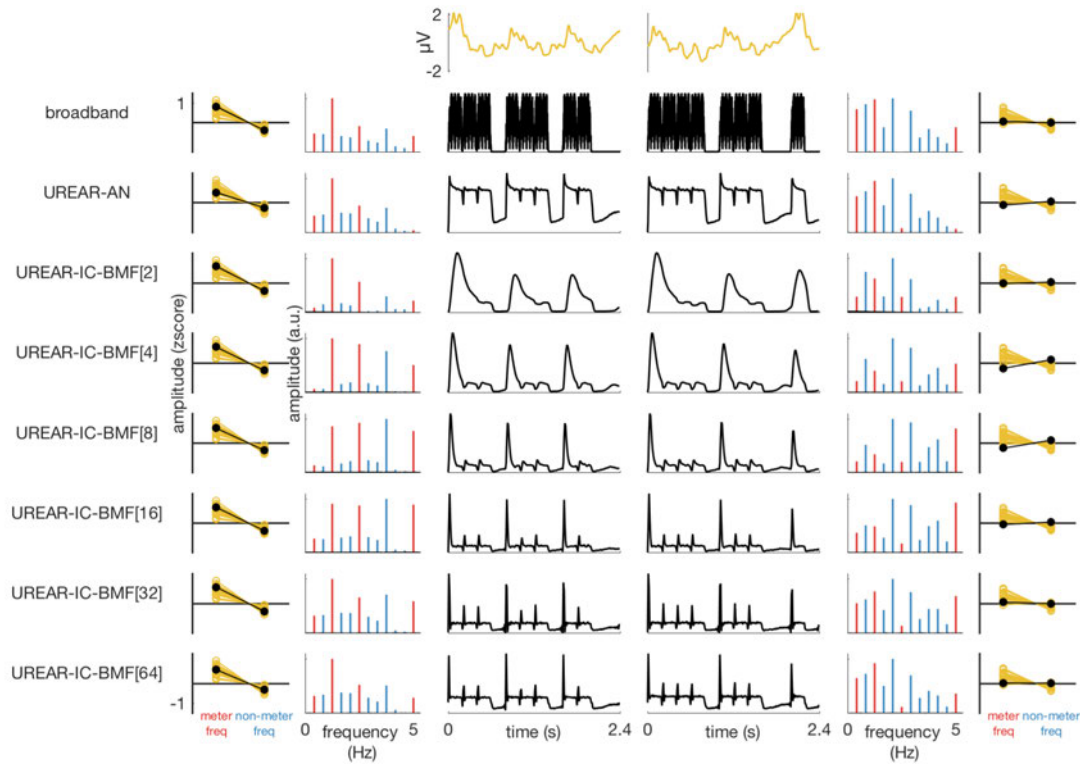


Figure 2.5. Comparison of higher-level EEG responses with models of auditory subcortical processing. Data for the high meter contrast rhythms are shown on the left, and data for the low meter contrast rhythm are shown on the right. The higher-level EEG response to one pattern cycle averaged across all pattern repetitions, trials, tasks, and participants is depicted on the top in yellow. This response was extracted after low-pass filtering at 30 Hz, and averaging 9 frontocentral channels (F1, F2, Fz, C1, C2, Cz, FC1, FC2, FCz). Each row below the EEG response corresponds to the output of one model of subcortical auditory processing. The model labels are shown on the left (depending on the parameter settings, the IC cell simulated with UR_EAR had a specific best modulation frequency, which is listed in square brackets in Hz). In the center of the figure are the responses of the models to one cycle of the rhythmic pattern depicted as mean firing rate across time. The mean rate over time was transformed into the frequency domain using FFT, and the resulting spectra are shown next to the time-domain responses. Meter-related frequencies are shown in red, and meter-unrelated frequencies in blue. On the sides of the figure are the z-scored spectral amplitudes averaged separately across meter-related and meter-unrelated frequencies. The yellow data points represent EEG responses of individual participants (z-scores averaged across the 3 tasks), and the black data points represent the auditory model. While the z-scores at meter-related frequencies did not differ reliably between the models and the EEG responses for the high meter contrast rhythm on the left, the EEG response at these frequencies was selectively enhanced for the low meter contrast rhythm (right).

2.4.4 Comparison of EEG responses with models of subcortical auditory processing

For the high meter contrast rhythm, the relative prominence of meter-related frequencies in the higher-level EEG responses was explained by all considered models of subcortical auditory processing (see Figure 2.5 and Table 2.1). The mean z-score at meter-related frequencies measured in the elicited EEG was not significantly different from either model, except there was a significantly greater z-score for the EEG responses in the pitch task when compared to the UR_EAR model of the auditory nerve response. However, for the low meter contrast rhythm, meter-related frequencies were consistently more prominent in the elicited EEG responses than predicted by all models of subcortical auditory processing. Importantly, this was the case even when participants were carrying out the visual task. These results were further corroborated by a control analysis showing that meter-related frequencies in the higher-level responses were robustly enhanced even when the highest meter frequency (5 Hz) was excluded from the analysis (see Table 2.S2).

As expected, the same set of comparisons for the early auditory responses showed no clear differences from the models of subcortical auditory processing (see Table 2.S3). Even though the meter frequencies in the early auditory responses were significantly above some auditory models after excluding the highest meter-related frequency in the control analysis (UR-EAR-IC with best modulation frequencies 4, 8, 16 Hz, see Table 2.S4), this was not systematic (in contrast to the higher-level responses), and was related to a selective suppression of the lower meter frequencies in these models compared to the rest of the models (see Figure 2.5).

Together, these results indicate that, as expected, the contrast at meter frequencies in the EEG response to the rhythmic input with prominent meter frequencies in its acoustic structure was mainly driven by low-level processing of this input. However, these early processing stages cannot fully explain the EEG response to the rhythmic input with less prominent meter frequencies in its acoustic structure. Most importantly, the processes responsible for the selective enhancement of meter frequencies for this latter rhythm seemed to be involved to a similar degree across the three attentional tasks.

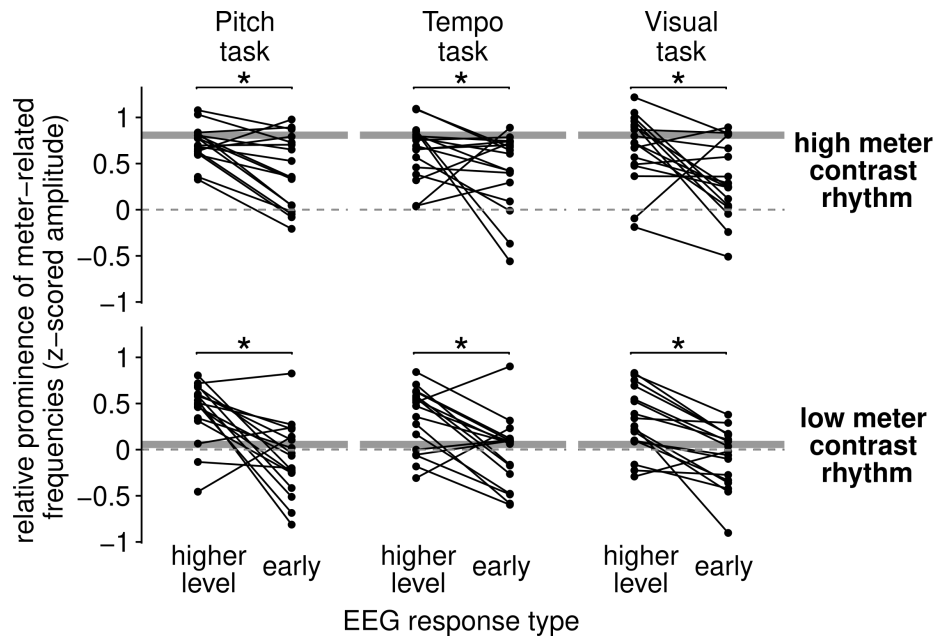


Figure 2.6. Comparison of prominence of meter-related frequencies in the higher-level and early auditory EEG responses. The mean z-scored amplitude at meter-related frequencies is plotted separately for each task, rhythm, and EEG response type. Individual data points represent participants. Horizontal continuous grey lines correspond to the mean z-scores at meter frequencies taken from the broadband amplitude envelope of the corresponding acoustic stimulus. The horizontal dashed grey lines represent zero (i.e. equal relative prominence of meter-related and meter-unrelated frequencies). The meter-related frequencies were consistently more prominent in the higher-level EEG responses across all rhythms and task (main effect of response type, indicated by asterisks).

2.4.5 Comparison of higher-level and early auditory EEG responses

A mixed model with factors Rhythm, Task and Response revealed a main effect of Rhythm ($F_{1,176} = 68.03$, $P < 0.0001$, $BF_{10} > 100$), as expected from the separate analyses of the higher-level and early auditory responses above, suggesting that both types of response were sensitive to the physical structure of the auditory input. There was also a main effect of Response ($F_{1,176} = 58.5$, $P < 0.0001$, $BF_{10} > 100$). As shown in Figure 2.6, the relative prominence of meter frequencies was consistently larger in the higher-level responses across all tasks and both rhythms. Moreover, this could not be easily explained by differences in the low-pass characteristic of the responses, as the same results were obtained in a control analysis where the highest meter-related frequency (i.e. 5 Hz) was excluded (see Supplementary Materials). This suggests that there was a significant

enhancement of meter-related frequencies at the later processing stage indexed by the higher-level responses.

2.5 Discussion

Our results show that while attentional focus affects the overall sensitivity of the brain to auditory rhythmic inputs, it has little influence on the selective contrast at meter periodicities in the elicited neural response. Moreover, while the magnitude of this selective contrast in the EEG response was readily explained by low-level auditory processing for the rhythm with prominent meter frequencies in the acoustic input, this was not the case for the rhythm that lacked prominent contrast at meter periodicities in its physical structure. Together, these results suggest the critical engagement of high-level processes that shape the neural representation of a rhythmic input by selectively enhancing contrast at meter periodicities across behavioral contexts even in rhythms where this contrast is not prominent. These results add to the evidence that rhythm perception is shaped by a range of processes including higher-level cortical stages, with different degrees of flexibility and automaticity.

2.5.1 Wide range of low-level and higher-level processes in meter perception

A number of studies have consistently shown that the brain can selectively enhance meter-related frequencies, particularly for low meter contrast rhythms. However, the behavioral context varied across these studies. Some asked participants to report small occasional changes in the duration of a sound event making up the rhythm (Nozaradan et al., 2012, 2017b; Lenc et al., 2018). While those small changes in duration of single time intervals are difficult to detect at a local scale, their detection is easier on a global scale, as the deviant interval results in a misalignment of the subsequent sequence with the perceived meter. Therefore, this task implicitly encourages participants to rely on an internal metric structure (Schulze, 1978; Jones and Yee, 1997; Grube and Griffiths, 2009). Similarly, some studies asked participants to focus on the tempo (overall perceived speed) of the sequences, either searching for local transient changes (Lenc et al., 2020), or for later comparison with a test sequence (Tal et al., 2017). Other studies have asked participants to report any temporal

irregularities in the stimuli while none were actually present (Nozaradan et al., 2018), or to simply attend to the sequences (Nozaradan et al., 2016a). Our results suggest that selective neural enhancement of meter frequencies can take place even without attention directly focused on the temporal properties of the stimulus (as observed during the pitch task). Interestingly, this enhancement was also observed when sound was ignored altogether (during the visual task), i.e., even when the overall signal-to-noise ratio (SNR) of the response was decreased. Nevertheless, our results inform future frequency-tagging studies, which may find it advantageous to employ behavioral tasks that encourage participants to focus on the auditory stimuli, even when the task-relevant dimension is orthogonal to rhythm processing, to increase the SNR and facilitate estimation of response properties.

While our results suggest that neural responses to rhythmic input might involve processes largely independent on attentional focus, this does not imply that neural processing of rhythm is fixed and inflexible. A number of recent studies have shown that selective neural enhancement of meter periodicities reflects flexible processes, which are sensitive to mental imagery (Nozaradan et al., 2011; Li et al., 2019), non-temporal features of the acoustic input (Lenc et al., 2018), prior experience (Chemin et al., 2014), and recent context (Lenc et al., 2020). Importantly, our current findings show that the internal transformation of a rhythmic input towards a particular metric category can be flexibly enhanced, possibly changed, but it is difficult to suppress completely: it has an automatic component. Whether the robust component can be changed by long-term exposure remains to be seen in future studies (Hannon et al., 2012b, 2012a; London et al., 2017; van der Weij et al., 2017; Polak et al., 2018).

These findings thus reveal similarities between meter processing and other higher-level auditory processes, such as auditory stream segregation or change detection (Sussman, 2017). While automatic in some contexts (Woods et al., 1992; Alho et al., 1994; Dyson et al., 2005; Teki et al., 2011, 2016; Masutomi et al., 2016), attention can boost or bias these processes (Haroush et al., 2010; Auksztulewicz and Friston, 2015; O'Sullivan et al., 2015; Costa-Faidella et al., 2017), particularly when the input provides ambiguous sensory cues to the system (Sussman et al., 2007; Gutschalk et al., 2015). Similarly, the brain response to isochronous auditory sequences might be spontaneously shaped by the intrinsic preference for binary structures (Brochard et al., 2003; Pablos Martin et al., 2007), but this can be biased towards different forms of organization by top-down processes dependent on

attentional resources (Nozaradan et al., 2011; Chemin et al., 2014; Celma-Miralles and Toro, 2019). Importantly, some higher-level auditory processes can be largely suppressed (especially by high load) but not completely abolished (Woldorff et al., 1991; Alain and Izenberg, 2003; Chait et al., 2012; Billig and Carlyon, 2016; Molloy et al., 2019). Together, these studies suggest that when assessing effects of attention on a particular perceptual phenomenon, it is important to keep in mind that (i) multiple processes can be involved (Chait et al., 2012), (ii) these different processes may be differentially affected by different kinds of tasks (Bidet-Caulet et al., 2007; Yerkes et al., 2019), and (iii) the effect of task may also depend on the sensory information provided by the stimulus (Gutschalk et al., 2015). With respect to meter perception, identification of the underlying internal processes and the types of resources they rely on remains worthwhile goal for future investigations (Lenc et al., 2018).

2.5.2 Dissociation between overall gain and selective contrast at meter periodicities

An overall decrease in sensitivity to sound input while carrying out a visual task has been previously reported by a large number of studies measuring early event-related potentials (Woods et al., 1992; Alho et al., 1994; Okamoto et al., 2011), frequency tagged responses (Keitel et al., 2011, 2013; Riecke et al., 2014), or BOLD activations in sensory cortices (Petkov et al., 2004; Shomstein and Yantis, 2004; Johnson and Zatorre, 2005; Riecke et al., 2017). In line with these studies, we observed increased overall gain of the higher-level EEG responses during the auditory tasks in comparison to the visual task. In other words, there was a general increase in the amplitude of neural activity time-locked to the rhythmic auditory stimulus. However, this non-specific enhancement could simply represent a proportional increase of magnitude across the response spectrum, thus being equivalent to a multiplicative enhancement of the response in the time domain. In that case, the relative contrast in the response at meter periodicities should necessarily remain constant.

As opposed to such non-specific enhanced gain of the whole response, a change in the selective contrast at meter periodicities would demonstrate selective increase at meter-related frequencies. The important distinction between overall gain and selective contrast at meter frequencies has been often neglected in studies claiming to measure “neural entrainment to meter” (Tierney and Kraus, 2014; e.g. Hickey et al., 2020). Our results

provide a cautionary example, whereby non-specific increase in response to the sound trivially explained by attention could have been misinterpreted as an enhancement of meter periodicities in brain responses. Instead, our method allowed us to dissociate between these two accounts, exploiting the fact that energy in the modulation spectra of our complex rhythmic stimuli is not solely distributed across meter-related frequencies. For this reason, using non-isochronous rhythms, particularly rhythms with less energy at meter frequencies, is advantageous over strictly isochronous stimuli where changes in overall response gain and selective contrast enhancement cannot be differentiated. In addition, using z-score normalization instead of the difference in raw spectral magnitude makes the measure robust to multiplicative gain (which would yield greatest raw magnitude increase at frequencies already prominent in the spectra).

2.5.3 Robust responses at meter periodicities even with low meter contrast in the input

Despite the significant differences in the gain of the higher-level EEG responses, the selective contrast at meter frequencies was not affected by task. However, this finding could be trivially explained by passive matching of stimulus modulation structure in the neural response. While faithful tracking of stimulus envelope is fundamental for auditory perception (Peelle et al., 2013; Di Liberto et al., 2018; Etard and Reichenbach, 2019; Ghinst et al., 2019), the brain must go beyond one-to-one representation of the sensory input to achieve adaptive behavior (Kuchibhotla and Bathellier, 2018). Thus, the sensory input is continuously transformed within the brain towards higher-level categories (Ley et al., 2014; Brodbeck et al., 2018; Rossion et al., 2020; Sankaran et al., 2020; Yin et al., 2020). Such transformations are critical for timing perception already at the level of single intervals (Desain and Honing, 2003; Jacoby and McDermott, 2017), but also patterns of intervals (Notter et al., 2018), and for meter perception, where a range of physically different acoustic inputs can be mapped onto the same set of periodic pulses (Nozaradan et al., 2017a). To assess whether the internal processes involved in this transformation were engaged even when attention was withdrawn from the auditory input, we compared the higher-level EEG responses to simulated representation of the auditory input across different auditory subcortical stages. While keeping in mind that absence of evidence is not evidence of absence, our results suggests that well-described low-level nonlinearities in

early auditory pathway cannot fully account for the cortical brain response to the low meter contrast rhythm.

This was further confirmed by the early sensory responses measured in the same EEG as the high-level responses. Even though the tagged frequencies used to identify the early sensory responses in the current study were most likely not in a high enough frequency range to strictly isolate brainstem responses from cortical responses (Coffey et al., 2016; Holmes et al., 2018), these responses may have preferentially captured contributions from primary auditory fields, as well as subcortical nuclei (Chandrasekaran and Kraus, 2010; Nourski and Brugge, 2011). Therefore, comparing EEG responses tagged at low frequencies (higher-level responses) vs. high frequencies (early auditory responses) may still be a useful way to separate sound representation in early auditory cortices from responses originating in a wide network of structures involved in rhythm processing (Patel and Iversen, 2014; Merchant et al., 2015a).

Our results are in line with Nozaradan et al. (2018) who observed a similar enhancement of meter-related frequencies in the higher-level EEG responses compared to the early auditory EEG responses. While they only observed higher-level response enhancement for the low meter contrast rhythm, this was the case for both rhythms in the current study. However, if our current results were related to low SNR for early auditory responses resulting in attenuation of the most prominent peaks in the spectra, one would expect to find opposite effects on meter-related frequencies for the two rhythms due to the differences in their physical structure. Moreover, non-selective attenuation of high frequencies in the early auditory responses alone did not fully explain the smaller prominence of meter frequencies when compared to the higher-level EEG responses. Finally, the prominence of meter frequencies in the early auditory responses was strongly modulated by the type of rhythm, showing sensitivity to the acoustic structure of the input. Therefore, together with the output of subcortical auditory processing models, our results provide evidence that (i) higher-level processes further enhance contrast at meter frequencies, especially when these meter frequencies are not prominent in the auditory input, and (ii) these processes remain active even when overall responsiveness to sound input is decreased (e.g. during a demanding visual task).

Recently, it has been proposed that low-level nonlinearities such as adaptation, amplitude-modulation tuning, and heightened sensitivity to contrast in early stages of the auditory

pathway could predict whether and what metric structure is perceived in a rhythmic stimulus (Rajendran et al., 2017; Zuk et al., 2018), and the consistency of perceived meter across individual listeners (Rajendran et al., 2020). While these low-level phenomena are definitely important in shaping rhythm perception, such models are inherently limited to enhancement of sensitivity to contrast that is *already present* in the physical input. Indeed, these models are based on the long-standing assumption in psychology and neuroscience (likely driven by over-emphasis on Western classical and popular music) that meter perception is driven by temporal contrasts defined by acoustic properties of the sound input (Longuet Higgins and Lee, 1982; Povel and Essens, 1985; Jones and Boltz, 1989; Palmer and Krumhansl, 1990; Drake et al., 2000; Toiviainen and Snyder, 2003). Strong arguments against these assumptions have been recently raised by a number of authors (see, e.g. London et al., 2017; van der Weij et al., 2017). Indeed, such models will unlikely explain perception of musical genres where the phase (e.g. reggae, ska, swing, mazurka) or period (e.g. tresillo, cascara, or rumba clave in afro-cuban music) of the perceived metric structure is weakly cued in the temporal distribution of features in the physical sound. Instead, over-constrained views may in the end lose explanatory power by ignoring diversity and flexibility in the cognitive phenomenon across cultures in pursuit of a reductionist mechanistic explanation.

The weak explanatory power of biologically plausible models of subcortical auditory processing to account for our EEG results adds to the evidence that meter perception involves higher-level transformations of the input, providing flexibility within (Repp, 2007; Repp et al., 2008; Chemin et al., 2014; Lenc et al., 2020) and across individuals (McKinney and Moelants, 2006; Martens, 2011; Hannon et al., 2012a; Kalender et al., 2013; Polak et al., 2018; Witek et al., 2020). Instead of offering a mechanistic explanation for the current EEG results, we emphasize the need for more data and powerful designs, in order to thoroughly describe the perceptual phenomenon in question, and how it is shaped by input features, behavioral goals, context, exposure, and learning. Similarly, we do not claim that the contrast at meter frequencies measured in EEG responses is one-to-one with meter perception in a phenomenological sense. At the same time, it is important to note that all measures of perception are indirect (including behavioral measures), and critically depend on the definition of the perceptual phenomenon (see also Rossion et al., 2020). If meter is defined as the perception of pulses that are time-locked to the temporal structure of the

stimulus, and if pulse is understood as something that consistently occurs at regularly-spaced time points *and not otherwise* (thus creating a temporal contrast), our EEG measure is directly relevant for meter processing. Moreover, our analysis was directly informed by tapping data from previous studies, thus constraining the set of behaviorally relevant periodicities based on an additional measure of meter perception.

The fact that we observed significantly enhanced responses at meter frequencies in the low meter contrast rhythm irrespective of attentional focus may seem inconsistent with the fMRI study of Chapin et al. (Chapin et al., 2010). In that study, participants listened to rhythms that had few acoustic cues to meter periodicities but still elicited stable meter perception. The authors observed larger BOLD responses within a network of structures typically associated with meter perception when participants were actively listening (memorizing the rhythm for subsequent reproduction) than when they were memorizing a visual array of letters. However, this result could potentially be explained by non-specific changes in the BOLD response due to auditory stimulation interacting with attention, or task-related motor preparation. Based on the current results, we suggest that the processing of low meter contrast rhythms may not be inherently different from high meter contrast rhythms. Indeed, rhythms with little acoustic cues to meter are ubiquitous across cultures (Cohn, 2016; London et al., 2017; Witek, 2017; Câmara and Danielsen, 2018). Hence, such rhythmic inputs may help to reveal the transformations that take place within the brain when the sensory input is mapped onto an internal metric representation, and eventually behavioral output, while controlling for acoustic or low-level confounds.

2.5.4 Evidence for robust meter processing complementary to MMN studies of passive listening

Our observation that processes related to meter perception are engaged robustly across behavioral contexts is consistent with previous studies using the mismatch-negativity event-related potential, or MMN (Ladinig et al., 2009; Winkler et al., 2009; Bouwer et al., 2014, 2016). Complementary to these studies, we observed task-independent neural enhancement of meter frequencies even for the low meter contrast rhythm, while MMN responses have only been assessed for rhythms with very prominent acoustic cues to meter periodicities. Because MMN paradigms rely on assumptions about predictions and regularity

violations, which are not well-defined for meter perception (see e.g. London et al., 2017), MMN studies involve uncertainty about the type of deviation that may elicit differential responses depending on its timing relative to the perceived pulse (Bouwer and Honing, 2015). MMN studies therefore rely on statistically linking the perceived pulse with salient sound events, and hence limit themselves to study of high meter contrast rhythms. Moreover, typical MMN studies employ passive listening with low cognitive load and no control of participant's attentional focus (Sussman et al., 2014), whereas we directly manipulated the attentional state of the listener with active demanding tasks that significantly affected the overall magnitude of the EEG responses and also parieto-occipital alpha power (see Supplementary Materials), which is an established index of crossmodal attentional engagement (Fu et al., 2001; Jensen and Mazaheri, 2010; Mo et al., 2011; Mazaheri et al., 2014). Our results thus represent an important step towards describing whether processes involved in meter perception depend on limited resources, which may be shared across modalities and cognitive domains (Marois and Ivanoff, 2005; Chait et al., 2012; Murphy et al., 2017; Molloy et al., 2020).

2.5.5 Conclusions

The human auditory system possesses a remarkable capacity to carry out high-level processes with limited attentional resources (Murphy et al., 2017). The current study provides evidence that this may also be the case for processes involved in meter perception in the context of musical rhythm. Our results indicate that the brain selectively emphasizes perceptually relevant periodicities even when overall sensitivity to sound is decreased due to a distracting task. Moreover, such perceptual emphasis occurs when the periodicities are not prominent in the sensory input, and their enhancement is not readily accounted for by low-level auditory processing. Therefore, these robust neural processes to auditory rhythms may support the spontaneity of meter perception when listening to a variety of musical inputs, while still allowing for flexibility and context dependence in meter perception within and across individuals.

2.6 Supplementary Material

2.6.1 Control analysis of higher-level and early auditory EEG responses excluding the highest (5 Hz) frequency

To make sure that the differences in the prominence of meter-related frequencies between higher-level and early auditory EEG responses were not solely driven by low-pass biases, we re-calculated the relative prominence of meter frequencies after excluding the highest frequency of interest (5 Hz) from the set. This frequency corresponded to rate of individual events in the rhythms, but captured also harmonics of the slower perceived metric pulses. Yet, the amplitude at this frequency would be affected most prominently if higher frequencies in the response were broadly attenuated irrespective of their contribution to the contrast at meter periodicities (e.g. within a neural network behaving like a simple low-pass filter). A mixed model with factors Rhythm, Task and Response still revealed a main effect of Rhythm ($F_{1,176} = 53.4$, $P < 0.0001$, $BF_{10} > 100$) and Response ($F_{1,176} = 35.0$, $P < 0.0001$, $BF_{10} > 100$), thus replicating the results from the main analysis. This indicates that the enhanced selective contrast at meter periodicities in the higher-level responses cannot be fully explained by non-selective enhancement of higher frequencies (i.e. without considering their relevance for the perceived meter).

Table 2.S1. Comparison of mean early auditory response amplitude averaged across all sidebands against zero using non-parametric Wilcoxon signed rank tests.

distortion product frequency	rhythm	task	mean	p	
168	high meter	pitch	0.048039	0.007	**
	contrast	tempo	0.044929	0.007	**
	rhythm	visual	0.056336	0.057	.
	low meter	pitch	0.056919	0.003	**
	contrast	tempo	0.053233	0.007	**
	rhythm	visual	0.068828	0.0003	***
189	high meter	pitch	0.008148	0.057	.
	contrast	tempo	-0.000029	0.463	
	rhythm	visual	0.015129	0.007	**
	low meter	pitch	0.007007	0.156	
	contrast	tempo	0.007147	0.155	
	rhythm	visual	0.011202	0.243	
357	high meter	pitch	0.002491	0.057	.
	contrast	tempo	0.004959	0.019	*
	rhythm	visual	0.001892	0.329	
	low meter	pitch	0.000452	0.452	
	contrast	tempo	0.001185	0.089	.
	rhythm	visual	0.000780	0.440	

. P < 0.1, * P < 0.05, ** P < 0.01, *** P < 0.001, FDR corrected

Table 2.S2. Control analysis of mean z-scored amplitude at meter-related frequencies without taking the highest frequency (5 Hz) into account. Comparison between higher-level EEG responses and models of auditory subcortical processing using one-sample t-tests.

EEG variable	rhythm	model	task	mean_model	mean_eeg	sd_eeg	t	df	p
higher-level response	high meter contrast rhythm	broadband	pitch	1.06	0.84	0.23	-4.03	16	1.000
			tempo	1.06	0.78	0.39	-2.98	16	1.000
			visual	1.06	0.76	0.48	-2.63	16	1.000
		UREAR_AN	pitch	0.97	0.84	0.23	-2.27	16	1.000
			tempo	0.97	0.78	0.39	-1.96	16	1.000
			visual	0.97	0.76	0.48	-1.79	16	1.000
		UREAR_IC_BMF2	pitch	1.06	0.84	0.23	-4.02	16	1.000
			tempo	1.06	0.78	0.39	-2.97	16	1.000
			visual	1.06	0.76	0.48	-2.62	16	1.000
		UREAR_IC_BMF4	pitch	0.90	0.84	0.23	-1.08	16	1.000
			tempo	0.90	0.78	0.39	-1.28	16	1.000
			visual	0.90	0.76	0.48	-1.23	16	1.000
		UREAR_IC_BMF8	pitch	0.76	0.84	0.23	1.48	16	0.171
			tempo	0.76	0.78	0.39	0.20	16	0.748
			visual	0.76	0.76	0.48	-0.01	16	0.865
		UREAR_IC_BMF16	pitch	0.79	0.84	0.23	0.94	16	0.338
			tempo	0.79	0.78	0.39	-0.11	16	0.899
			visual	0.79	0.76	0.48	-0.27	16	0.967
		UREAR_IC_BMF32	pitch	0.91	0.84	0.23	-1.20	16	1.000
			tempo	0.91	0.78	0.39	-1.35	16	1.000
			visual	0.91	0.76	0.48	-1.29	16	1.000
		UREAR_IC_BMF64	pitch	0.94	0.84	0.23	-1.80	16	1.000
			tempo	0.94	0.78	0.39	-1.69	16	1.000
			visual	0.94	0.76	0.48	-1.57	16	1.000
	low meter contrast rhythm	broadband	pitch	0.13	0.51	0.40	4.01	16	0.002 **
			tempo	0.13	0.40	0.39	2.85	16	0.015 *
			visual	0.13	0.25	0.48	1.05	16	0.317
		UREAR_AN	pitch	0.14	0.51	0.40	3.86	16	0.002 **
			tempo	0.14	0.40	0.39	2.70	16	0.019 *
			visual	0.14	0.25	0.48	0.93	16	0.338
		UREAR_IC_BMF2	pitch	0.02	0.51	0.40	5.11	16	0.0002 ***
			tempo	0.02	0.40	0.39	3.95	16	0.002 **
			visual	0.02	0.25	0.48	1.96	16	0.077 .
		UREAR_IC_BMF4	pitch	-0.46	0.51	0.40	10.16	16	<0.0001 ***
			tempo	-0.46	0.40	0.39	9.03	16	<0.0001 ***
			visual	-0.46	0.25	0.48	6.14	16	<0.0001 ***
		UREAR_IC_BMF8	pitch	-0.65	0.51	0.40	12.05	16	<0.0001 ***
			tempo	-0.65	0.40	0.39	10.94	16	<0.0001 ***
			visual	-0.65	0.25	0.48	7.71	16	<0.0001 ***
		UREAR_IC_BMF16	pitch	-0.49	0.51	0.40	10.46	16	<0.0001 ***
			tempo	-0.49	0.40	0.39	9.33	16	<0.0001 ***
			visual	-0.49	0.25	0.48	6.38	16	<0.0001 ***
		UREAR_IC_BMF32	pitch	-0.08	0.51	0.40	6.18	16	<0.0001 ***
			tempo	-0.08	0.40	0.39	5.04	16	0.0002 ***
			visual	-0.08	0.25	0.48	2.85	16	0.015 *
		UREAR_IC_BMF64	pitch	0.13	0.51	0.40	3.98	16	0.002 **
			tempo	0.13	0.40	0.39	2.82	16	0.015 *
			visual	0.13	0.25	0.48	1.03	16	0.317

. P < 0.1, * P < 0.05, ** P < 0.01, *** P < 0.001, FDR corrected

Table 2.S3. Comparison of mean z-scored amplitude at meter-related frequencies between early auditory EEG responses and models of auditory subcortical processing using one-sample t-tests.

EEG variable	rhythm	model	task	mean_model	mean_eeg	sd_eeg	t	df	p
early auditory response	high meter contrast rhythm	broadband	pitch	0.81	0.41	0.39	-4.20	16	1.0
			tempo	0.81	0.42	0.42	-3.80	16	1.0
			visual	0.81	0.32	0.41	-5.00	16	1.0
		UREAR_AN	pitch	0.52	0.41	0.39	-1.15	16	1.0
			tempo	0.52	0.42	0.42	-0.93	16	1.0
			visual	0.52	0.32	0.41	-2.04	16	1.0
		UREAR_IC_BMF2	pitch	0.83	0.41	0.39	-4.38	16	1.0
			tempo	0.83	0.42	0.42	-3.98	16	1.0
			visual	0.83	0.32	0.41	-5.17	16	1.0
		UREAR_IC_BMF4	pitch	0.79	0.41	0.39	-4.00	16	1.0
			tempo	0.79	0.42	0.42	-3.61	16	1.0
			visual	0.79	0.32	0.41	-4.80	16	1.0
		UREAR_IC_BMF8	pitch	0.75	0.41	0.39	-3.60	16	1.0
			tempo	0.75	0.42	0.42	-3.24	16	1.0
			visual	0.75	0.32	0.41	-4.42	16	1.0
		UREAR_IC_BMF16	pitch	0.80	0.41	0.39	-4.09	16	1.0
			tempo	0.80	0.42	0.42	-3.69	16	1.0
			visual	0.80	0.32	0.41	-4.89	16	1.0
		UREAR_IC_BMF32	pitch	0.81	0.41	0.39	-4.25	16	1.0
			tempo	0.81	0.42	0.42	-3.85	16	1.0
			visual	0.81	0.32	0.41	-5.04	16	1.0
		UREAR_IC_BMF64	pitch	0.68	0.41	0.39	-2.89	16	1.0
			tempo	0.68	0.42	0.42	-2.56	16	1.0
			visual	0.68	0.32	0.41	-3.73	16	1.0
	low meter contrast rhythm	broadband	pitch	0.06	-0.10	0.40	-1.58	16	1.0
			tempo	0.06	-0.04	0.38	-1.03	16	1.0
			visual	0.06	-0.09	0.33	-1.84	16	1.0
		UREAR_AN	pitch	-0.13	-0.10	0.40	0.31	16	1.0
			tempo	-0.13	-0.04	0.38	0.95	16	1.0
			visual	-0.13	-0.09	0.33	0.45	16	1.0
		UREAR_IC_BMF2	pitch	-0.04	-0.10	0.40	-0.54	16	1.0
			tempo	-0.04	-0.04	0.38	0.06	16	1.0
			visual	-0.04	-0.09	0.33	-0.58	16	1.0
		UREAR_IC_BMF4	pitch	-0.29	-0.10	0.40	2.00	16	0.3
			tempo	-0.29	-0.04	0.38	2.72	16	0.2
			visual	-0.29	-0.09	0.33	2.49	16	0.2
		UREAR_IC_BMF8	pitch	-0.27	-0.10	0.40	1.79	16	0.4
			tempo	-0.27	-0.04	0.38	2.50	16	0.2
			visual	-0.27	-0.09	0.33	2.24	16	0.2
		UREAR_IC_BMF16	pitch	-0.08	-0.10	0.40	-0.13	16	1.0
			tempo	-0.08	-0.04	0.38	0.48	16	1.0
			visual	-0.08	-0.09	0.33	-0.09	16	1.0
		UREAR_IC_BMF32	pitch	0.05	-0.10	0.40	-1.53	16	1.0
			tempo	0.05	-0.04	0.38	-0.98	16	1.0
			visual	0.05	-0.09	0.33	-1.77	16	1.0
		UREAR_IC_BMF64	pitch	0.01	-0.10	0.40	-1.07	16	1.0
			tempo	0.01	-0.04	0.38	-0.50	16	1.0
			visual	0.01	-0.09	0.33	-1.22	16	1.0

. P < 0.1, * P < 0.05, ** P < 0.01, *** P < 0.001, FDR corrected

Table 2.S4. Control analysis of mean z-scored amplitude at meter-related frequencies without taking the highest frequency (5 Hz) into account. Comparison between early auditory EEG responses and models of auditory subcortical processing using one-sample t-tests.

EEG variable	rhythm	model	task	mean_model	mean_eeg	sd_eeg	t	df	p
early auditory response	high meter contrast rhythm	broadband	pitch	1.06	0.51	0.45	-5.07	16	1.000
			tempo	1.06	0.60	0.47	-4.07	16	1.000
			visual	1.06	0.31	0.68	-4.52	16	1.000
		UREAR_AN	pitch	0.97	0.51	0.45	-4.18	16	1.000
			tempo	0.97	0.60	0.47	-3.23	16	1.000
			visual	0.97	0.31	0.68	-3.93	16	1.000
		UREAR_IC_BMF2	pitch	1.06	0.51	0.45	-5.06	16	1.000
			tempo	1.06	0.60	0.47	-4.06	16	1.000
			visual	1.06	0.31	0.68	-4.51	16	1.000
		UREAR_IC_BMF4	pitch	0.90	0.51	0.45	-3.58	16	1.000
			tempo	0.90	0.60	0.47	-2.66	16	1.000
			visual	0.90	0.31	0.68	-3.54	16	1.000
		UREAR_IC_BMF8	pitch	0.76	0.51	0.45	-2.29	16	1.000
			tempo	0.76	0.60	0.47	-1.42	16	1.000
			visual	0.76	0.31	0.68	-2.68	16	1.000
		UREAR_IC_BMF16	pitch	0.79	0.51	0.45	-2.57	16	1.000
			tempo	0.79	0.60	0.47	-1.68	16	1.000
			visual	0.79	0.31	0.68	-2.86	16	1.000
		UREAR_IC_BMF32	pitch	0.91	0.51	0.45	-3.64	16	1.000
			tempo	0.91	0.60	0.47	-2.71	16	1.000
			visual	0.91	0.31	0.68	-3.58	16	1.000
		UREAR_IC_BMF64	pitch	0.94	0.51	0.45	-3.95	16	1.000
			tempo	0.94	0.60	0.47	-3.00	16	1.000
			visual	0.94	0.31	0.68	-3.78	16	1.000
	low meter contrast rhythm	broadband	pitch	0.13	-0.03	0.48	-1.39	16	1.000
			tempo	0.13	0.02	0.45	-0.99	16	1.000
			visual	0.13	-0.06	0.39	-1.94	16	1.000
		UREAR_AN	pitch	0.14	-0.03	0.48	-1.51	16	1.000
			tempo	0.14	0.02	0.45	-1.11	16	1.000
			visual	0.14	-0.06	0.39	-2.08	16	1.000
		UREAR_IC_BMF2	pitch	0.02	-0.03	0.48	-0.48	16	1.000
			tempo	0.02	0.02	0.45	-0.03	16	1.000
			visual	0.02	-0.06	0.39	-0.82	16	1.000
		UREAR_IC_BMF4	pitch	-0.46	-0.03	0.48	3.71	16	0.005 **
			tempo	-0.46	0.02	0.45	4.38	16	0.002 **
			visual	-0.46	-0.06	0.39	4.29	16	0.002 **
		UREAR_IC_BMF8	pitch	-0.65	-0.03	0.48	5.28	16	0.0006 ***
			tempo	-0.65	0.02	0.45	6.04	16	0.0002 ***
			visual	-0.65	-0.06	0.39	6.21	16	0.0002 ***
		UREAR_IC_BMF16	pitch	-0.49	-0.03	0.48	3.95	16	0.003 **
			tempo	-0.49	0.02	0.45	4.64	16	0.001 **
			visual	-0.49	-0.06	0.39	4.59	16	0.001 **
		UREAR_IC_BMF32	pitch	-0.08	-0.03	0.48	0.41	16	1.000
			tempo	-0.08	0.02	0.45	0.91	16	0.901
			visual	-0.08	-0.06	0.39	0.27	16	1.000
		UREAR_IC_BMF64	pitch	0.13	-0.03	0.48	-1.41	16	1.000
			tempo	0.13	0.02	0.45	-1.01	16	1.000
			visual	0.13	-0.06	0.39	-1.96	16	1.000

. P < 0.1, * P < 0.05, ** P < 0.01, *** P < 0.001, FDR corrected

2.6.2 EEG oscillatory alpha activity

Parieto-occipital alpha power was measured as an independent index of attentional engagement during the visual task, compared to the two auditory tasks. To separate the oscillatory activity from the $1/f$ background, the time-domain preprocessed data were subjected to irregular-resampling auto-spectral analysis (IRASA) (Wen and Liu, 2016). This method utilizes the fact that irregular resampling with non-integer factors results in shifts of the oscillatory component along the frequency axis, whereas the $1/f$ component remains constant. The procedure was carried out separately for each channel, condition, and participant. The data from each trial were segmented into 15 overlapping windows that were equally spaced throughout the trial. The number of samples in each window corresponded to the largest power of 2 that did not exceed 90% of trial duration. For each window, the auto-power spectrum was estimated using FFT after multiplication with a Hann function. This was performed for the original sampling rate, and also after resampling using pairs of resampling factors f and $1/f$ (where f was taken from 0.1 to 0.9 in steps of 0.05). The geometric mean of the auto-power spectra was taken across each pair of resampling factors. The power spectrum of the fractal component was estimated as the median-average spectrum across all values of f , separately for each window. The power spectrum of the fractal component was then averaged across the 15 windows and subtracted from the average power spectrum of the original signal (without any resampling), to obtain an estimate of the oscillatory component.

In the resulting oscillatory power spectra, all frequencies that were expected to contain neural activity elicited by the acoustic stimulus (i.e. harmonics of the pattern repetition rate, 0.416 Hz) were set to zero. Subsequently, the power in the alpha range was quantified by taking the mean power between 8 and 12 Hz (Iemi et al., 2017; van Diepen and Mazaheri, 2017; Van Diepen et al., 2019), separately for each condition. Alpha power was averaged across 18 parieto-occipital channels (Iz, O1, Oz, O2, PO7, PO3, POz, PO4, PO8, P9, P7, P3, P1, Pz, P2, P4, P8, P10) that were expected to show largest effects of cross-modal attention based on previous studies (Fu et al., 2001; Mazaheri et al., 2014; van Diepen and Mazaheri, 2017).

The power of alpha oscillatory activity from the parieto-occipital electrodes was significantly modulated by task ($F_{2,80} = 10$, $P = 0.0001$, $BF_{10} > 100$). As shown in Figure 2.S1, the power

significantly decreased during the Visual task compared to the Tempo task ($\beta = -0.03$, $t_{82} = -4.39$, $P = 0.0001$, 95% CI = [-0.05, -0.01]) and Pitch task ($\beta = -0.02$, $t_{82} = -3.08$, $P = 0.008$, 95% CI = [-0.04, -0.005]).

Visual inspection of the data suggested that the effect might have been only present for participants who had higher baseline alpha power. This was supported by a significant improvement of the model after adding the interaction between Task and alpha power in the Visual task as a continuous predictor ($F_{3,72.5} = 11.72$, $P < 0.0001$). Across participants, higher power in the Visual-task condition was related to greater power increase in the Tempo task ($\beta = 1.41$, $t_{45.3} = 5.41$, $P < 0.0001$, 95% CI = [0.73, 1.98]) and in the Pitch task ($\beta = 1.16$, $t_{45.3} = 4.47$, $P < 0.0001$, 95% CI = [0.49, 1.73]).

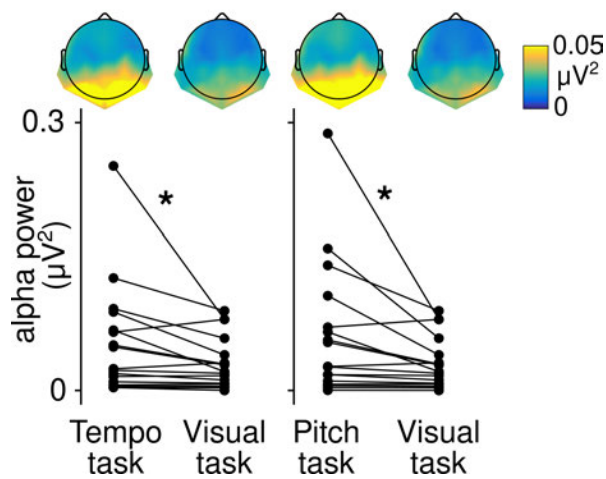


Figure 2.S1. EEG power at the alpha frequency elicited across the different tasks at posterior channels. Alpha power was significantly smaller ($p < 0.01$, marked by asterisks) during the visual task, and the magnitude of the effect depended on the baseline alpha response across participants.

3 Study 2: Neural tracking of the musical beat is enhanced by low-frequency sounds

This study has been published in the Proceedings of the National Academy of Sciences in 2018.

The results of Study 1 raised the question of whether there are factors that could further boost the enhancement of contrast at meter periodicities in the brain response. To this end, Study 2 explored the role of spectral content of the sounds delivering the rhythmic information. This was inspired by the observation that music composers and performers often preferentially use bass sounds to carry rhythmic foundations of music and make people dance. The results showed that a rhythmic input conveyed by bass sounds leads to enhanced contrast at meter periodicities in the neural response, compared to an identical input delivered by high-pitched sounds. The fact that this enhanced transformation was observed only when the sound input lacked prominent contrast at meter periodicities suggested that the results cannot be simply explained by fixed low-level physiological mechanisms that passively enhance specific modulation frequencies in the input. Instead, these results suggest that bass sounds engage higher-level mechanisms that flexibly transform the input when it lacks prominent contrast at meter periodicities.

The reader may notice subtle differences in the way the terms “beat” and “meter” are used in Study 2 and their definitions proposed in section 1.1. Particularly, Study 2 defines beat as the primary pulse whereas meter is defined as resulting from subdivisions and groupings of the beat. Moreover, part of the analysis treats the “beat pulse” separately from the whole meter, and does not include harmonics of this pulse when measuring its representation in different signals. The reason for these inconsistencies is that my thinking about meter has evolved over time, and the manuscript for Study 2 was written up and published early in my candidature. However, importantly, the main part of the analysis (where all meter-related frequencies are taken into account) is consistent with the way frequency-tagging approach is described in section 1.2.2, and the discussion and conclusions of the study are in line with the concepts developed in section 1 of the current thesis.

3.1 Abstract

Music makes us move, and using bass instruments to build the rhythmic foundations of music is especially effective at inducing people to dance to periodic pulse-like beats. Here, we show that this culturally widespread practice may exploit a neurophysiological mechanism whereby low-frequency sounds shape the neural representations of rhythmic input by boosting selective locking to the beat. Cortical activity was captured using

electroencephalography while participants listened to a regular rhythm or a relatively complex syncopated rhythm conveyed either by low tones (130 Hz) or high tones (1236.8 Hz). We found that cortical activity at the frequency of the perceived beat is selectively enhanced compared to other frequencies in the EEG spectrum when rhythms are conveyed by bass sounds. This effect is unlikely to arise from early cochlear processes, as revealed by auditory physiological modeling, and was particularly pronounced for the complex rhythm requiring endogenous generation of the beat. The effect is likewise not attributable to differences in perceived loudness between low and high tones, as a control experiment manipulating sound intensity alone did not yield similar results. Finally, the privileged role of bass sounds is contingent on allocation of attentional resources to the temporal properties of the stimulus, as revealed by a further control experiment examining the role of behavioral task. Together, our results provide a neurobiological basis for the convention of using bass instruments to carry the rhythmic foundations of music and drive people to move to the beat.

3.2 Significance Statement

Bass sounds play a special role in conveying the rhythm and stimulating motor entrainment to the beat of music. However, the biological roots of this culturally widespread musical practice remain mysterious, despite its fundamental relevance in the sciences and arts, and also for music-assisted clinical rehabilitation of motor disorders. Here, we show that this musical convention may exploit a neurophysiological mechanism whereby low-frequency sounds shape neural representations of rhythmic input at the cortical level by boosting selective neural locking to the beat, thus explaining the privileged role of bass sounds in driving people move along with the musical beat.

3.3 Introduction

Music powerfully compels humans to move, showcasing our remarkable ability to perceive and produce rhythmic signals (Phillips-Silver et al., 2010). Rhythm is often considered the most basic aspect of music, and is increasingly regarded as a fundamental organizing principle of brain function. Yet, the neurobiological mechanisms underlying entrainment to musical rhythm remain unclear, despite the broad relevance of the question in the sciences and arts. Clarifying these mechanisms is also timely given the growing interest in music-assisted practices for the clinical rehabilitation of cognitive and motor disorders caused by brain damage (Thaut et al., 2015).

Typically, people are attracted to move to music in time with a periodic pulse-like beat, for example by bobbing the head or tapping the foot to the beat of the music. The perceived beat and meter (i.e., hierarchically nested periodicities corresponding to grouping or subdivision of the beat period) are thus used to organize and predict the timing of incoming rhythmic input (Essens and Povel, 1985) and to guide synchronous movement (Toiviainen et al., 2010). Notably, the perceived beats sometimes coincide with silent intervals instead of accented acoustic events, as in syncopated rhythms (a hallmark of jazz), revealing remarkable flexibility with respect to the incoming rhythmic input in human perceptual-motor coupling (Large et al., 2015). However, specific acoustic features such as bass sounds seem particularly well suited to convey the rhythm of music and support rhythmic motor entrainment (Hove et al., 2007; Burger et al., 2018). Indeed, in musical practice, bass instruments are conventionally used as a rhythmic foundation, whereas high-pitched instruments carry the melodic content (Lerdahl and Jackendoff, 1983; Trainor et al., 2014). Bass sounds are also crucial in music that encourages listeners to move (Pressing, 2002; Stupacher et al., 2016).

There has been a recent debate as to whether evolutionarily shaped properties of the auditory system lead to superior temporal encoding for bass sounds (Hove et al., 2014; Wojtczak et al., 2017). One study using electroencephalography (EEG) recorded brain responses elicited by misaligned tone onsets in an isochronous sequence of simultaneous low- and high-pitched tones (Hove et al., 2014). Greater sensitivity to the temporal misalignment of low tones was observed when they were presented earlier than expected, which suggested better time encoding for low sounds. These results were replicated and

extended by Wojtczak et al. (2017), who showed that the effect was related to greater tolerance for low-frequency sounds lagging high-frequency sounds than vice versa. However, these studies only provide indirect evidence for the effect of low sounds on internal entrainment to rhythm, inferred from brain responses to deviant sounds. Moreover, they stop short of resolving the issue because, as noted by Wojtczak et al. (2017), these brain responses are not necessarily informative about the processing of the global temporal structure of rhythmic inputs.

A promising approach to capture the internal representations of rhythm more directly involves the combination of EEG with frequency-tagging. This approach involves measuring brain activity elicited at frequencies corresponding to the temporal structure of the rhythmic input (Regan, 1989; Picton et al., 2003; Nozaradan, 2014; Nozaradan et al., 2017a). In a number of studies using this technique, an increase in brain activity has been observed at specific frequencies corresponding to the perceived beat and meter of musical rhythms (Nozaradan et al., 2011, 2012, 2016a; Stupacher et al., 2017; Tal et al., 2017). Evidence for the functional significance of this neural selectivity comes from work showing that the magnitude of beat- and meter-related brain responses correlates with individual differences in rhythmic motor behavior and is modulated by contextual factors influencing beat and meter perception (Chemin et al., 2014; Cirelli et al., 2016; Nozaradan et al., 2016b, 2017b).

The current study aimed to use this approach to provide first evidence for a privileged effect of bass sounds in the neural processing of rhythm, especially in boosting cortical activity at beat- and meter-related frequencies. The EEG was recorded from human participants while they listened to isochronous and non-isochronous rhythms conveyed either by low (130 Hz) or high (1236.8 Hz) pure tones. The isochronous rhythm provided a baseline test of neural entrainment. The non-isochronous rhythms, which included a regular unsyncopated and a relatively complex syncopated rhythm, contained combinations of tones and silent intervals positioned to imply hierarchical metric structure. The unsyncopated rhythm was expected to induce the perception of a periodic beat that corresponded closely to the physical arrangement of sound onsets and silent intervals making up the rhythm, whereas the syncopated rhythm was expected to induce the perception of a beat that matched physical cues to a lesser extent, thus requiring more endogenous generation of the beat and meter (Nozaradan et al., 2012, 2016b, 2016a, 2017b). Theoretical beat periods were confirmed in a tapping session conducted after the EEG session, in which participants were asked to tap

along with the beat that they perceived in the rhythms, as a behavioral index of entrainment to the beat. Importantly, these different rhythms allowed us to test the low tone benefit on cortical activity at the beat and meter frequencies even when the input lacked prominent acoustic energy at these frequencies. Frequency-domain analysis of the EEG was performed to obtain a direct fine-grained characterization of the mapping between rhythmic stimulus and EEG response. Additional analyses including auditory physiological modeling were performed to examine the degree to which the effect of bass sounds may be explained by cochlear properties.

3.4 Results

Sound analysis. The isochronous rhythm and the two non-isochronous (unsyncopated and syncopated) rhythms carried by low or high tones were analyzed using a cochlear model to (i) determine frequencies that could be expected in the EEG response, and (ii) estimate an early representation of the sound input. This model consisted of a gammatone auditory filterbank that converted acoustic input into a multi-channel representation of basilar membrane motion (Patterson and Holdsworth, 1996), followed by a simulation of hair cell dynamics. The model yields an estimate of spiking responses in the auditory nerve with rate intensity functions and adaptation closely following neurophysiological data (Meddis, 1986). The envelope modulation spectrum obtained for the isochronous rhythms consisted of a peak at the frequency of single events (5 Hz) and its harmonics. For the unsyncopated and syncopated rhythms, the obtained envelope spectra contained 12 distinct peaks, corresponding to the repetition frequency of the whole pattern (0.416 Hz) and its harmonics up to the frequency of repetitions of single events (5 Hz) (see Fig. 3.1). The magnitudes of responses at these 12 frequencies were converted into z-scores (see Materials and Methods). This standardization procedure allowed the magnitude at each frequency to be assessed relative to the other frequencies, and, thereby, allowed us to determine how much one (here, the beat frequency at 1.25 Hz) or a subgroup of frequencies (meter-related frequencies at 1.25, $1.25/3$, 1.25×2 , 1.25×4 Hz; see Materials and Methods) stood out prominently relative to the entire set of frequencies (see, e.g., 18). This procedure has the further advantage of offering the possibility to objectively measure the degree of relative

transformation, i.e., the distance between an input (corresponding to the cochlear model) and an output (corresponding to the obtained EEG responses), irrespective of the difference in their unit and scale (Nozaradan et al., 2012, 2017a, 2017b, 2018).

Behavioral tasks. During the EEG session, participants were asked to detect and identify deviant tones with duration lengthened or shortened by 20% (40 ms), to encourage attentive listening specifically to the temporal structure of the auditory stimuli. While participants were generally able to identify the temporal deviants (see SI Appendix, Table 3.S1), there was a significant interaction between rhythm and tone frequency [$F(2,26) = 3.55$, $P = 0.04$, $\eta_G^2 = 0.04$]. Post-hoc t-tests revealed significantly lower performance in the low-tone compared to high-tone syncopated rhythm [$t(13) = 2.97$, $P = 0.03$, $d = 0.79$], suggesting higher task difficulty especially for the syncopated rhythm delivered with low tones (Moore et al., 1993; Grube and Griffiths, 2009).

The beat-tapping task performed after the EEG session (SI Appendix) generally confirmed the theoretical assumption about entrainment to the beat based on preferential grouping by four events (Povel and Essens, 1985; Nozaradan et al., 2012, 2016b). Moreover, there were no statistically significant differences in the mean inter-tap interval and its variability between conditions (SI Appendix).

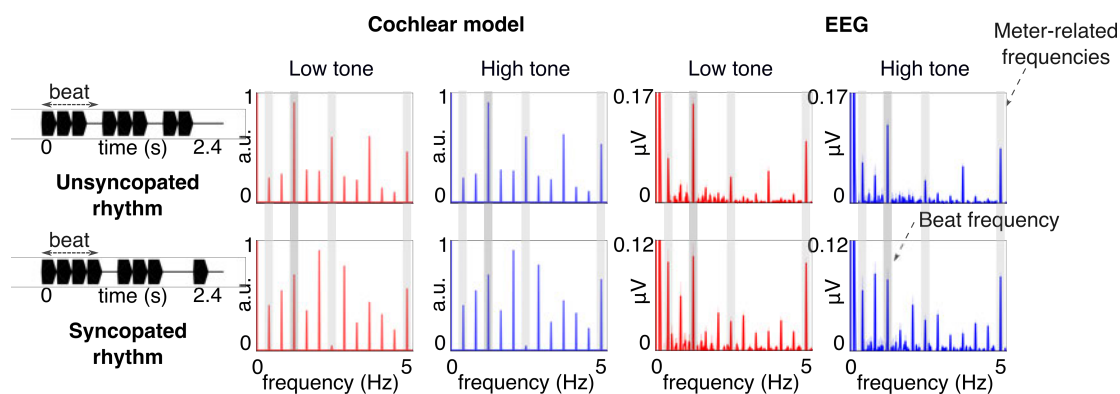


Figure 3.1. Spectra of the acoustic stimuli (processed through the cochlear model) and EEG responses (averaged across all channels and participants; $n = 14$; shaded regions indicate SEMs; see Morey, 2008). The waveform of one cycle for each rhythm (2.4-s duration) is depicted in black (Left) with the beat period indicated. The rhythms were continuously repeated to form 60-s sequences, and these sequences were presented eight times per condition. The cochlear model spectrum contains peaks at frequencies related to

the beat (1.25 Hz; dark-gray vertical stripes) and meter (1.25 Hz/3, $\times 2$, $\times 4$; light-gray vertical stripes), and also at frequencies unrelated to the beat and meter. The EEG response includes peaks at the frequencies contained in the cochlear model output; however, the difference between the average amplitude of peaks at frequencies related vs. unrelated to the beat and meter is increased in the low-tone compared with high-tone conditions (see Relative Enhancement at Beat and Meter Frequencies and Fig. 3.2). Note the scaling difference in plots of EEG responses for unsyncopated and syncopated rhythms.

Frequency-domain analysis of EEG. As shown in Figure 3.1, the rhythmic stimuli elicited frequency-tagged EEG responses at the 12 frequencies expected based on the results of sound analysis with the cochlear model, with topographies similar to previous work (Nozaradan et al., 2012, 2016b; see Fig. 3.2).

Overall magnitude of the EEG response. We first evaluated whether the overall magnitude (μV) of the EEG responses differed between low- and high-tone conditions for the three different rhythms (SI Appendix, Table 3.S1). Overall magnitude was computed for each participant and condition by summing the amplitude of the frequency-tagged responses across the frequencies that we expected to be elicited based on the sound analysis of the rhythms. The resultant measure of overall response magnitude provides an index of the general capacity of the central nervous system to respond to the rhythms and the modulation of this capacity by tone frequency, regardless of the relevance of frequency-tagged components related to the beat and meter. For the isochronous rhythm, there was a significantly larger overall response in the low-tone condition [$t(13) = 3.68$, $P = 0.008$, $d = 0.98$], in line with previous work using isochronous trains of tones or sinusoidally amplitude-modulated tones (Wunderlich and Cone-Wesson, 2001; Ross et al., 2003). In contrast, there were no significant differences between the high-tone and low-tone condition for the unsyncopated and syncopated rhythm ($P_s > 0.34$). This suggests that the global enhancement of the responses by low tones might only be present for isochronous rhythms.

Relative enhancement at beat and meter frequencies. The main goal of the study was to examine the relative amplitude at specific beat- and meter-relevant frequency components for the unsyncopated and syncopated rhythms conveyed by high or low tones. A 2x2 repeated-measures ANOVA revealed greater relative amplitude at the beat frequency (z-score of the amplitude at 1.25 Hz) in the low-tone condition for both types of rhythm (main effect of tone frequency, Fig. 3.3 top) [$F(1,13) = 9.46$, $P = 0.009$, $\eta_G^2 = 0.11$] (see Materials

and Methods and SI Appendix for tests of the validity of this standardization procedure). Furthermore, when all meter frequencies were taken into account (mean z-scored amplitude at 1.25, 1.25/3, 1.25x2 and 1.25x4 Hz), the 2x2 ANOVA revealed a significant interaction between tone frequency and rhythm [$F(1,13) = 5.23$, $P = 0.04$, $\eta_G^2 = 0.05$], indicating greater relative amplitude at meter frequencies (Fig. 3.3 bottom) in the low-tone condition for the syncopated rhythm [$t(13) = 3.79$, $P = 0.004$, $d = 1.01$], but not for the unsyncopated rhythm ($P = 0.24$, $d = 0.45$).

Finally, we evaluated the extent to which early cochlear processes could potentially explain the relative enhancement of EEG response at the beat and meter frequencies elicited by low tones. The difference in z-scored EEG response amplitudes between the low- and high-tone conditions was compared with the corresponding difference in the z-scored cochlear model output. For the response at the beat frequency, difference scores were significantly larger in the EEG response compared with the cochlear model for the syncopated rhythm [$t(13) = 1.4$, $P = 0.04$, $d = 0.8$], but not for the unsyncopated rhythm ($P = 0.73$, $d = 0.38$). A similar pattern was revealed for the mean response at meter-related frequencies, with a significantly greater difference score for the syncopated rhythm [$t(13) = 4.17$, $P = 0.004$, $d = 1.11$], but not for the unsyncopated rhythm ($P = 0.28$, $d = 0.53$).

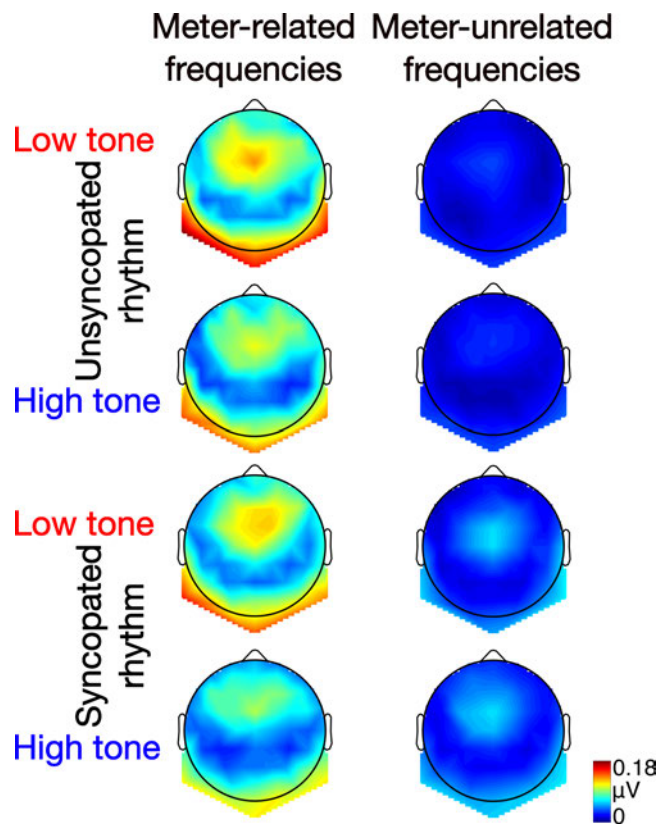


Figure 3.2. Grand average topographies ($n = 14$) of neural activity measured at meter-related (left column) and meter-unrelated (right column) frequencies for the unsyncopated and syncopated rhythm conveyed by low or high tones.

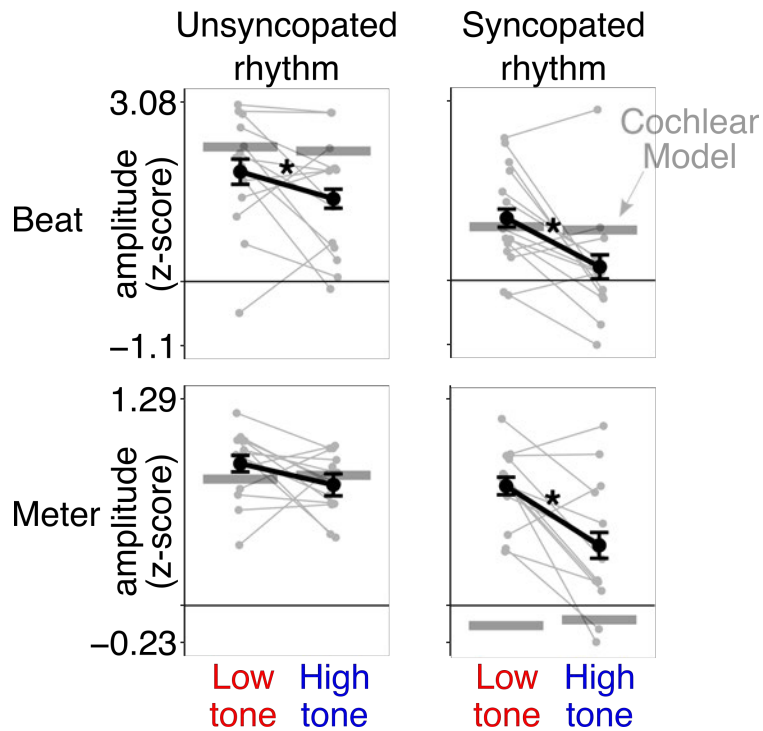


Figure 3.3. Effect of tone frequency on the selective enhancement of EEG activity at beat- and meter-related frequencies. Shown separately are z-scores for the beat frequency (top) and mean z-scores for meter-related frequencies (bottom) averaged across participants are shown separately for the unsyncopated (left) and syncopated (right) rhythm. Error bars indicate SEM (Morey, 2008). Asterisks indicate significant differences ($P < 0.05$). Responses from individual participants are shown as grey points linked by lines. The horizontal lines represent z-score values obtained from the cochlear model. The low tone led to significant neural enhancement of the beat frequency in both rhythms. The low tone also elicited enhanced EEG response at meter frequencies, but only in the syncopated rhythm. There was no significant modulation of meter-related responses by tone frequency for the unsyncopated rhythm.

3.5 Discussion

The results show that rhythmic stimulation by bass sounds leads to enhanced neural representation of the beat and meter. EEG and behavioral responses were collected from participants presented with auditory rhythms conveyed either by low- or high-frequency tones. As hypothesized, we observed a selective enhancement of neural activity at the beat frequency for rhythms conveyed by low tones compared to high tones. When taking into consideration all meter frequencies, this low tone benefit was only significant for the syncopated rhythm requiring relatively more endogenous generation of meter. Moreover,

the low tone benefit was not attributable to differences in perceived loudness between low and high tones, as a control experiment manipulating sound intensity alone did not yield similar results (SI Appendix, Control Experiment 1: Effect of Sound Intensity). Finally, the low tone benefit appears to require the allocation of attention to temporal features of the stimulus, as the effect did not occur in a control experiment where attention was directed to non-temporal features (SI Appendix, Control Experiment 2: Effect of Behavioral Task).

It has been proposed that the preference for low-register instruments in conveying the rhythm within multi-voiced music derives from more precise temporal encoding of low-pitched sounds due to masking effects in the auditory periphery (Hove et al., 2014). An alternative account holds that lower-pitched sounds are not necessarily subject to more precise temporal encoding but to greater tolerance for perception of simultaneity when low sounds lag behind higher-pitched sounds (Wojtczak et al., 2017). Here we demonstrate that the privileged status of rhythmic bass sounds is not exclusively dependent on multi-voice factors, such as masking or perceptual simultaneity, as it can be observed even without the presence of other instrumental voices.

Analysis of our stimuli with a physiologically plausible cochlear model (Meddis, 1986; Patterson and Holdsworth, 1996) indicated that nonlinear processes at the early stages of the auditory pathway are unlikely to account for the observed low tone benefit and the interaction with syncopation. Moreover, the effect is not explained by greater activation of auditory neurons due to greater loudness, as the intensity of low and high tones was adjusted to evoke similar loudness percepts (Moore et al., 2016). Possible residual loudness differences between low and high tones are also unlikely to account for the observed effect of tone frequency. This was confirmed in a control experiment (SI Appendix, Control Experiment 1: Effect of Sound Intensity) showing that manipulating sound intensity alone (70 vs. 80 dB sound pressure level) did not influence the neural representation of the beat and meter.

Instead, the low tone benefit observed here could be explained by a greater recruitment of brain structures involved in movement planning and control, including motor cortical regions (Kung et al., 2013; Patel and Iversen, 2014; Merchant et al., 2015a; Todd and Lee, 2015b; Burunat et al., 2017; Morillon and Baillet, 2017), the cerebellum, and basal ganglia (Nozaradan et al., 2017b). These structures may be recruited via functional interconnections

between the ascending auditory pathway and a vestibular sensory-motor network (including a striatal circuit involved in beat-based processing) that is particularly responsive to bass acoustic frequencies (Trainor and Unrau, 2009; Trainor et al., 2009; Todd and Lee, 2015a). The involvement of these sensory-motor areas thus constitutes a plausible mechanism for the observed low tone benefit, as these areas have been shown to be critically involved in predictive beat perception (Patel and Iversen, 2014), to contribute to the selective enhancement of EEG responses at the beat frequency (Nozaradan et al., 2017b), and to be activated by vestibular input (Trainor et al., 2009; Todd and Lee, 2015a). It should be noted that direct activation of the human vestibular organ by bass sounds occurs only at higher intensities (above ~95 dB sound pressure level) than those employed in the current study (Todd et al., 2000). However, functional interactions between auditory and vestibular sensory-motor networks in response to low-frequency rhythms can arise centrally (Trainor and Unrau, 2009). These neural connections presumably develop from the onset of hearing in the fetus through the continuous experience of correlated auditory and vestibular sensory-motor input (e.g., the sound of the mother's footsteps coupled with walking motion) (Trainor and Unrau, 2009).

3.5.1 Low-tone benefit in syncopated rhythm

In accordance with previous studies (Wunderlich and Cone-Wesson, 2001; Ross et al., 2003), overall larger magnitudes of EEG response were obtained with low tones compared with high tones in the isochronous rhythm. This general effect was not observed in non-isochronous rhythms, suggesting that as the stimulus becomes temporally more complex there is no longer a simple relationship between the overall response magnitude and tone frequency. Therefore, to fully capture the effect of tone frequency on the neural activity to complex rhythms, higher-level properties of the stimulus such as onset structure, which plays a role in inducing the perception of beat and meter (Povel and Essens, 1985), need to be taken into account. This was achieved here through a finer-grained frequency analysis focused on neural activity elicited at beat- and meter-related frequencies. Enhanced activity at the beat frequency was observed with low tones, irrespective of the rhythmic complexity of the stimulus. However, when taking into consideration all meter frequencies, neural

activity was enhanced with low tones only in the syncopated rhythm, whose envelope did not contain prominent peaks of energy at meter-related frequencies.

These findings corroborate the hypothesis that bass sounds stimulate greater involvement of top-down, endogenous processes, possibly via stronger engagement of motor brain structures (Kung et al., 2013; Patel and Iversen, 2014; Todd and Lee, 2015b, 2015a). Activation of a widely distributed sensory-motor network may thus have facilitated the selective neural enhancement of meter relevant frequencies in the current study, especially when listening to the low-tone syncopated rhythm, which relies heavily on endogenous processes. This association between rhythmic syncopation, low-frequency tones, and recruitment of a sensory-motor network, could explain why musical genres specifically tailored to induce a strong urge to move to the beat (i.e., groove-based music such as funk) often contain a syncopated bass line (e.g. Witek, 2017). Accordingly, syncopation is perceived as more prominent when produced by a bass drum than a hihat cymbal (Witek et al., 2014a), and rhythmically complex bass lines are rated as increasingly likely to make people dance (Wesolowski and Hofmann, 2016).

3.5.2 Critical role of temporal attention

The greater involvement of endogenous processes in the syncopated rhythm carried by low tone could be driven by an increase in endogenously generated predictions or attention to stimulus timing necessitated by carrying out the temporal deviant identification task. An internally generated periodic beat constitutes a reference used to encode temporal intervals in the incoming rhythmic stimulus (Essens and Povel, 1985; Povel and Essens, 1985; Large and Jones, 1999), and therefore contributes to successful identification of changes in tone duration (Bergeson and Trehub, 2006; Grube and Griffiths, 2009). In the current study, such an endogenous mechanism might have been especially utilized in the condition where the rhythmic structure of the input matched the beat and meter percept to a lesser extent (Povel and Essens, 1985; Grube and Griffiths, 2009), and where low tones made the identification of fine duration changes more difficult due to lower temporal resolution of the auditory system with low-frequency tones (Moore et al., 1993). The relative contribution of these endogenous temporal processes to the low tone benefit observed here was addressed in SI Appendix Control Experiment 2: Effect of Behavioral

Task, where participants were instructed to detect and identify any changes in broadly defined sound properties (pitch, tempo and loudness) when in fact none were present. This experiment yielded no effect of tone frequency, suggesting that the low tone benefit only occurred when the behavioral task required attention to be focused on temporal properties of the stimulus. Hence, even though a widespread sensory-motor network supporting endogenous meter generation can be directly activated by bass sounds when the intensities exceed the vestibular threshold (Todd and Lee, 2015), attending to temporal features of the sound is critical to the low tone benefit at intensities beneath the vestibular threshold. Similarly, the association between temporal attention and vestibular sensory-motor activation may occur in music and dance contexts aimed at encouraging people to move to music. Indeed, the intention to move along with a stimulus is likely to direct attention to stimulus timing, and the resulting body movement, in turn, enhances vestibular activation.

3.5.3 Conclusion

The present study provides direct evidence for selective brain processing of musical rhythm conveyed by bass sounds, thus furthering our understanding of the neurobiological bases of rhythmic entrainment. We propose that the selective increase in cortical activity at beat- and meter-related frequencies elicited by low tones may explain the special role of bass instruments for delivering rhythmic information and inducing sensory-motor entrainment in widespread musical traditions. Our findings also pave the way for future investigations of how acoustic frequency content, combined with other features such as timbre and intensity, may efficiently entrain neural populations by increasing functional coupling in a distributed auditory-motor network. A fruitful avenue for probing this network further is through techniques with greater spatial resolution, such as human intracerebral recordings (Nozaradan et al., 2016a). Ultimately, identifying sound properties that enhance neural tracking of the musical beat is timely, given the growing use of rhythmic auditory stimulation for the clinical rehabilitation of cognitive and motor neurological disorders (Hove and Keller, 2015).

3.6 Materials and methods

Participants. Fourteen healthy individuals (mean age = 28.4, SD = 6.1; 10 females) with various levels of musical training (mean = 6.9, SD = 5.6, range 0-14 years) participated in the study after providing written informed consent. All participants reported normal hearing and no history of neurological or psychiatric disease. The study was approved by the Research Ethics Committee of Western Sydney University.

Auditory stimuli. The auditory stimuli were created in Matlab R2016b (MathWorks, USA) and presented binaurally through insert earphones with an approximately flat frequency response over the range of frequencies included in the stimuli (ER-2, Etymotic Research). The stimuli consisted of three different rhythms of 2.4 s duration looped continuously for 60 s. All rhythms comprised 12 events, each individual event lasting 200 ms. The structure of each rhythm was based on a specific patterning of sound (pure tones; 10 ms rise and 50 ms fall linear ramps) and silent events (amplitude at 0), as depicted in Figure 3.1. The carrier frequency of the pure tone was either 130 Hz (low-tone frequency) or 1236.8 Hz (high-tone frequency; 39 semitones higher than the low-tone frequency). These two frequencies were chosen to fall within spectral bands where rhythmic fluctuations either correlate (100-200 Hz) or do not correlate (800-1600 Hz) with sensory-motor entrainment to music, as indicated by previous research (Burger et al., 2012, 2018). To take into account the differential sensitivity of the human auditory system across the frequency range, the loudness of low and high tones was equalized to 70 phons using the time-varying loudness model of Glasberg and Moore (2016; by matching the maximum short-term loudness of a single 200-ms high-tone and low-tone sound), and held constant across participants.

One stimulus rhythm consisted of an isochronous train of tones with no silent events. The two other rhythms were selected based on previous evidence that they induce a periodic beat based on grouping by 4 events (i.e., $4 \times 200 \text{ ms} = 800 \text{ ms} = 1.25 \text{ Hz}$ beat frequency) (Nozaradan et al., 2012, 2016b, 2018). Related metric levels corresponded to subdivisions of the beat period by 2 (2.5 Hz) and 4 (i.e., 200-ms single event = 5 Hz), and grouping of the beat period by 3 (i.e., 2.4-s rhythm = 0.416 Hz). One rhythm was designed to be unsyncopated, as a sound event coincided with every beat in almost all possible beat positions (syncopation score = 1; calculated as in Longuet-Higgins and Lee, 1984). The internal representation of beat and meter should thus match physical cues in this rhythm.

The other rhythm was syncopated, as it involved some beat positions coinciding with silent events rather than sound events (syncopation score = 4). The internal representation of beat and meter should thus match external cues to a lesser extent than in the unsyncopated rhythm.

Experimental design and procedure. Crossing tone frequency (low, high) and rhythm (isochronous, unsyncopated, syncopated) yielded six conditions that were presented in separate blocks. The order of the 6 blocks was randomized, with the restriction that at least one high-tone and one low-tone block occurred within the first three blocks. These blocks were presented in an EEG session followed by a tapping session, with the same block order for the two sessions. Each block consisted of 8 trials in the EEG session (2-4 s silence, followed by the 60-s stimulus), and 2 trials in the tapping session.

EEG session and behavioral task. In each trial of the EEG session, the duration of the steady-state portion of one randomly chosen sound event was either increased or decreased by 20% (40 ms), yielding 4 “longer” and 4 “shorter” deviant tones in each block. Participants were asked to detect the deviant tone and report after each trial whether it was longer or shorter than other tones comprising the rhythm. These deviants could appear only in the 3 repetitions of the rhythm prior to the last repetition, and were restricted to 3 possible positions within each rhythm. In the unsyncopated and syncopated rhythms, these positions corresponded to sound events directly followed by a silent event. This was done to minimize the differences in task difficulty between unsyncopated and syncopated rhythms, as the perception of duration might differ according to the context in which a deviant tone appears (i.e. whether it is preceded and followed by tones or silences). For the isochronous rhythm, 3 random positions were chosen. The last 4 repetitions of the rhythms of all trials were excluded from further EEG analyses. The primary purpose of the deviant identification task was to ensure that participants were attending to the temporal properties of the auditory stimuli. To test whether the difficulty of deviant identification varied across conditions, percent-correct responses were compared using a repeated-measures ANOVA with factors tone frequency (low, high) and rhythm (isochronous, unsyncopated, syncopated).

Participants were seated in a comfortable chair and asked to avoid any unnecessary movement or muscle contraction, and to keep their eyes fixated on a marker displayed on

the wall ~1 m in front of them. Examples of the “longer” and “shorter” deviant tones were provided before the session to ensure that all participants understood the task.

Stimulus sound analysis with cochlear model. The cochlear model used to analyze the stimuli applied a Patterson-Holdsworth ERB filter bank with 128 channels (Patterson and Holdsworth, 1996), followed by Meddis’ inner hair-cell model (Meddis, 1986), as implemented in the Auditory Toolbox for Matlab (Slaney, 1998). The output of the cochlear model was subsequently transformed into the frequency-domain using the fast Fourier transform (FFT) and averaged across channels. For the unsyncopated and syncopated rhythms, the magnitudes obtained from the resultant modulation spectrum were then expressed as z-scores, as follows: $(x - \text{mean across the 12 frequencies}) / \text{SD across the 12 frequencies}$ (Nozaradan et al., 2012, 2017b, 2018; Chemin et al., 2014; Cirelli et al., 2016).

EEG acquisition and preprocessing. The EEG was recorded using a Biosemi Active-Two system (Biosemi, Amsterdam, Netherlands) with 64 Ag-AgCl electrodes placed on the scalp according to the international 10/20 system. The signals were referenced to the CMS (Common Mode Sense) electrode and digitized at 2048 Hz sampling rate. Details of EEG data preprocessing are in SI Appendix. The cleaned EEG data were segmented from 0 to 50.4 s relative to the trial onset (i.e., exactly 21 repetitions of the rhythm, thus excluding repetitions of the rhythm where the deviant tones could appear), re-referenced to the common average, and averaged across trials in the time-domain separately for each condition and participant (Nozaradan et al., 2011, 2017b). The EEG preprocessing was carried out using Letswave6 (<http://www.notions.org/letswave>) and Matlab. Further statistical analyses were carried out using R (version 3.4.1, <https://www.R-project.org>), with Greenhouse-Geisser correction applied when the assumption of sphericity was violated and Bonferroni-corrected post-hoc tests to further examine significant effects.

Frequency-domain analysis of EEG responses. For each condition and participant, the obtained averaged waveforms were transformed into the frequency-domain using FFT, yielding a spectrum of signal amplitudes (in μV) ranging from 0 to 1024 Hz, with a frequency resolution of 0.0198 Hz. Within the obtained frequency spectra, the signal amplitude can be expected to correspond to the sum of (i) EEG responses elicited by the stimulus, and (ii) unrelated residual background noise. To obtain valid estimates of the responses, the contribution of noise was minimized by subtracting, at each frequency bin, the average amplitude at the neighboring bins (2nd to 5th on both sides) (Mouraux et al., 2011;

Nozaradan et al., 2012). For each condition and participant, noise-subtracted spectra were then averaged across all channels to avoid electrode-selection bias (as the response may originate from a widespread cortical network), and to account for individual differences in response topography (Tierney and Kraus, 2014; Nozaradan et al., 2016b). The noise-subtracted, channel-averaged amplitudes at the expected frequencies (based on sound analysis with the cochlear model) in response to each stimulus were then measured for each condition and participant at the exact frequency bin of each expected response (note that the length of the analyzed epochs contained an integer number of rhythm cycles, so that the frequency bins were centered exactly at the frequencies of the expected responses).

Overall magnitude of the EEG response. The overall magnitude of the EEG response in each condition was measured as the sum of amplitudes at the 12 frequencies expected in response to the unsyncopated and syncopated rhythms, and at 5 Hz and harmonics for the isochronous rhythm (only harmonics up to 45 Hz with amplitude significantly above 0 μ V in the noise-subtracted EEG spectra were considered for each condition). Significance of the harmonics was assessed using the non-subtracted amplitude spectra, averaged over all electrodes and participants (Rossion et al., 2015). Responses were tested by z-scoring the amplitude at each harmonic with a baseline defined as 20 neighboring bins (second - 11th on each side), using the formula $z(x) = (x - \text{baseline mean})/\text{baseline SD}$. Using this test, eight successive harmonics were considered significant for the low-tone and nine for the high-tone isochronous condition, as they had z-scores greater than 2.32 (i.e. $p < 0.01$, one-sample one-tailed t-test, testing signal > noise). To test whether the overall response was enhanced when the same rhythm was conveyed by low vs. high tones, three separate paired-samples t-tests were conducted on the isochronous, unsyncopated and syncopated rhythm.

Relative amplitude at beat and meter frequencies. To assess the relative prominence of the specific frequencies in the EEG response to the unsyncopated and syncopated rhythm, amplitudes at the 12 expected frequencies elicited by each rhythm were converted into z-scores, similarly to the analysis using the cochlear model (Nozaradan et al., 2012; Chemin et al., 2014; Cirelli et al., 2016) (see also SI Appendix for a control analysis using different normalization method). The z-score at the beat frequency (1.25 Hz) was taken as a measure of relative amplitude at the beat frequency. The greater this value, the more the beat frequency stood out relative to the entire set of frequency components elicited by the

rhythm (Nozaradan et al., 2016b). Additionally, the z-scores were averaged across frequencies that were related (0.416, 1.25, 2.5, 5 Hz) or unrelated (the remaining 8 frequencies) to the theoretically expected beat and meter for these rhythms (Nozaradan et al., 2012). The greater the average z-score across meter frequencies, the more prominent was the response at meter frequencies relative to all elicited frequencies. Z-score values at beat and meter frequencies were compared across conditions using a 2x2 repeated-measures ANOVAs with factors tone frequency (low, high) and rhythm (unsyncopated, syncopated).

Finally, the z-scored EEG response at the beat frequency in the high-tone condition was subtracted from the response in the low-tone condition, separately for the unsyncopated and syncopated rhythm. These difference scores were compared with the corresponding difference scores calculated from the z-scored magnitudes of the cochlear model output using one-sample t-tests. The same comparison was conducted with the averaged response at the meter-related frequencies. These comparisons between cochlear model output and EEG responses are based on the assumption that if the EEG response is driven solely by early cochlear processes, the change in relative prominence between the low- and high-tone conditions should be similar in the cochlear model output and in the EEG response.

3.7 Supplementary Information

3.7.1 Main Experiment

EEG acquisition and preprocessing. The continuous EEG recordings were high-pass filtered offline at 0.1 Hz (4th order Butterworth filter) to remove very slow drifts from the signals. Artifacts produced by eye blinks were identified and removed participant-wise with independent component analysis (Jung et al., 2000) using the Runica algorithm (Bell and Sejnowski, 1995; Makeig, 2002) applied to the concatenated epochs from all blocks segmented from 0 to 60 s relative to the trial onset. A single independent component related to eye blinks was selected and removed for each participant based on visual inspection of its waveform and topography. If the amount of variance explained by the component was less than at least 10 other components, then it was not removed from the

signal (1 participant). Channels containing excessive artifacts or noise were linearly interpolated using the 3 closest channels (1 channel interpolated in 3 participants).

Beat tapping session. The main goal of the beat-tapping task performed after the EEG session was to confirm theoretical assumptions about entrainment to the beat based on a preferential grouping by four events (i.e. beat period of 800 ms) for the present rhythmic sequences and tempo (Nozaradan et al., 2012, 2016b, 2018), and to examine possible differences in this preferential grouping between low- and high-tone conditions. Participants were asked to tap the index finger of their preferred hand in time with the regular, isochronous, pulse-like beat that they perceived in the rhythm. The experimenter provided a short pop-music example of a beat and then short examples of tapping to the unsyncopated and syncopated rhythm (according to theoretical beat frequency). It was emphasized that there were multiple plausible pulses and starting positions, and participants were encouraged to keep the beat as naturally as possible throughout the trials. The tapping was performed on a custom-built response box containing a piezoelectric sensor that registered taps, which were recorded as an audio file using PsychToolbox, version 3.0.14 (Brainard, 1997).

Tap-times were extracted by locating the peaks in the signal recorded from the response box. The first and last taps of each trial were discarded from further analyses. The mean and coefficient of variation (SD/mean) of inter-tap intervals (ITIs) were calculated for each trial and averaged for each condition and participant. Mean ITI provides an index of the perceived beat period while the coefficient of variation is a measure of beat tapping variability throughout the trial. Repeated-measures ANOVAs were performed for the mean ITI and the coefficient of variation, with factors tone frequency (low, high) and rhythm (isochronous, unsyncopated, syncopated).

The mean ITI for all three types of rhythm was predominantly 800 ms (as expected based on previous work; see Nozaradan et al., 2012, 2016b) which corresponds to grouping by four events (see Table 3.S1 and Fig. 3.S1). For the isochronous rhythm, a number of participants also tapped at a faster period of 400 ms (grouping by two events). No differences in the perceived beat period or its variability were observed between the high- and low-tone conditions for each type of rhythm, as revealed by non-significant effects involving the factor tone frequency in the ANOVAs on mean ITI and the coefficient of variation ($P_s > 0.11$).

Validity of the z-score standardization. Z-scoring the amplitudes at the 12 frequencies elicited by each rhythm (unsyncopated and syncopated) allows the relative enhancement of a particular subset of frequencies across different units, scales, and inputs to be compared. Nevertheless, z-scoring might be prone to biases to the extent that the absolute magnitudes of participants' responses are dependent on the degree of amplitude variation they exhibit across the frequency components included in the calculation. Accordingly, larger z-scores might be assigned to participants with less variability in their responses, leading to unequal weighting of participants in the group statistics independently of the experimentally manipulated factors. To ensure that this was not the case in the current study, the variability of EEG response amplitudes across the 12 peaks was subjected to a 2 x 2 ANOVA with factors tone frequency (low, high) and rhythm (unsyncopated, syncopated). There was no significant effect of tone frequency [$F(1,13) = 0.24$, $P = 0.64$, $\eta_G^2 < 0.01$] and no interaction [$F(1,13) = 1.26$, $P = 0.28$, $\eta_G^2 < 0.01$], suggesting that the observed effect of low tone in the main analysis was not due to z-scoring of the EEG amplitudes.

To address this issue further, we tested an alternative normalization method that was not dependent on the variability of EEG response amplitudes. In this method, response amplitudes were normalized by the maximum amplitude value across the 12 peaks elicited by each rhythm separately for each participant, thus rescaling the amplitudes to 1. A 2x2 ANOVA with factors tone frequency (low, high) and rhythm (unsyncopated, syncopated) revealed greater relative amplitude at the beat frequency (main effect of tone frequency) [$F(1,13) = 7.76$, $P = 0.015$, $\eta_G^2 = 0.11$] and a significant interaction between the factors rhythm and tone frequency when taking the mean across meter frequencies [$F(1,13) = 7.69$, $P = 0.016$, $\eta_G^2 = 0.06$]. Similarly, the magnitudes of the cochlear model output were rescaled to 1 and difference scores between the low- and high-tone conditions were calculated separately for the beat and meter frequencies, and the unsyncopated and syncopated rhythm. The comparison to the corresponding difference scores calculated from the EEG response amplitudes rescaled to 1 revealed significantly larger difference score in the syncopated rhythm for the beat frequency [$t(13) = 3.11$, $P = 0.03$, $d = 0.83$] and mean meter frequencies [$t(13) = 3.82$, $P = 0.009$, $d = 1.02$]. For the unsyncopated rhythm, there was no significant effect for the beat frequency nor for meter frequencies ($P_s = 1$). The similar

outcome of this alternative analysis to the main analysis indicates that the z-score procedure did not artificially bias the results of the current study.

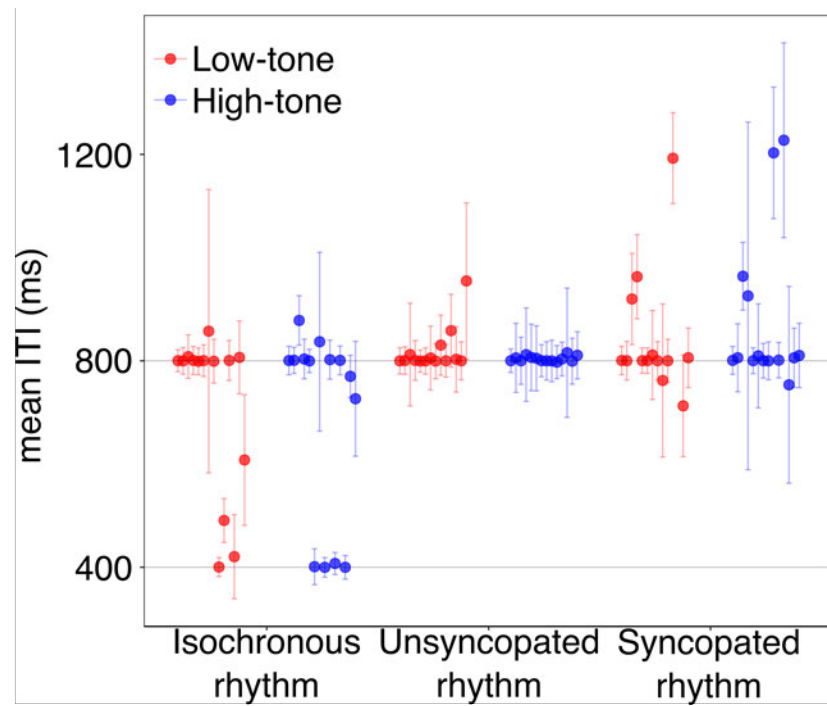


Figure 3.S1. Tapping responses. Mean inter-tap interval (ITI) for each participant in each condition, depicted as single data points. One data point was removed for display purposes from the low-tone syncopated condition (the participant tapped with period 2.4 s, i.e. repetition of the whole rhythmic pattern). Coefficient of ITI variation is shown as error bars for each condition and participant. Meter frequencies are shown as horizontal lines at 400 ms (grouping by 2 events) and 800 ms (grouping by 4 events). Participants' tapping predominantly converged toward grouping by 4 events, with some participants tapping a faster beat (grouping by 2) for the isochronous rhythm. No significant differences in the mean ITI or the coefficient of variation were observed between low-tone (red) and high-tone (blue) conditions.

Table 3.S1. Descriptive statistics for the main experiment.

	Isochronous rhythm		Unsyncopated rhythm		Syncopated rhythm	
	Low-tone	High-tone	Low-tone	High-tone	Low-tone	High-tone
Temporal deviant identification task (mean % correct)	83.92 ± 24.23	79.46 ± 27.12	78.57 ± 16.57	83.93 ± 15.83	69.64 ± 20.04	84.82 ± 15.64
Tapping tempo (median ITI ± IQR in ms)	800.05 ± 145.0	800.17 ± 314.8	800.56 ± 10.51	801.83 ± 6.41	800.48 ± 92.46	805.92 ± 96.57
Tapping variability (mean coefficient of variation)	8.9 ± 8.67	6.64 ± 5.05	6.24 ± 3.60	6.55 ± 3.32	7.89 ± 4.67	10.75 ± 9.58
EEG overall response magnitude (in µV)	0.25 ± 0.12	0.18 ± 0.09	0.49 ± 0.20	0.43 ± 0.18	0.56 ± 0.21	0.54 ± 0.25
EEG beat frequency (z-score)	-	-	1.88 ± 0.98	1.42 ± 0.99	1.07 ± 0.88	0.23 ± 0.94
EEG mean meter frequencies (z-score)	-	-	0.9 ± 0.22	0.76 ± 0.19	0.75 ± 0.23	0.38 ± 0.40

Mean ± SD; ITI, inter-tap interval; IQR, interquartile range.

3.7.2 Control Experiment 1: Effect of sound intensity

In the main experiment, low-tone and high-tone carrier frequencies were equalized in loudness to account for the differential sensitivity of the human auditory system across the frequency range. It is, nevertheless, possible that residual loudness differences, with low-tone rhythms being perceived as louder than high-tone rhythms due to possible over-correction by the psychoacoustic model, could partly account for the effects observed in the main experiment. However, to our knowledge, there is no evidence as to whether louder rhythmic sequences induce overall larger responses or a selective enhancement at specific frequencies coinciding with perceived beat and meter. Control Experiment 1 addressed these questions by directly manipulating sound intensity.

3.7.2.1 Materials and Methods

All materials and methods were the same as in the main experiment, except as indicated below.

Participants. Thirteen individuals were recruited (mean age = 26.8, SD = 8.4, 8 females), three of whom had participated in the main experiment. The amount of formal musical training ranged from 0 (5 participants) to 17 years.

Auditory stimuli. The auditory rhythms were conveyed by a pure tone at 400.1 Hz (i.e. the geometric mean between high and low tone of the main experiment). Instead of manipulating tone frequency, we manipulated sound intensity by delivering each rhythm at either 80 dB SPL (“loud condition”) or at 70 dB SPL (“soft condition”). The 10 dB difference (corresponding approximately to doubling of the perceived loudness (see e.g. Hartmann, 1997) was thus expected to be much larger than any possible residual loudness difference between low-tone and high-tone conditions in the main experiment. Importantly, the intensities were kept below the vestibular threshold (90-95 dB), as was the case in the main experiment.

Data analysis. During preprocessing, an independent component containing eyeblink-related artifacts was not removed for four participants. One channel was interpolated in three participants.

3.7.2.2 Results

The mean percentages of correct responses in the behavioral task (Table 3.S2) were comparable to the main experiment. However, the ANOVA comparing behavioral responses across conditions did not show a significant main effect or interactions involving tone intensity ($P_s > 0.62$), although some participants reported the task to be more demanding in the soft condition.

The overall magnitude of the EEG response (in μV) was not significantly different between the loud and soft conditions in either rhythm ($P_s > 0.63$). Importantly, the EEG response at the beat frequency was not significantly different in loud and soft conditions (no main effect of tone intensity, $P = 0.77$), and there was no interaction between the factors tone intensity and rhythm ($P = 0.95$). Similar results were obtained for the EEG responses at meter-related frequencies ($P_s > 0.51$). Together, these results do not support the alternative hypothesis that sound intensity might have been a confounding factor driving the effects observed in the main experiment. Conversely, these results suggest that, at sound intensities well above the detection threshold (but below the vestibular threshold), the global response magnitude and the EEG responses at beat and meter frequencies are not affected by differences in sound pressure level.

Table 3.S2. Descriptive statistics for Control Experiment 1.

	Isochronous rhythm		Unsyncopated rhythm		Syncopated rhythm	
	Loud	Soft	Loud	Soft	Loud	Soft
Temporal deviant identification task (mean % correct)	83.33 ± 15.39	84.38 ± 16.1	81.25 ± 22.3	80.21 ± 21.62	86.46 ± 17.24	81.83 ± 24.25
EEG overall response magnitude (in μV)	0.2 ± 0.08	0.19 ± 0.08	0.44 ± 0.17	0.39 ± 0.15	0.53 ± 0.18	0.48 ± 0.19
EEG beat frequency (z-score)	-	-	2.1 ± 0.59	2 ± 0.69	0.98 ± 0.98	0.92 ± 1.2
EEG mean meter frequencies (z-score)	-	-	0.83 ± 0.16	0.76 ± 0.24	0.58 ± 0.22	0.61 ± 0.24

Mean ± SD.

3.7.3 Control Experiment 2: Effect of behavioral task

Control Experiment 2 was conducted to address the relative contribution of endogenously generated beat-based predictions associated with the temporal deviant identification task on the effect of tone frequency observed in the main experiment. In the main experiment, the nature of the deviant identification task required focusing attention on fine-grained timing in the stimulus rhythm. It has been shown that detection performance of temporal perturbations is better in highly metrical rhythms (such as unsyncopated rhythms) compared to weakly metrical rhythms (such as syncopated rhythms, e.g. Bergeson and Trehub, 2006; Grube and Griffiths, 2009). This is because highly metrical rhythms induce stronger representation of metric structure, where periodic beats are utilized to precisely encode temporal properties of the stimulus (Povel and Essens, 1985). Furthermore, the fine temporal resolution of the auditory system is slightly lower for low-frequency sounds compared to high-frequency sounds (Moore et al., 1993). These two factors in combination might have resulted in greater demands for endogenous generation of the meter in order to carry out the task in the low-tone syncopated condition of the main experiment.

In the present control experiment, the behavioral task was adapted so that participants focused their attention on broadly defined properties of the auditory stimulus (pitch, tempo, and loudness), and not on the fine temporal properties as in the main experiment. That is, whereas instructions in the main experiment encouraged focus on fine-grained event timing relative to the perceived beat (to identify shorter vs. longer deviants), instructions in the current experiment encouraged general vigilance rather than attention specifically to temporal structure. These task instructions, combined with the fact that no actual changes were present in any of the trials, were assumed to guarantee similar demands for endogenous meter generation across conditions.

3.7.3.1 Materials and Methods

Materials and methods were the same as in the main experiment, except as indicated below.

Participants. Fifteen individuals were recruited (9 females, mean age = 27.5, SD = 8.7), none of whom had participated in the main experiment. The number of years of formal musical training ranged from 0 to 17 years (mean = 2.6, SD = 5.3).

Behavioral task. The rhythmic sequences did not contain any shorter or longer sound events, in contrast with the main experiment. However, to ensure that participants were generally attentive, they were asked to listen carefully to the stimuli and report after each trial any change in pitch, loudness, or tempo that was perceived. There were in fact no actual changes in any of the trials (and therefore quantitative assessment of ‘identification’ performance was not conducted). However, it can be noted that, possibly due to the repetitive nature and long duration of the stimuli, participants reported hearing very subtle (apparently illusory) changes in most trials.

Data analysis. During preprocessing, an independent component containing eyeblink-related artifacts was not removed for four participants, because the variance explained by the component was smaller than at least for 10 other components. One to three channels were interpolated in six participants.

3.7.3.2 Results

The overall magnitude of the EEG response was commensurate in the low-tone and high-tone conditions for all rhythms types, as revealed by the paired-samples t-tests ($P_s = 1$, Bonferroni-corrected). As for the relative amplitude at beat frequency, there were no significant differences between conditions ($P_s > 0.52$). Similarly, the ANOVA on meter-related frequencies revealed no significant main effects or interactions ($P_s > 0.29$). These results suggest that the effect of tone frequency is dependent on the attentional focus of the listener. Together, results of the main experiment and Control Experiment 2 suggest that the EEG response at beat and meter frequencies is boosted when the behavioral task requires focusing attention on temporal properties of the stimulus, particularly in syncopated rhythms conveyed by bass sounds. Note that these results do not imply that attention exclusively to tone duration per se is necessary for the low tone benefit. In everyday contexts, attention is also directed to temporal properties of rhythm when listening to expressively timed performances, where micro-timing variations are a key determinant of performance quality (Repp, 1992; Gabrielsson, 2003; Istók et al., 2013), coordinating body movements with music while dancing (Burger et al., 2014), or synchronizing with others during group music making (Keller et al., 2014).

Table 3.S3. Descriptive statistics for Control Experiment 2.

	Isochronous rhythm		Unsyncopated rhythm		Syncopated rhythm	
	Low-tone	High-tone	Low-tone	High-tone	Low-tone	High-tone
EEG overall response magnitude (in μV)	0.15 ± 0.07	0.15 ± 0.05	0.35 ± 0.16	0.37 ± 0.19	0.43 ± 0.17	0.42 ± 0.14
EEG beat frequency (z-score)	-	-	0.11 ± 0.96	0.05 ± 1.05	0.02 ± 0.91	-0.15 ± 0.92
EEG mean meter frequencies (z-score)	-	-	0.01 ± 0.55	0.03 ± 0.5	-0.04 ± 0.47	-0.04 ± 0.4

Mean \pm SD.

4 Study 3: Neural and behavioral evidence for frequency-selective context effects in rhythm processing in humans

This study has been published in Cerebral Cortex Communications in 2020.

To further examine the flexibility of the transformation, Study 3 investigated the role of directly preceding auditory input by presenting participants with acoustic sequences that gradually changed from high to low contrast at meter periodicities or the other way around. The results showed that the enhancement of meter periodicities in the neural response and movement can be boosted when the directly preceding input contains high contrast at these periodicities, and this context effect depends on the musical training of the listener. Thus, this study demonstrated that the transformation relevant for meter perception cannot be predicted solely from the current acoustic input, but depends on recent past, and this effect may be flexibly modulated based on long-term experience.

4.1 Abstract

When listening to music, people often perceive and move along with a periodic meter. However, the dynamics of mapping between meter perception and the acoustic cues to meter periodicities in the sensory input remain largely unknown. To capture these dynamics, we recorded the EEG while non-musician and musician participants listened to nonrepeating rhythmic sequences where acoustic cues to meter frequencies either gradually decreased (from regular to degraded) or increased (from degraded to regular). The results revealed greater neural activity selectively elicited at meter frequencies when the sequence gradually changed from regular to degraded compared to the opposite. Importantly, this effect was unlikely to arise from overall gain, or low-level auditory processing, as revealed by physiological modeling. Moreover, the context effect was more pronounced in non-musicians, who also demonstrated facilitated sensory-motor synchronization with the meter for sequences that started as regular. In contrast, musicians showed weaker effects of recent context in their neural responses and robust ability to move along with the meter irrespective of stimulus degradation. Together, our results demonstrate that brain activity elicited by rhythm does not only reflect passive tracking of stimulus features, but represents continuous integration of sensory input with recent context.

4.2 Introduction

One of the biggest challenges in understanding brain function is to explain how stable perception is experienced from continuously changing, ambiguous sensory input. To achieve such robustness, it has been proposed that the brain uses prior experience to instantiate expectations, which dynamically interact with the incoming input to shape perception (Dolan et al., 1997; Ahissar and Hochstein, 2004; Eger et al., 2007; Esterman and Yantis, 2010; Melloni et al., 2011; Holdgraf et al., 2016; de Lange et al., 2018). In particular, stimulus history, that is, recent context, plays a key role in supporting stable perception, especially in the face of degraded sensory input (Snyder et al., 2015). Effects of recent context involve a form of attraction whereby the perception of the current sensory input is biased towards recently encountered stimuli (Lieberman et al., 2016; Cicchini et al., 2018). Such effects have been reported in perception of simple features (Raviv et al., 2012; Fischer and Whitney, 2014; Arzounian et al., 2017; Chambers et al., 2017), but also higher level attributes (Cicchini et al., 2014; Lieberman et al., 2014; Suárez-Pinilla et al., 2018; Xia et al., 2018), scene perception (Snyder and Weintraub, 2013; Manassi et al., 2017), and reproduction of single temporal intervals (Jazayeri and Shadlen, 2010; Cicchini et al., 2012). Similar robustness to input degradation seems to be present in perception of rhythms (sequences of events in time). When listening to rhythms, particularly in musical contexts, humans often spontaneously organize the incoming sounds in time according to a perceived nested set of periodic pulses, usually referred to as meter (Cohn, 2020). Meter perception is considered a cornerstone of temporal prediction and sensory-motor synchronization with rhythm (Toiviainen et al., 2010; Vuust et al., 2018). Traditionally, it has been assumed that whether (and what) metric structure is perceived depends on the acoustic cues in the stimulus, namely distribution of salient acoustic events with respect to the putative pulse positions (Essens and Povel, 1985; Parncutt, 1994; Toiviainen and Snyder, 2003; Tomic and Janata, 2008; Large and Snyder, 2009). In other words, the more “pulse-like” the physical structure of the sensory input (i.e. the more salient acoustic events are preferentially concentrated at pulse positions), the more likely a meter is perceived. However, recent evidence shows that meter perception is quite robust to input deviations from a pulse-like template (Repp et al., 2008; Sioros et al., 2014; Witek et al., 2014b; Câmara and Danielsen, 2018), and mapping between the sensory input and perceptual experience is not

straightforward (London et al., 2017; van der Weij et al., 2017). This indicates that meter constitutes a high-level perceptual phenomenon that shows a degree of flexibility and stability with respect to the physical stimulus.

In line with this view, a growing body of evidence suggests that meter perception is related to fluctuations of neural activity time-locked to the perceived metric pulses (Nozaradan et al., 2018, 2011, 2012, 2016a, 2016b, 2017a, 2017b; Chemin et al., 2014; Tierney and Kraus, 2014; Tal et al., 2017; Lenc et al., 2018; Hickey et al., 2020; Kaneshiro et al., 2020). Importantly, instead of passively tracking the rhythmic structure of the acoustic input, the elicited neural activity is transformed towards selectively tracking the perceived meter, particularly when input deviates from the pulse-like template (Nozaradan et al., 2017a). This is manifested as selective enhancement of brain activity elicited at frequencies corresponding to the rates of the perceived metric pulses, relative to activity at other frequencies that are unrelated to the perceived meter but can be nonetheless prominent in the acoustic input (Nozaradan et al., 2011, 2012; Tal et al., 2017). This transformation has been observed already in the human auditory cortex (Nozaradan et al., 2016a, 2018), and possibly involves functional connections within an extended cortico-subcortico-cortical network (Nozaradan et al., 2017b). However, how sensory and endogenous signals are continuously weighted to build this neural representation of rhythm remains unknown. The current study addresses this question by directly testing the influence of recent history of auditory stimulation on the selective neural tracking of the perceived meter.

Similarly to other perceptual domains, effects of recent context are arguably at play during meter perception (London, 2004). It has been proposed by a number of music theorists that once a stable meter has been established, it tends to withstand ambiguities produced by the continuously changing rhythmic surface of music (Cooper and Meyer, 1963; Lerdahl and Jackendoff, 1983). While there is evidence suggesting that meter induced by a recent input can affect perception of subsequent time intervals (Desain and Honing, 2003; McAuley and Jones, 2003), the persistence of meter in the face of a degraded sensory input remains unclear (the general term "degradation" refers here to an input deviation from a template, i.e. how much sensory cues support a particular perceptual interpretation).

In the current study, we tested the impact of recent context on meter processing by creating auditory sequences gradually changing from a regular rhythm (onset structure

matching the pulse-like template of a given meter) to a degraded rhythm (irregular onset structure completely ambiguous with respect to the given meter). We also created flipped versions of these sequences, yielding sequences gradually changing from degraded to regular. EEG activity was recorded from participants while listening to these sequences without overt movement. After the EEG session, participants were asked to tap with the hand in time with the perceived pulse of an additional set of sequences constructed with the same algorithm as those used in the EEG session. This behavioral measure therefore indicated the induced metric periodicities across both sets of sequences. Because the envelope spectra of the stimuli were strictly identical across the original and flipped sequences, different EEG spectra across the two sequence directions would provide direct evidence for context-dependent neural representations of rhythm. This context effect would be informative about how the relative contribution of sensory and endogenous signals continuously shape neural representation of dynamic input, particularly when the sensory information is degraded. We compared groups of musicians and non-musicians, with the hypothesis that formal musical training would provide the listener with robust ability to perceive meter irrespective of sensory input degradation, thus decreasing sensitivity to recent context (Cicchini et al., 2012).

4.3 Materials and Methods

4.3.1 Participants

Thirty-two healthy volunteers participated in the study after providing written informed consent. The sample consisted of a group of individuals with no formal musical training ($N = 16$, mean age = 21.1 y, $SD = 5.1$ y, 9 females), and a group of musically trained participants ($N = 16$, mean age = 24.1 y, $SD = 5.4$ y, 13 females) with various levels of musical training (mean = 7.2 y, $SD = 4.9$ y). All participants reported normal hearing and no history of neurological or psychiatric disease. The study was approved by the Research Ethics Committee of Western Sydney University.

4.3.2 Data and code availability

Experimental stimuli and data are publicly available online (Lenc et al., 2020) at <https://doi.org/10.6084/m9.figshare.11366120>.

4.3.3 Auditory stimulation

We created rhythmic patterns by assigning a grid of twelve 200-ms events, wherein 8 events were filled with sounds (440 Hz pure tone, 10 ms linear onset and offset ramp) and 4 events with silence in all possible permutations. After removing phase-shifted versions of the same pattern, this resulted in 43 unique patterns. To quantify how well the arrangement of sound events matched a pulse-like metric template, each pattern was analyzed with a model of syncopation proposed by Longuet-Higgins and Lee (1984), as implemented in the synpy package (Song et al., 2015). The syncopation scores were calculated assuming metrical structure comprising nested pulses with rates corresponding to 2, 4, and 12 events respectively (such as in a 3/4 meter). Given these particular pulse rates (i.e. meter frequencies), there were 12 possible ways to align the metric template with each analyzed rhythmic pattern (i.e. 12 meter phases, starting on either of the 12 events constituting the rhythmic patterns). For patterns with highly regular arrangements of sound intervals, the close match of the rhythmic structure and metric template for certain alignments would necessarily result in poor match for other alignments. In contrast, for patterns with highly ambiguous structure, there would be no single alignment resulting in close match between the rhythmic structure and the metric template. Therefore, we used the range of syncopation scores across the 12 possible meter phases (the highest minus the lowest score) as a measure of the regularity of each rhythmic pattern. This value also describes the degree of phase-stability of the meter induced by each pattern. While patterns with large ranges of syncopation strongly encourage perception of particular meter phases over others, there is no such preference for patterns with small syncopation ranges (Povel and Essens, 1985; Fitch and Rosenfeld, 2007). Based on this analysis, the 43 patterns were then categorized into 8 groups (syncopation ranges {8,7,6,5,4,3,2,1}, omitting the single rhythm with range of 9), i.e., from large syncopation range (regular patterns) to small syncopation range (ambiguous patterns).

Next, we created 57.6-sec long sequences, by concatenating 24 patterns randomly chosen (with repetition) from the 43 patterns in such a way that the range of syncopation decreased continuously throughout the sequence. To do so, three different patterns were chosen in each of the eight syncopation range groups from range value 8 to 1. This yielded $3 \times 8 = 24$ patterns per sequence in total, with gradually decreasing meter phase stability. After randomly choosing a pattern within the desired syncopation-range group, its particular phase was chosen so that the syncopation score continuously increased throughout the sequence, i.e. increasing degradation with respect to the meter induced by the patterns (syncopation scores $\{-1,-1,0,1,2,3,4,4\}$ for the eight syncopation range groups). This resulted in a sequence that gradually transformed from regular to degraded without structural changes likely to trigger mental phase-shifts that would markedly reduce the perceived syncopation (e.g. Fitch and Rosenfeld, 2007).

In order to construct sequences with a gradual change in the opposite direction (from degraded to regular), we created a time-inverted version of each 57.6-s sequence, so that the first event became the last event. We also added 2 sound events at the beginning and end of the sequence, which were excluded from the analyses (see Figure 4.1). This prevented spurious differences in the neural response between sequence directions, which could otherwise arise due to increased transient responses to sound events at the beginning of each sequence.

Fifteen unique sequences and their respective inverted versions were generated, forming stimuli for two experimental conditions: the original sequences that evolved from low to high syncopation (regular-to-degraded condition) and their inverted versions that progressed from high to low syncopation (degraded-to-regular condition). Five additional sequences and their inverted versions were constructed for the tapping session. The auditory stimuli were created in Matlab R2016b (The MathWorks, Natick, MA) and presented binaurally through insert earphones (ER-2; Etymotic Research, Elk Grove Village, IL) at 75 dB SPL using PsychToolbox, version 3.0.14 (Brainard, 1997).

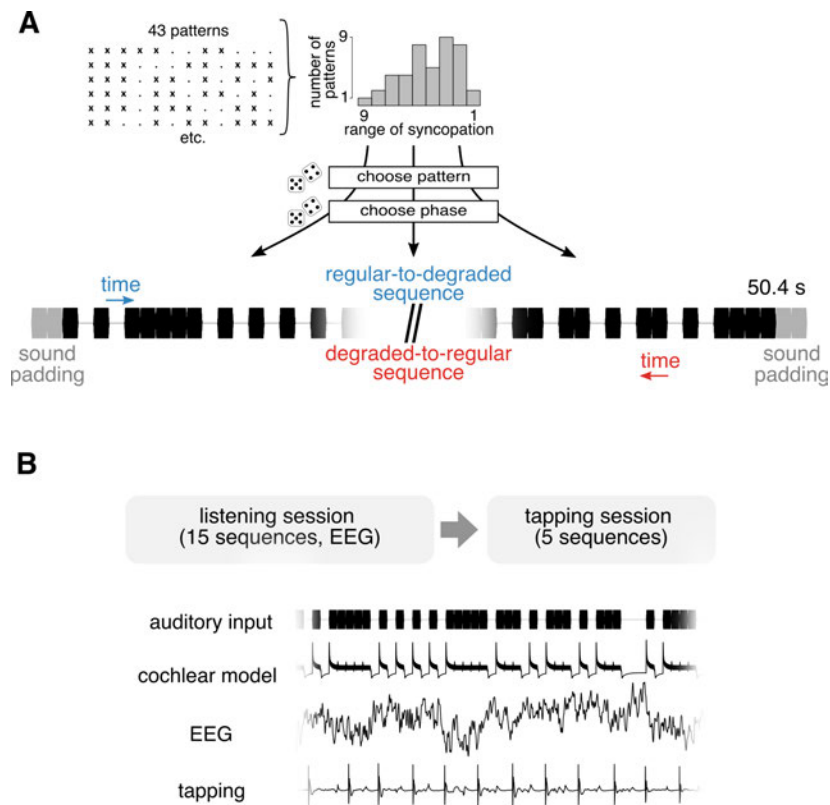


Figure 4.1. Illustration of the sequence generation method. A. Examples of individual constituent patterns used to construct the sequences. Each pattern contains 8 sounds (depicted as “x”), and 4 silences (depicted as “.”). The patterns were categorized based on the range of syncopation across all 12 possible meter phases (calculated separately for each pattern). Sequences were constructed by randomly sampling patterns according to their range of syncopation. After a pattern was selected, its particular phase (i.e. starting point) was sampled according to the particular syncopation score required. Bottom part of the panel depicts an example of a beginning and an end (padded with two sounds) of a single sequence. B. Top panel. Schematic of the experimental design. In the first session, participants listened to 15 sequences and their inverted versions without overt movement, and the EEG was recorded. This was followed by the second session, where participants tapped to five additional sequences and their inverted versions. B. Bottom panel. Examples of different signals (in the time domain) analyzed in the current study.

4.3.4 Stimulus analysis

Syncopation score. To calculate the evolution of syncopation scores across the generated sequences, the sequences were divided into 14.4-sec-long segments (72 events per segment) with 50% overlap, yielding 7 distinct segments per sequence. To evaluate whether the corresponding segments in the original and inverted sequences differed in their degree

of degradation with respect to the metric template, syncopation scores proposed by Longuet-Higgins and Lee (1984) were calculated for each segment, assuming meter with nested pulses at the rates of 2 and 4 events. This corresponded to the meter used during sequence construction without the slowest pulse, as the individual constituent patterns were not repetitively looped in the sequence. Importantly, syncopation scores are dependent on the particular alignment of the metric template with the analyzed rhythmic pattern (i.e. meter phase). However, the phase of the perceived metric structure was unknown in the current experimental design. Therefore, syncopation scores for each segment were calculated separately after moving the analysis window by -2 to 2 events relative to the first event of the segment (thus including the padding sounds for the first and last segment of each sequence). The minimum syncopation score across the phase shifts was taken, assuming that listeners have a tendency to align their perceptual metric organization in a way that yields the lowest syncopation (Povel and Essens, 1985; Fitch and Rosenfeld, 2007). Syncopation scores were compared across conditions using a linear mixed model with Direction (regular-to-degraded vs. degraded-to-regular) and Segment (1-7) as fixed effects. In this test and further statistical tests, for all models including the factor Segment as a fixed effect, the order of segments from the degraded-to-regular condition was always reversed in order to compare responses to the exact inverted versions of the same rhythmic stimulus.

The analysis of the syncopation scores calculated for the 15 stimulus sequences used in the EEG session (Figure 4.S1) yielded a significant interaction between the factor Direction and Segment ($F_{6,182} = 10.06$, $P < 0.0001$, $BF_{10} > 100$), suggesting that across trials, inversion of the sequences affected only certain segments. Post-hoc contrasts revealed that the syncopation score was significantly higher for the degraded-to-regular condition in segment 2 ($\beta = -2.33$, $t_{182} = -4.7$, $P < 0.0001$, 95% CI = [-3.31, -1.35]) and 3 ($\beta = -2.67$, $t_{182} = -5.37$, $P < 0.0001$, 95% CI = [-3.65, -1.69]), and for the regular-to-degraded condition in segment 4 ($\beta = 1.60$, $t_{182} = 3.22$, $P = 0.01$, 95% CI = [0.62, 2.58]). Even though these results suggest that the inversion procedure did not perfectly preserve the theoretically expected amount of syncopation in the sequences, the direction of the effect was opposite to the effect of context we expected to find in the EEG responses. In other words, according to the syncopation scores, there should be slightly better match between the input and metric template in the middle segment in the degraded-to-regular condition.

The procedure used to construct the auditory stimuli in the current study was based on variations in syncopation that assumed a specific metrical interpretation ($\{2,4\}$ meter with nested pulses at rates of 2 and 4 events). However, there are other possible metrical interpretations of the sequences, which were not considered during stimulus construction. To ensure that the stimulus sequences did indeed change, in theory, from an unambiguous $\{2,4\}$ meter into highly syncopated sequences instead of converging onto a different meter, we calculated the evolution of syncopation scores across the sequence for two other possible metrical interpretations ($\{3,6\}$ meter with nested pulses at rates of 3 and 6 events; $\{2,6\}$ meter with rates of 2 and 6 events). These three different metrical interpretations, ($\{2,4\}$, $\{3,6\}$ and $\{2,6\}$) constitute the simplest nested groupings of the events based on grouping by two or three events. If the sequences modulated into a different meter, then we would expect to find monotonically decreasing syncopation scores for that meter as the sequence progressed from regular to degraded. As shown in Figure 4.S2, this was not the case for the two other tested meters, further validating the stimulus construction method that was used.

Cochlear model. The main motivation for using the exact inversions of the regular-to-degraded sequences to generate the degraded-to-regular sequences was to ensure that the envelope magnitude spectra of the original and inverted sequence were identical (due to the properties of the Discrete Fourier Transform). This way, differences between the original and inverted sequences in the EEG response across corresponding segments can only be explained by recent stimulus history. To ensure that other nonlinearities in the auditory system (such as adaptation) were not likely to explain the differences between the original and inverted sequences in the EEG response, the stimuli were analysed with a cochlear model. The model consisted of a Patterson-Holdsworth ERB filter bank with 100 channels (Patterson and Holdsworth, 1996), followed by Meddis' hair-cell model (Meddis, 1986), as implemented in the Auditory Toolbox for Matlab (Slaney, 1998). The output of this model is designed to approximate sound representation in the auditory nerve, after narrowband filtering at the level of cochlea and nonlinearities introduced at the hair-cell level (adaptation, compression). The output of the cochlear model for each trial and sequence direction was segmented into seven 14.4-sec-long segments with 50% overlap (as for calculation of the syncopation scores). The obtained time-domain signals were averaged

across trials separately for each 14.4-s segment and sequence direction, and transformed into the frequency-domain using fast Fourier transform (FFT, yielding a spectral resolution of $1/14.4$ s, i.e. approximately 0.069 Hz). The resulting magnitude spectra were then averaged across cochlear channels.

As depicted in Figure 4.2, none of the obtained spectra showed clear peaks emerging from the spectral background, except at the frequency of individual events (5 Hz), and half this rate (2.5 Hz). This was due to the fact that none of the patterns making up the sequences were consistently repeated within the sequence, thus yielding no prominent periodicities in the sequences except those related to individual events and successions of two events. As the sequences gradually transformed from regular to degraded, the prominence of the peak at 2.5 Hz decreased over the segments, and the spectral energy spread across other frequencies, thus indicating, as intended, the absence of prominent cues to any particular higher-order structure beyond the event rate.

To make sure that the output of the cochlear model was not significantly different between sequence directions, especially at the frequencies related to the induced meter, we measured the amplitude at specific frequencies in the obtained spectra. These frequencies corresponded to different possible groupings of the events comprising the sequence, i.e., considering cycles of 12 events (0.416 Hz) and 16 events (0.312 Hz) and their harmonics up to 5 Hz (individual event frequency). From this set ($N = 21$ frequencies), a subset of frequencies was categorized as related to the induced meter (1.25, 2.5 and 5 Hz, as confirmed by the tapping session; see section 4.3.8). These meter-related frequencies represent nested grouping of the individual event rate (5 Hz) by 2 (2.5 Hz) and 2 (1.25 Hz), thus corresponding to the meter used to construct the sequences (as for the syncopation score calculation). All other frequencies were considered meter-unrelated. The amplitude at each frequency was extracted either at the exact frequency, if a bin was centred at that frequency (14 frequencies), or otherwise as a maximum value from the two closest bins. The 21 extracted amplitudes were z-scored as follows: $(x - \text{mean across the 21 frequencies}) / \text{SD across the 21 frequencies}$. This standardization evaluated the magnitude at each frequency relative to the other frequencies, and therefore allowed us to quantify how much a particular subset of frequencies (here meter-related frequencies) stood out relative to the whole set of frequencies. Because this measure is invariant to differences in unit and scale, it also enabled us to objectively measure the relative distance between stimulus

representation at the earliest stages of the auditory pathway (estimated with the cochlear model) and the elicited EEG response.

The relative prominence of meter-related frequencies in the cochlear model output (considering the whole set of 21 extracted frequencies) was calculated as a mean z score at 1.25, 2.5 and 5 Hz. These meter-related z scores were compared between the two sequence directions across segments to ensure that the inversion of the stimulus was unlikely to introduce significant differences in the prominence of meter frequencies at the earliest stages of the auditory pathway. For this comparison, the z-scored amplitudes were extracted in the way described above but separately for each trial (i.e. without first averaging across trials in the time domain), and fitted with a mixed model (fixed effects Direction and Segment). There were no significant differences between the original and inverted condition (main effect of Direction, $F_{1,182} = 0.01$, $P = 0.92$, $BF_{10} = 0.15$; interaction of Direction and Segment, $F_{6,182} = 0.64$, $P = 0.7$, $BF_{10} = 0.07$). This result suggests that nonlinearities at the early stages of the auditory pathway are unlikely to account for any effects of context in the EEG responses.

The same analyses performed on the 5 sequences used in the tapping session suggested similar differences in syncopation scores, including higher syncopation score for degraded-to-regular condition in segment 2 ($\beta = -2.8$, $t_{52} = -2.96$, $P = 0.03$, 95% CI = [-4.7, -0.9]) and 3 ($\beta = -3$, $t_{52} = -3.17$, $P = 0.02$, 95% CI = [-4.9, -1.1]), and no significant effects involving the factor Direction for the analysis with cochlear model ($P_s > 0.82$, $BF_{510} < 0.25$).

4.3.5 Experimental design and procedure

The experiment consisted of an EEG and a tapping session directly following each other. In the EEG session, participants were presented with the 15 sequences and their inverted versions in random order with regular-to-degraded and degraded-to-regular trials alternating (counterbalanced across participants). Participants were seated in a comfortable chair with their head resting on a support, and asked to avoid any unnecessary movement. The support made contact with the head just below the most inferiorly positioned electrodes in order to prevent artifacts in the recorded EEG signals. Participants were asked to focus on the regular pulse in the auditory stimuli, and after each trial, to rate (on a scale from 1 to 5) how difficult on average they thought it would be to tap along the pulse in that

trial. To further encourage attention to the temporal properties of the stimuli, participants were also asked to detect slight transient decrease of tempo randomly inserted in two additional trials that were not included in the analyses. Before the EEG session, the experimenter provided examples of pulse in popular music and artificially constructed rhythms, to make sure participants understood the task.

After the EEG session, participants were presented with five additional sequences and the respective inverted versions (as for the EEG session, with random order, sequence direction alternating, counterbalanced across participants), and were asked to tap the regular pulse they perceived in the sequences using the index finger of the preferred hand. Participants were instructed to tap any pulse they perceived in the rhythmic sequence, as long as the pulse they tapped was (i) isochronous and (ii) synchronized to the stimulus sequence. They were allowed to start and stop tapping within a trial depending on whether they perceived a periodic pulse or not, and change the period or phase of the pulse at any point. Tapping was performed on a custom-built response box containing a piezoelectric sensor that converted the mechanical vibrations of the box due to the impact of the finger into electrical signals, which were recorded as audio files.

4.3.6 EEG recording and preprocessing

The EEG was recorded using a Biosemi Active-Two system (Biosemi, Amsterdam, Netherlands) with 64 Ag-AgCl electrodes placed on the scalp according to the international 10/20 system, and two additional electrodes attached to the mastoids. Head movements were monitored using an accelerometer with two axes (front-back and left-right) attached to the EEG cap and recorded as 2 additional channels. The signals were digitized at a 2048-Hz sampling rate and downsampled to 512 Hz offline.

The continuous EEG signals were high-pass filtered at 0.1 Hz (4th order Butterworth filter) to remove slow drifts from the signal. Independent component analysis (Bell and Sejnowski, 1995; Jung et al., 2000) was used to identify and remove artifacts related to eye blinks and horizontal eye movements based on visual inspection of their typical waveform shape and topographic distribution (2 components removed for 14 participants, 1 component for 18 participants). Channels containing excessive artifacts or noise were linearly interpolated using the 3 closest channels (1 channel interpolated for 2 participants, 4 channels for 1

participant). The cleaned EEG data were segmented into 57.6-s long epochs, starting from 0.4 s relative to trial onset (i.e. discarding the two padding sound events, see above section 4.3.3 and Figure 4.1). If an epoch contained excessive artifacts it was discarded from further analyses (1 epoch for 1 participant), as well as the epoch for the trial with inverted version of the corresponding stimulus sequence. The epochs were then further segmented into seven 14.4-sec long segments with 50% overlap (as for the auditory stimulus analysis), re-referenced to the common average, and averaged across trials in the time domain separately for each sequence direction, segment, and participant. Time-domain averaging was performed to increase the signal-to-noise ratio of the neural response by cancelling signals that were not time-locked to the stimulus (Mouraux et al., 2011; Nozaradan et al., 2011, 2012). The EEG preprocessing was carried out using Letswave6 (www.letswave.org) and Matlab.

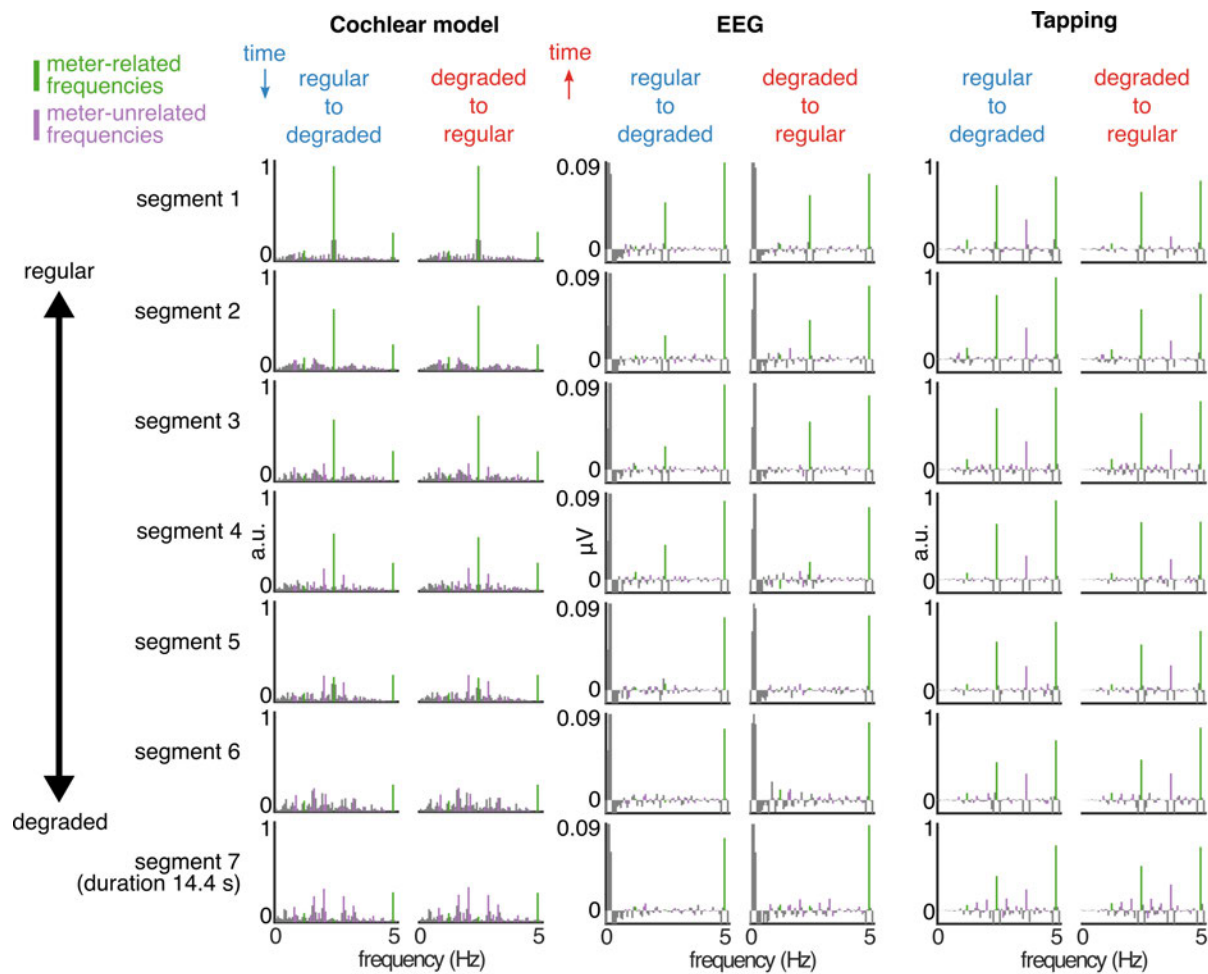


Figure 4.2. Cochlear model, EEG, and tapping spectra. The data are averaged across trials and plotted separately for each segment and sequence direction. The segments from the degraded-to-regular condition are displayed in reverse order to facilitate comparison across conditions (this way the segments with the same stimulus envelope spectra are aligned). The cochlear model output (Left) shows highly similar spectra across sequence directions, with decreasing prominence of meter-related frequencies (green) and increasing prominence of meter-unrelated frequencies (purple) as the sequence changes from regular to degraded. The EEG response (Middle) averaged across all channels and participants contains peaks at the frequencies present in the cochlear model output, with decreasing prominence of meter-related frequencies in the degraded segments. The tapping response (Right) averaged across participants shows prominent peaks at meter-related frequencies even in the degraded segments.

4.3.7 Frequency-domain analysis of EEG response

For each participant, sequence direction, and segment, the EEG signals were transformed into the frequency domain using FFT. The obtained EEG spectra can be assumed to consist

of a superposition of (i) responses to the stimulus concentrated into narrow peaks and (ii) residual background noise smoothly spread across the entire frequency range. To obtain valid estimates of the responses, the contribution of noise was minimized by subtracting, at each frequency bin, the average amplitude in the 2nd neighboring bin either side of it (Mouraux et al., 2011; Xu et al., 2017).

Because the meter-unrelated frequencies did not form prominent narrow peaks in the output of the cochlear model, it was important to ensure that the noise subtraction would not selectively suppress meter-unrelated frequencies in the EEG spectra (which could lead to spurious increase in the relative prominence of meter frequencies if there was an overall increase in response gain). A control analysis conducted on the EEG spectra obtained without noise subtraction yielded similar results to the analysis incorporating noise subtraction (see Supplementary Results), showing that this processing step alone could not explain our results. The noise-subtracted spectra were averaged across all channels to avoid electrode-selection bias and to account for individual differences in response topography.

To assess the relative prominence of the specific frequencies in the EEG responses elicited by the auditory stimuli, amplitudes at the 21 frequencies corresponding to different possible metric interpretations were then extracted from the spectra and z-scored in the same way as for the auditory stimulus analysis. A higher z score at a specific frequency indicates more prominent amplitude at that frequency relative to the whole set of 21 frequencies in the EEG response. Mean z-scored amplitude at frequencies related to the induced meter (5 Hz, 2.5 Hz and 1.25 Hz, as theoretically expected based on the sequence generation algorithm and as indicated by tapping analysis) was taken as a relative measure of selective neural tracking of the meter periodicities (control analysis with raw EEG amplitudes yielded similar results to the analysis with z scores, see Supplementary Results). The mean meter-related z-scored amplitudes were compared across sequence directions and segments, by fitting a mixed model (fixed effects Direction, Segment, and Musical Training). We expected to find a decrease in the prominence of meter-related frequencies in the segments with higher degradation, as in the auditory stimulus. Importantly, we used additional post-hoc contrasts to test whether the EEG response was affected by the direction of the sequence, by comparing the prominence of meter frequencies in segment one (most regular rhythm) to all subsequent segments, separately for each sequence direction. We hypothesized that in the regular-to-degraded condition, the decrease would take place in segments with higher

amounts of degradation compared to the degraded-to-regular condition. We also directly compared segments that had identical sound envelope spectra across sequence directions, to assess whether the EEG response at meter-related frequencies would be enhanced for particular segments in the regular-to-degraded condition.

To further show that cochlear processing was unlikely to explain the effect of context in the EEG responses, the two signals were directly compared after standardization (z-scoring). In order to use the same processing pipeline for the EEG and cochlear model (see section 4.3.4), the cochlear model spectra were noise-subtracted (2nd bin on each side) before z-scoring the magnitudes across the meter-related and meter-unrelated frequencies. Subsequently, the difference in meter-related z scores between the cochlear model and the EEG response was calculated separately for each sequence direction, segment, and participant. The difference scores were compared between sequence directions, segments, and levels of musical training with a mixed model, and post-hoc contrasts compared the difference score between directions separately for each segment. Hence, if the EEG responses were fully explained by cochlear processing, the obtained scores should not significantly differ between the two sequence directions.

4.3.8 Tapping analysis

Tap times were extracted by locating points in the continuous signal from the tapping sensor where the (i) amplitude was increasing, (ii) amplitude exceeded a threshold set manually for each participant, and (iii) the amount of time from the previous detected point was larger than a constant set manually for each participant. These points corresponded to the tap onsets, i.e. the times where the finger hit the response box.

To quantitatively evaluate the meter periodicities participants synchronized to, the median inter-tap interval (ITI) was calculated separately for each sequence direction and participant. The value was then compared to three possible meters each consisting of 3 nested periodicities (nested pulses at rates of {2,4}, {2,6}, and {3,6} events, corresponding to periods {200, 400, 800} ms; {200, 400, 1200} ms; and {200, 600, 1200} ms, respectively) by taking the minimum percent difference between the median ITI and the three possible periodicities comprising each meter. This minimum difference score was compared across meters and sequence directions using a mixed model. The meter that yielded the smallest

difference score was considered to be the meter predominantly induced by the stimulus construction method.

To assess how well participants synchronized to the meter periodicities, it was important to consider the challenges stemming from the nature of the tapping task, whereby participants were free to tap any periodic pulse they perceived and could start and stop tapping at different points within a trial. Therefore there was no a priori information about the particular period and phase they were tapping, and the number of executed taps could differ between trials. Additionally, the tapped period and phase could change between and within individual analysis windows, without necessarily implying poor synchronization to the meter.

To provide a measure of synchronization insensitive to infrequent changes in tapping phase within the analysis windows, an ITI-error index was calculated separately for each participant, sequence direction, segment, and trial. This was done by first removing ITIs longer than 2 seconds and finding the minimum percent difference between the median ITI and the three periodicities within the predominantly induced meter (i.e. 200, 400, 800 ms, see Results section). The period closest to the median ITI was considered the pulse chosen by the participant for the analyzed window, and ITI-error was calculated as percent difference between this period and each individual ITI. The ITI-errors were averaged across trials and analyzed using a mixed model with Direction, Segment, and Musical Training as fixed effects. If the participant tapped with a fixed period corresponding to one of the metric pulses, but changed the alignment of this pulse with respect to the rhythmic stimulus at some point in the analysis window, ITI-error would remain low. Hence, the main advantage of this measure was its robustness to changes in tapping phase. However, if the participant changed the tapping period within the analysis window to another metric pulse, the ITI-error would become high.

Thus, in order to account for this, the tapping was also analyzed in the frequency domain. This evaluated synchronization at meter-related frequencies at the level of behavioral output with a method directly comparable to the auditory stimuli and EEG responses. The main advantage this frequency-domain analysis was its robustness to changes in tapping period within the analysis window, as tapping either metrical pulse would result in energy distributed solely across meter-related frequencies. However, the method was sensitive to phase changes, as changes in tapping phase within the analysis window would lead to

decreased Fourier magnitude at the tapping frequency. This is in contrast with ITI-error, which was robust to phase changes but sensitive to changes in tapping period. Moreover, continuous signals from the tapping box contained information about tapping intensity (amount of accentuation of each tap), thus potentially revealing periodicities in the behavioral response that would remain hidden when analyzing ITIs. Continuous signals from the response box recorded during the tapping session were segmented the same way as the EEG signals, averaged across trials in the time domain, and transformed into the frequency-domain using FFT. The contribution of background noise was minimized, as for the EEG, by subtracting the average magnitude in the 2nd neighboring bin either side of each frequency-bin. The resulting magnitude spectra were averaged across trials, and magnitudes at meter-related and meter-unrelated frequencies were extracted and z-scored as for the EEG analysis. Mean z-scored amplitudes at meter-related frequencies were compared across segments, sequence directions, and levels of musical training, by fitting a mixed model. The persistence of the tapping synchronization across different amounts of syncopation was assessed using post-hoc contrasts that compared the prominence of meter-related frequencies in the first segment to all subsequent segments. To further understand the evolution of the tapping response over segments, the prominence of meter frequencies was also compared across all pairs of successive segments.

4.3.9 Head movement analysis

To evaluate the extent to which unintentional head movement artifacts could explain the observed EEG results, the data from the accelerometer were segmented the same way as EEG signals and transformed into the frequency-domain separately for each movement axis. The resulting spectra were averaged across the two axes, and mean magnitudes at meter-related frequencies were extracted and further analyzed as for the EEG responses. This control analysis confirmed that the observed EEG effects were unlikely to be explained by head movement artifacts (see Supplementary Results).

4.3.10 Statistical analyses

The statistical analyses were performed using linear mixed models with lme4 package in R (Bates et al., 2014). Each participant was included as a random-effect intercept (in case of

stimulus analyses, the intercept was modeled as a random variable across trials). For models including the factor Segment as a fixed effect, the order of segments from the degraded-to-regular condition was always reversed in order to compare responses to the inverted version of the same acoustic stimulus. Post-hoc multiple comparisons were computed using emmeans package (Lenth, 2018). The Kenward-Roger approach was used to approximate degrees of freedom and Bonferroni correction was used to adjust for multiple comparisons. Complementary to the null-hypothesis significance tests with mixed models, we also calculated Bayes factors to quantify the evidence in favor of the alternative hypothesis over the null hypothesis (BF_{10}), as implemented in the package BayesFactor for R (Morey and Rouder, 2014).

4.4 Results

4.4.1 Tapping

Median ITI analysis. The tapping task confirmed theoretical expectations about the meter periodicities induced by the auditory stimulus sequences. The difference between the median ITI and possible meter periodicities varied significantly across the different possible meters ($F_{2,155} = 19.65$, $P < 0.0001$, $BF_{10} > 100$). Post-hoc comparisons showed that the median ITI was significantly closer to the {2,2} meter than the {3,6} meter ($\beta = -13.22$, $t_{157} = -5.39$, $P < 0.0001$, 95% CI = [-19.15, -7.28]) and {2,6} meter ($\beta = -13.54$, $t_{157} = -5.52$, $P < 0.0001$, 95% CI = [-19.47, -7.61]). These results further justify the selection of meter-related frequencies (5Hz, 5 Hz/2 and 5 Hz/4, corresponding to the rates of one, two, and four individual events respectively) for the frequency-domain analyses.

Frequency-domain analysis. The spectra of continuous signals from the tapping sensor exhibited prominent peaks at meter-related frequencies (Figure 4.2). As depicted in Figure 4.3, the prominence of these frequencies in the tapping spectra evolved across segments differently for musicians and non-musicians ($F_{6,390} = 5.53$, $P < 0.0001$, $BF_{10} > 100$). When comparing the two groups separately for each segment, meter frequencies were more prominent for musicians in segments 5 ($\beta = 0.62$, $t_{55.83} = 3.39$, $P = 0.009$, 95% CI = [0.25, 0.99]), 6 ($\beta = 0.77$, $t_{55.83} = 4.2$, $P = 0.001$, 95% CI = [0.4, 1.14]) and 7 ($\beta = 0.91$, $t_{55.83} = 4.94$, $P < 0.0001$, 95% CI = [0.54, 1.28]). This was due to the fact that for non-musicians, meter

frequencies significantly decreased in segments 5 ($\beta = -0.44$, $t_{396} = -4.29$, $P = 0.001$, 95% CI = [-0.65, -0.24]), 6 ($\beta = -0.79$, $t_{396} = -7.61$, $P < 0.0001$, 95% CI = [-0.99, -0.58]) and 7 ($\beta = -0.9$, $t_{396} = -8.66$, $P < 0.0001$, 95% CI = [-1.1, -0.69]) when compared to segment 1, while for musicians there was only a trend towards a decrease in segment 6 ($\beta = -0.31$, $t_{396} = -2.97$, $P = 0.04$, 95% CI = [-0.51, -0.10]). This indicates that the ability of non-musicians to synchronize their tapping at meter frequencies deteriorated significantly once the degradation in the sensory input exceeded a critical level.

There was also a significant interaction between musical training and condition ($F_{1,390} = 6.25$, $P = 0.01$, $BF_{10} = 2.4$). While the overall prominence of meter frequencies was larger in the tapping of musicians for both sequence directions, this difference was more pronounced in the degraded-to-regular condition ($\beta = 0.63$, $t_{33.82} = 3.87$, $P = 0.001$, 95% CI = [0.3, 0.96]) than the regular-to-degraded condition ($\beta = 0.43$, $t_{34.11} = 2.66$, $P = 0.02$, 95% CI = [0.1, 0.76]). This was due to the fact that non-musicians showed overall smaller prominence of meter frequencies in the degraded-to-regular condition compared to the regular-to-degraded condition ($\beta = 0.16$, $t_{396} = 2.95$, $P = 0.01$, 95% CI = [0.05, 0.27]).

ITI-error analysis. ITI-error index values further confirmed the results from the frequency domain analysis (interaction between Direction and Musical Training, $F_{1,390} = 10.97$, $P = 0.001$, $BF_{10} = 28.31$), by revealing significantly less tapping error in the regular-to-degraded condition compared to the degraded-to-regular condition for non-musicians ($\beta = -0.04$, $t_{396} = -4.66$, $P < 0.0001$, 95% CI = [-0.05, -0.02]) (Figure 4.S3). Interestingly, there was no effect of Segment in the analysis of ITI-error ($P_s > 0.25$, $BF_{s10} < 0.09$). This suggests that the fast deterioration of non-musicians' tapping in the degraded segments, as observed in the frequency-domain analysis of tapping, was partly related to frequent changes in tapping phase. Taken together, these results suggest that non-musicians' tapping to the meter generally improved when the rhythm evolved from regular to degraded compared to the opposite direction, whereas musicians showed precise and stable tapping synchronization across all levels of degradation.

4.4.2 Frequency-domain analysis of EEG

EEG responses were elicited at frequencies that were expected on the basis of the auditory stimulus analysis (Figure 4.2), with typical fronto-central topographies (Figure 4.4), as

previously observed for responses to repeating auditory rhythms (Nozaradan et al., 2012; Lenc et al., 2018).

The main aim of the current study was to examine the effect of context on the relative amplitude of EEG responses at meter-related frequencies (Figure 4.3). The direction of the sequence affected the prominence of meter-related frequencies (mean z-scored amplitudes) in the EEG response (interaction between Direction and Segment, $F_{6,390} = 4.26$, $P = 0.0004$, $BF_{10} = 33.70$). Directly contrasting the corresponding segments between the two sequence directions revealed significantly larger meter frequencies for segment 4 ($\beta = 0.37$, $t_{396} = 4.16$, $P = 0.0002$, 95% CI = [0.20, 0.55]) in the regular-to-degraded condition compared to the opposite sequence direction. This was due to greater persistence of the response in the regular-to-degraded condition, as degradation increased. Table 4.1 shows the response across segments compared to the first segment, separately for musicians and non-musicians. For non-musicians, the response significantly decreased in segment 5, 6, and 7 in the regular-to-degraded condition. However, for the degraded-to-regular condition, there was a significant decrease already in segment 4, followed by segment 5, 6, and 7. In other words, in the segment with medium amount of degradation, the meter-related frequencies were more prominent in the EEG when regular, as opposed to degraded, input preceded this segment. Similar, although less pronounced, pattern of results was observed for musicians (decrease in segments 5 and 6 for regular-to-degraded and segments 4, 5, and 6 in the opposite direction). However, despite this apparent difference between musicians and non-musicians, the three-way interaction between sequence direction, segment, and musical training was not significant ($F_{6,390} = 0.71$, $P = 0.64$, $BF_{10} = 0.07$), suggesting that context affected the neural response similarly across groups.

Furthermore, there was an interaction between musical training and segment ($F_{6,390} = 4.35$, $P = 0.0003$, $BF_{10} = 41.70$). However, this effect seemed primarily driven by greater selective response at meter-related frequencies in segment 7 for musicians, which did not reach significance in the post hoc contrasts ($\beta = 0.30$, $t_{109.06} = 2.57$, $P = 0.08$, 95% CI = [0.07, 0.54]). Finally, musical training interacted with sequence direction ($F_{1,390} = 9.03$, $P = 0.003$, $BF_{10} = 6.51$). However, post hoc contrasts did not reveal significant differences between musicians and non-musicians in either condition ($P_s > 0.13$).

A number of control analyses were done to confirm that the sequence direction effects observed here were not spurious (see Supplementary Results). Specifically, these control

analyses showed that the context effect (i) could not be explained by head movement artifact or (ii) low-level nonlinear auditory processing of the inputs, and (iii) was not a spurious effect of the standardization, or (iv) noise subtraction procedure applied to the EEG data.

Table 4.1. Prominence of meter-related frequencies in the EEG response compared between the first and all subsequent segments, separately for the two sequence directions, and for musicians (N = 16) and non-musicians (N = 16).

musical training	direction	contrast segments	estimate	df	t	lower CI	upper CI	p-value	
non-musicians	egu a o deg aded	2	0.5	390	.9	0.49	0.9	.00	
		3	0.2	390	.63	0.54	0.3	.00	
		4	0.7	390	.36	0.5	0.6	.00	
		5	0.43	390	3.40	0.77	0.09	0.02	*
		6	0.79	390	6.2	.2	0.45	<0.000	***
		7	0.84	390	6.62	.8	0.50	<0.000	***
	deg aded o egu a	2	0.2	390	0.9	0.45	0.22	.00	
		3	0.09	390	0.69	0.42	0.25	.00	
		4	0.55	390	4.33	0.89	0.2	0.0004	***
		5	0.76	390	5.95	.09	0.42	<0.000	***
		6	0.63	390	4.98	0.97	0.30	<0.000	***
		7	0.74	390	5.8	.07	0.40	<0.000	***
musicians	egu a o deg aded	2	0.2	390	0.94	0.46	0.22	.00	
		3	0.06	390	0.50	0.40	0.27	.00	
		4	0.0	390	0.06	0.34	0.33	.00	
		5	0.57	390	4.45	0.90	0.23	0.0003	***
		6	0.55	390	4.35	0.89	0.22	0.0004	***
		7	0.34	390	2.68	0.68	0.00	0.8	
	deg aded o egu a	2	0.8	390	.4	0.52	0.6	.00	
		3	0.03	390	0.2	0.36	0.3	.00	
		4	0.38	390	3.0	0.72	0.05	0.07	.
		5	0.53	390	4.2	0.87	0.20	0.0008	***
		6	0.39	390	3.09	0.73	0.06	0.05	.
		7	0.9	390	.52	0.53	0.4	.00	

CIs represent 95% confidence intervals. (. P < 0.1, * P < 0.05, ** P < 0.01, *** P < 0.001)

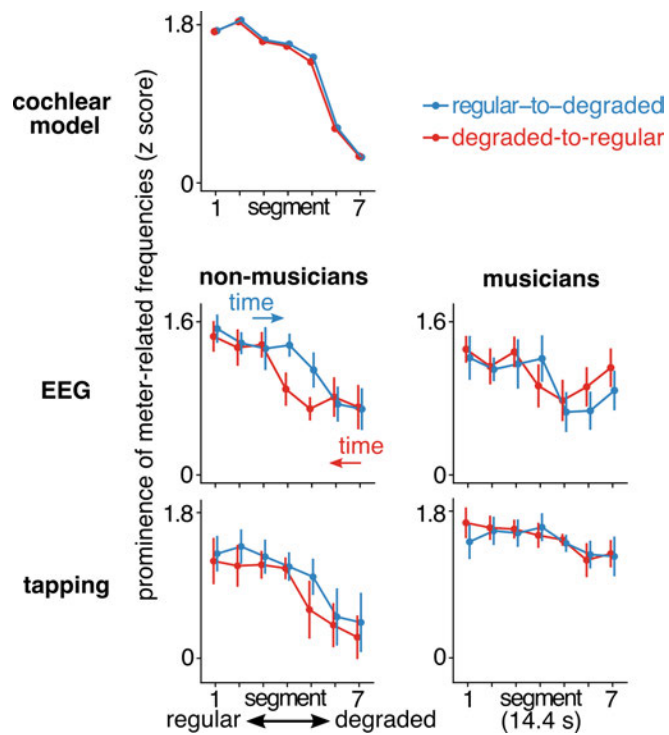


Figure 4.3. Mean z-scored amplitudes at meter-related frequencies in the cochlear model, EEG, and tapping response. The order of segments in the degraded-to-regular condition (red) is reversed to aid the comparison of segments with identical stimulus envelope spectra across conditions. Arrows indicate the direction of time for each condition. Mean values are shown as points, and error bars represent 95% confidence interval (Morey, 2008). (Top) Cochlear model output. As intended, the prominence of meter frequencies decreased as the degradation of the sequence increased. (Middle) EEG responses plotted separately for non-musicians (Left) and musicians (Right). Non-musicians showed enhanced EEG responses at meter frequencies in the middle segments of the regular-to-degraded condition (blue). The EEG responses of musicians were more similar across conditions. (Bottom) Tapping responses. For non-musicians (Left) the prominence of meter frequencies in the tapping decreased rapidly with increasing degradation. Musicians (Right) showed prominent meter frequencies in their tapping even in the degraded segments.

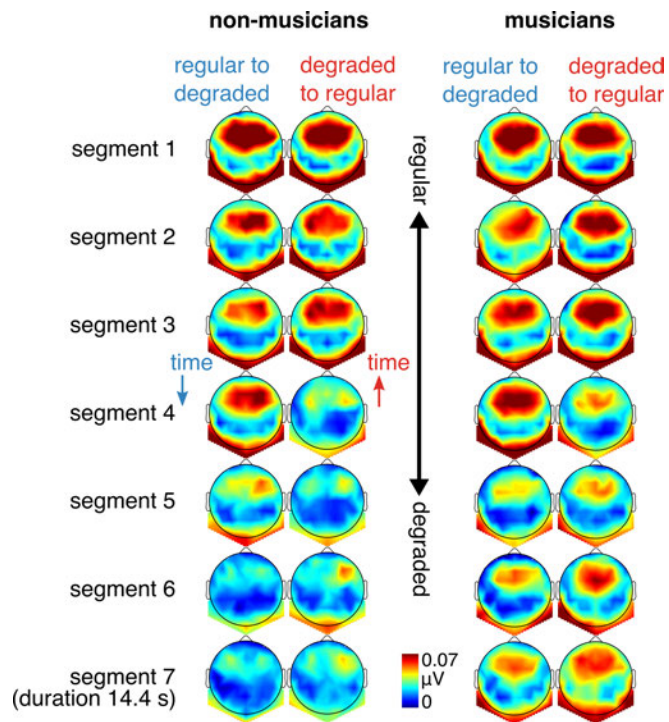


Figure 4.4. Topographies of the mean EEG amplitude at meter-related frequencies. Scalp distributions of responses across conditions and segments are shown separately for non-musicians (Left) and musicians (Right).

4.5 Discussion

Our results show direct evidence for sensitivity to recent auditory context in neural responses to rhythmic inputs. In the EEG, we observed a selective enhancement of meter-related frequencies that persisted when the acoustic cues guiding meter perception were gradually degraded in the stimulus. Conversely, these meter-related frequencies were less prominent in the neural response when the preceding input lacked acoustic cues to guide meter perception. Moreover, this context effect seemed stronger in participants with no formal musical training, who (i) demonstrated sensitivity to context in their ability to tap along with the meter, and (ii) whose tapping deteriorated when it was not supported by acoustic cues in the stimulus. In contrast, the context effect appeared weaker in musicians' EEG, and their tapping indicated a robust ability to maintain a meter despite stimulus degradation and independently of context. Together, these results demonstrate that perceptual organization of a rhythmic stimulus is not solely determined by low-level

features of the sensory input but also involves integration of prior experience, as reflected in the elicited neural activity.

Importantly, our stimulus design ensured that low-level input properties such as envelope spectra could not fully account for the observed neural responses. Moreover, the context effect observed here was unlikely to be explained by nonlinearities at the early stages of the auditory pathway (as indicated by the analysis of our stimuli with a biologically plausible model of the auditory periphery), or the overall gain (as we used a relative measure of response prominence). Instead, the context effect could be explained by selective neural enhancement of meter-related frequencies as a function of prior prominence of these frequencies in the sensory input.

4.5.1 No one-to-one mapping between sensory input and perception

Robust perception. Human perception shows remarkable robustness to degraded sensory input across domains (Shannon et al., 1995; Schwiedrzik et al., 2018). For instance, while under certain conditions the perception of a visual object or a speech utterance can be largely determined by the physical features of the sensory input, in real-world noisy situations the mapping between the input and perceptual experience is far from trivial. Our results show that similar processes may be at work in perceptual organization of rhythm, especially for individuals with musical training. We found that musicians were able to precisely synchronize their tapping to the perceived meter even when this meter could not be clearly determined from the stimulus features alone. This is in line with previous evidence that musical training generally leads to superior precision of meter representation (Rüsseler et al., 2002; Brochard et al., 2003; Geiser et al., 2010; Lappe et al., 2011), with a high degree of invariance with respect to the rhythmic stimulus (Repp, 2007, 2010; Repp et al., 2008; Su and Pöppel, 2012).

Sensitivity to context. Further evidence against a one-to-one mapping between acoustic input and perceptual output is provided by the effect of recent context we observed in the tapping and in the EEG response. These results suggest that perception of meter in degraded rhythmic input can be facilitated when the directly preceding input provides clear sensory cues to the meter periodicities. While effects of recent context have been

investigated in single-interval timing (Drake and Botte, 1993; Large, 2000b; McAuley and Jones, 2003; Jazayeri and Shadlen, 2010; Cicchini et al., 2012) and rhythmic pattern perception (Desain and Honing, 2003), they remain under-explored with respect to perceptual organization of rhythmic patterns (Cameron and Grahn, 2016). The current results thus constitute a step forward in our understanding of how the brain dynamically builds representation of complex patterns of time intervals.

The fact that these context effects were stronger in participants with no musical training is consistent with the hypothesis that influence of prior context increases as the uncertainty of the current representation increases (Cicchini and Burr, 2018; Cicchini et al., 2018). Non-musicians, whose meter perception was overall less robust to input degradation, would rely more on the recent context to make better sense of the degraded input (see Cicchini et al., 2012 for similar findings in single time interval reproduction). The context effect observed here is also similar to widely studied phenomena in visual object recognition and language domains, where perception of objects from impoverished inputs can be enhanced by prior exposure to the intact version of the stimulus (Bruner and Potter, 1964; Dolan et al., 1997; Kleinschmidt et al., 2002; Ahissar and Hochstein, 2004; Melloni et al., 2011; Teufel et al., 2015), or even through higher-level semantic context (Eger et al., 2007; Hervais-Adelman et al., 2008; Esterman and Yantis, 2010; Sohoglu et al., 2014; Stein and Peelen, 2015). Both types of perceptual enhancements have been linked to neural responses across a widespread network, involving sensory and frontal cortices (Kleinschmidt et al., 2002; Hegdé and Kersten, 2010; Sohoglu et al., 2012; Sohoglu and Davis, 2016). Moreover, there is evidence suggesting that the underlying mechanism might involve top-down modulations biasing processing of input features in sensory areas towards greater similarity with the expected category (Hsieh et al., 2010; Holdgraf et al., 2016; Leonard et al., 2016; St. John-Saaltink et al., 2016). While our method does not address the neural network mediating the context effects observed here, our results provide a new critical piece of knowledge on the integration of sensory input with context. That is, brain activity elicited at behaviorally-relevant frequencies is significantly modulated by the prominence of these frequencies in recent input. These findings may thus provide a basis to further investigations of the nature of neural representations of rhythmic input, using a similar design combined with a range of neuroimaging methods including intracerebral EEG (Grahn and Rowe, 2013; Chemin et al.,

2014; Rajendran et al., 2017; Mendoza et al., 2018; Narain et al., 2018; Gámez et al., 2019; Sohn et al., 2019).

4.5.2 Evidence against evoked responses passively tracking low-level acoustic features of the rhythmic input

Increasing evidence converges toward the view that during meter perception, the brain transforms the sensory input (a sequence of events in time) towards the metrical category (a nested set of periodic pulses), and this transformation can be observed as a selective increase of brain response at meter-related frequencies (Nozaradan et al., 2012, 2016a, 2017a, 2018). Importantly, this transformation is not fixed or mechanistic, but can be flexibly shaped by the spectral acoustic context (Lenc et al., 2018), prior body movement (Chemin et al., 2014), or behavioral goals (Nozaradan et al., 2011). Here we add to this evidence by showing that this transformation can be dynamically shaped by preceding input and even without overt movement. Together, these results thus provide strong evidence against the view that this selective increase of brain response at meter-related frequencies reflects passive tracking of low-level features of the rhythm (Large et al., 2015; Daube et al., 2019; Rimmele et al., 2020). Instead, the data suggest that this measure is (i) behaviorally relevant, and (ii) reflects transformation from acoustic features towards higher-level categories, in line with recent work on speech (Ding and Simon, 2012; Mesgarani and Chang, 2012; Di Liberto et al., 2015, 2019; Brodbeck et al., 2018; Broderick et al., 2018) and melody perception (Di Liberto et al., 2020a; Sankaran et al., 2020).

Moreover, the approach used in the current study goes beyond the common assumption that better alignment of neural response with stimulus envelope necessarily reflects better processing (Park et al., 2015; Etard and Reichenbach, 2019; Harding et al., 2019; Fiveash et al., 2020; Herff et al., 2020; Wollman et al., 2020). Specifically, instead of looking for precise reconstruction of low-level features such as envelope periodicity using e.g. input-output coherence or regression analysis, the current study aimed to investigate dynamic processes that continuously transform sensory input towards invariant perceptual categories (Ley et al., 2014; Kuchibhotla and Bathellier, 2018; Broderick et al., 2019; Yi et al., 2019). The input-output mapping approach used here allowed us to uncover these processes while ensuring

that the results are not driven by (i) acoustic confounds, (ii) overall gain of the response, or (iii) low-level nonlinear auditory processes.

4.5.3 Context effect is short-lived in neural activity but long-lasting in behavior

In the current study, the contextual enhancement of meter-related frequencies in the EEG was relatively short-lived, i.e. lasting around one 14-seconds long segment. These observations demonstrate that the influence of prior acoustic context on EEG responses might have a short time constant, only affecting the processing of directly following rhythmic material. Such short-lived integrative mechanism would thus make the system both robust to momentary changes in the sensory input (e.g. syncopation, Sioros et al., 2014) and flexible enough to adjust meter perception under persisting counterevidence from the sensory input (London, 2004; Fitch and Rosenfeld, 2007).

The short time constant observed here could also be due to the stimulus sequence design combined with a context effect restricted to inputs up to a certain level of input degradation. Indeed, while perception across domains is remarkably robust to sensory degradation, the perceptual system is limited in terms of the minimal amount of sensory cues required to elicit a percept (for evidence of these limits in meter perception see e.g. Nozaradan et al., 2012; Witek et al., 2014b; Vuust et al., 2018; Matthews et al., 2020). Even though prior context may significantly shift this limit, perceptual organization may be lost once the cues in the sensory input are too degraded. Consequently, the effects of prior context would be confined to inputs with medium amounts of degradation, thus explaining why we did not observe selective enhancement of meter frequencies in response to the most degraded sections of the sequences.

In contrast to the neural response, the effect of recent context in sensory-motor synchronization was spread across all segments. This difference between neural response and sensory-motor synchronization is in line with recent studies showing that synchronized movement can directly (Nozaradan et al., 2013, 2016c; Morillon and Baillet, 2017; Yon et al., 2018) and prospectively (Lahav et al., 2007; Chemin et al., 2014) affect sound processing in the brain. While it has been previously shown that overt movement can facilitate extraction of a periodic pulse from complex rhythmic sequences (Su and Pöppel, 2012), our results suggest that in certain situations, overt movement may impede extraction of a periodic

meter. This could be specific to situations similar to the current study, where the preceding movement is desynchronized, possibly preventing extraction of regularities gradually emerging in the sensory input. Alternatively, it could be that the location of the prior-context benefit within the sequences was variable across trials, yielding generally improved performance in the regular-to-degraded sequence after averaging. These possibilities remain to be investigated with larger samples allowing for more detailed tapping analyses.

4.5.4 Conclusion

Together, our results demonstrate that, similar to high-level perceptual organization in other domains, meter can emerge from highly complex and degraded sensory inputs. At the same time, the robustness to input degradation is limited (Witek et al., 2014b; Vuust et al., 2018) and these limits depend on context and prior experience. These observations highlight the predictive nature of perceptual processing and the importance of endogenous information (such as prior knowledge and expectations) in shaping the processing of sensory signals across domains (de Lange et al., 2018; Demarchi et al., 2019; Koelsch et al., 2019).

A common assumption in the neuroscientific literature is that meter perception can be predicted from the acoustic features of the rhythmic stimulus. In other words, rhythms with a good fit between the distribution of acoustic events and hypothetical pulses comprising meter (i.e. regular rhythms) are assumed to induce “strong” meter perception, whereas degraded rhythms are expected to induce “weak” or no meter perception (Povel and Essens, 1985; Grahn and Brett, 2007; Bengtsson et al., 2009; Grube and Griffiths, 2009; Kung et al., 2013). Together, our findings caution against a too strict stimulus-centered view, suggesting that prior experience at short and long timescales is critical to understand the mapping between sensory input and perception of rhythm. Indeed, internal processes dependent on prior experience provide a basis for cultural embeddedness of perception, which is fundamental for human musical behaviors (London et al., 2017; van der Weij et al., 2017).

4.6 Supplementary Materials

4.6.1 Supplementary Results: Control analyses of EEG data

Contribution of low-level nonlinear auditory processes. Direct comparison between the cochlear model and EEG data indicated that the results are unlikely to be fully explained by nonlinearities in the early stages of the auditory pathway. The difference in prominence of meter-related frequencies between EEG and cochlear model significantly depended on the direction of the sequence (interaction between Direction and Segment, $F_{6,838} = 8.42$, $P < 0.0001$, $BF_{10} > 100$) for segment 4 ($\beta = 0.36$, $t_{844} = 5.91$, $P < 0.0001$, 95% CI = [0.24, 0.48]). Similar to the main analysis, there was an interaction of musical training and segment ($F_{6,838} = 9.34$, $P < 0.0001$, $BF_{10} > 100$), driven by greater response at meter frequencies for musicians in segment 7 ($\beta = 0.30$, $t_{63.22} = 2.99$, $P = 0.03$, 95% CI = [0.10, 0.51]). Additionally, there was an interaction between musical training and sequence direction ($F_{1,838} = 19.4$, $P < 0.0001$, $BF_{10} > 100$). Again, directly contrasting musicians and non-musicians for each condition did not yield significant differences ($P_s > 0.12$).

Together, these results indicate that even after accounting for the response variability explained by the cochlear model, EEG responses at meter-related frequencies were significantly affected by sequence direction.

Raw amplitudes. Z-scoring EEG amplitude values across a set of frequencies was used to assess the selective variations in the amplitudes at meter-related frequencies in a manner that minimized the contribution of the overall gain. However, to demonstrate that this standardization procedure was not responsible for the context effect observed in the current study, we carried out a control analysis of the EEG response without any normalization, i.e. using raw amplitude values from the EEG spectra averaged over meter-related frequencies as the dependent measure. The amplitude at meter-related frequencies was affected by sequence direction (interaction between Direction and Segment, $F_{6,390} = 6.97$, $P < 0.0001$, $BF_{10} > 100$), and this was due to significantly larger amplitude at meter frequencies in the regular-to-degraded condition for segment 4 ($\beta = 0.37$, $t_{396} = 4.16$, $P = 0.0003$, 95% CI = [0.20, 0.55]). There was also a significant interaction between musical training and segment ($F_{6,390} = 4.44$, $P = 0.0002$, $BF_{10} = 39.56$), due to the marginally higher prominence of meter frequencies in segment 7 for musicians ($\beta = 0.30$, $t_{109.06} = 2.57$, $P =$

0.08, 95% CI = [0.07, 0.54]). These results suggest that z-scoring EEG amplitudes alone is not likely to explain the current results.

Without noise subtraction. The prominence of meter-related frequencies extracted from EEG spectra without noise subtraction was affected by sequence direction (interaction between Direction and Segment, $F_{6,390} = 5.30$, $P < 0.0001$, $BF_{10} > 100$), due to greater meter z score in segment 4 ($\beta = 0.22$, $t_{396} = 3.30$, $P = 0.007$, 95% CI = [0.09, 0.35]). Interaction between musical training and segment also reached significance ($F_{1,390} = 2.72$, $P = 0.01$, $BF_{10} = 1.45$), but there was no significant difference for either segment separately ($P_s > 0.47$). These results suggest that noise subtraction alone is not likely to explain the current results.

Head movement analysis. The prominence of meter-related frequencies in the head movement data was not significantly affected by sequence direction ($P_s > 0.44$, $BF_{s10} < 0.32$). There was a weak but significant main effect of Segment ($F_{6,390} = 2.32$, $P = 0.03$, $BF_{10} = 0.35$). A significant linear trend ($\beta = -1.04$, $t_{409} = -3.26$, $P = 0.001$, 95% CI = [-1.67, -0.41]) suggested that participants had a tendency to synchronize subtle head movements with the meter when the stimulus was more regular. Together, this control analysis suggests that the observed EEG effects are unlikely to be explained by head movement artifacts.

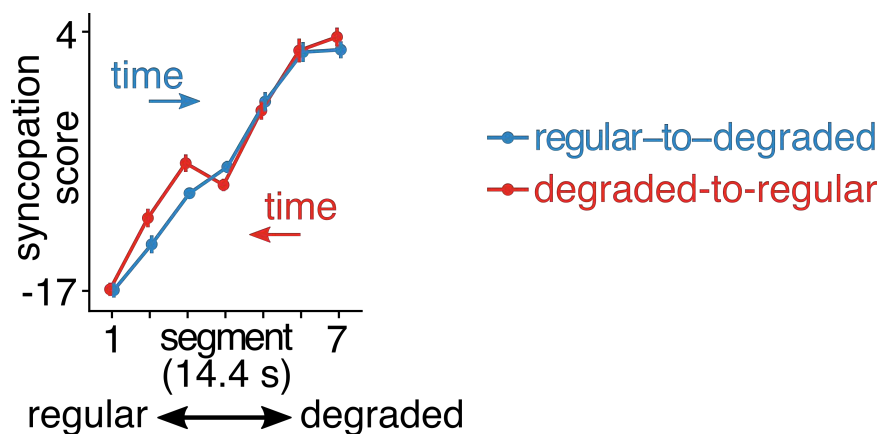


Figure 4.S1. Analysis of the stimulus sequences used in the EEG session. Syncopation scores are averaged across the 15 trials separately for each condition, with arrows indicating the direction of time for each condition. Error bars represent 95% confidence interval (Morey, 2008).

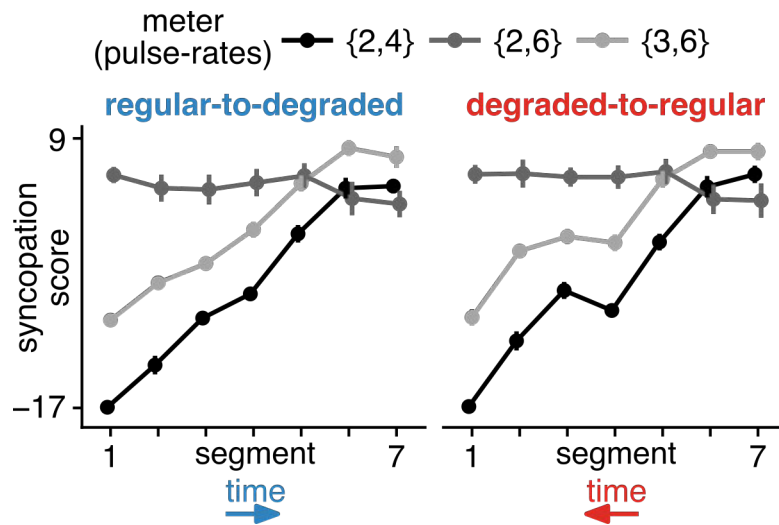


Figure 4.S2. Evolution of syncopation scores across segments, assuming different metrical interpretations of the sequences (nested pulses at rates of $\{2,4\}$, $\{3,6\}$, and $\{2,6\}$ events). Syncopation scores for $\{2,6\}$ -meter remained high throughout the sequence, whereas syncopation scores for $\{3,6\}$ -meter increased monotonically along with the $\{2,4\}$ -meter used in the main analysis. This suggests that the sequences did not change between different meters, but gradually changed from providing clear cues to a $\{2,4\}$ meter to containing little cues to any regular meter.

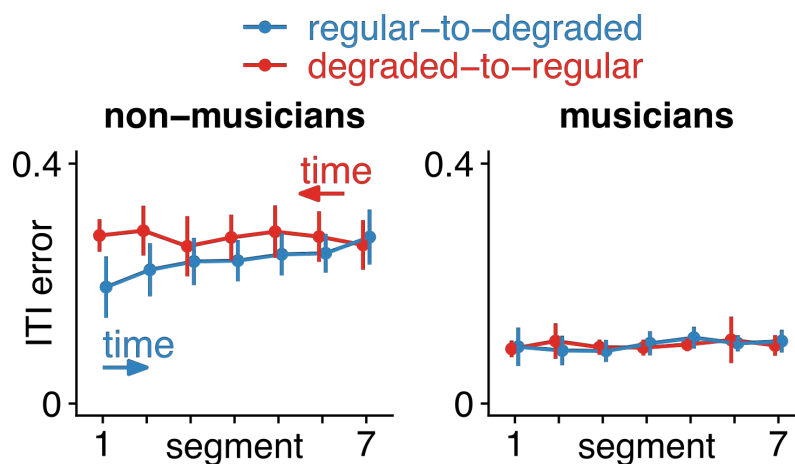


Figure 4.S3. Analysis of inter-tap interval (ITI) error across sequence directions and segments. The order of segments in the degraded-to-regular condition (red) is reversed in time (indicated by arrows) to aid the comparison of segments with equivalent amount of degradation. Mean values are shown as points, and error bars represent 95% confidence interval (Morey, 2008). (Left) Non-musicians performed worse (larger error

between their ITIs and meter periodicities) in the degraded-to-regular condition. (Right) Musicians' ITI error was low and stable over segments, suggesting that they produced ITIs close to the meter periodicities irrespective of stimulus degradation and context.

5 Discussion and Perspectives

In this thesis I explored meter as a higher-level perceptual organization of auditory rhythmic inputs in human listeners. In the first part, I aimed to provide a clear definition of the perceptual phenomenon, crucial for development of methods to measure the phenomenon in the brain and behavior. I proposed a way to approach meter perception from the viewpoint of transformation, or mapping between rhythmic inputs and internal representation of metric pulses. Based on this definition and approach to meter, I discussed direct methods to measure properties of signals relevant for meter perception. I introduced the frequency-tagging approach as a unique method to directly measure these properties in a range of complex signals, offering a number of important advantages over different direct methods. Using the frequency-tagging approach in a series of empirical studies, I aimed to contribute to the debate regarding the nature of the transformation between sound input and neural response relevant for meter perception.

Together, the current results provide evidence that multiple mechanisms are involved during meter processing. On one hand, the brain enhances periodic pulses in the input largely automatically, irrespective of attentional focus. This indicates that the mapping between rhythmic input and internal metric pulses must at least partially rely on a neural system efficient enough to support rapid transformation with little reliance on cognitive resources. This is in line with the fact that humans can often map rhythmic inputs onto metric categories within an extraordinarily short exposure time (which is not easily achieved in current computational models of meter processing, see e.g. Merchant et al., 2015a). Indeed, categorization and invariance with respect to the input seem to be supported by efficient neural systems across domains (Grootswagers et al., 2020; Rossion and Retter, 2020; Sankaran et al., 2020). In the case of meter perception, different robust processes may be engaged depending on the nature of the input. When the input lacks prominent contrast that can directly drive perception of metric pulses with moderate period durations, these pulses may be internally bootstrapped from slower or faster periodicities if these are available in the acoustic input. Firstly, the fundamental ability to subdivide slower pulses may be exploited when the input consists of a repeating pattern (as in Study 1 and 2) (Parncutt, 1994; Repp and Doggett, 2007; Repp, 2008). Secondly, the ability to interpolate and subsequently chunk fast pulses could be used to induce slower metric pulses if the input contains acoustic contrast marking a fast periodicity (Brochard et al., 2003; Repp et al. 2008; Nozaradan et al., 2011) (this was the case for the “syncopated” or “low-meter contrast”

rhythm in Study 1 and 2, and even in the most “degraded” parts of rhythmic sequences in Study 3). The specific periods of the endogenously generated metric pulses may be a result of interactions amongst intrinsic biases of the system such as preference for binary grouping and subdivision (Repp and Su, 2013), as well as preferred pulse period around 600 ms (Parncutt, 1994). These processes may be robustly engaged irrespective of attentional resources (see Study 1), and even when acoustic contrast marking pulses with intermediate period durations is completely absent in the input (Brochard et al., 2003; Bouwer and Honing, 2015).

On the other hand, the current results demonstrate that the enhancement of meter periodicities can be further boosted by bass sounds and depends on recent context and training. Therefore, besides robustness, the neural system seems to involve mechanisms that allow for considerable flexibility of the transformation from sound input towards internal representation of metric pulses. These results highlight the fact that there is no one-to-one relationship between the rhythmic input and the elicited neural or behavioral response (Nozaradan et al., 2017a). Instead, a range of endogenous and exogenous factors may shape the mapping of a rhythmic input onto metric pulses. Such neural mechanisms are crucial to support one-to-many mapping, that is, perception of different meters elicited by the same acoustic input depending on short- and long-term context (Repp, 2007; Phillips-Silver and Trainor, 2008; Chemin et al., 2014; Polak et al., 2018). Determining to what extent these mechanisms may be a result of long-term exposure and training (London et al., 2017; van der Weij et al., 2017), and whether their engagement critically depends on attention (see Study 2) remains an exciting avenue for future research.

Across the current experiments, I made sure that the results cannot be trivially explained by a number of confounds. Firstly, the overall gain of the response was accounted for by consistently using relative measures (z-scores) within the frequency-tagging approach. Secondly, by checking that the conclusions robustly hold for different ways to select meter-related and -unrelated frequencies (see section 1.2.2.5), I provided further support for the claim that the results were driven by changes in periodic contrast instead of non-specific broadband changes in the shape of the spectrum. In addition, across the studies, I ensured

that low-level nonlinearities in the auditory pathway cannot trivially account for the results by explicitly modeling them using state-of-art biophysical models.

One point that should be highlighted is that throughout the experiments I have strongly relied on the use of rhythmic stimuli with low acoustic contrast at meter periodicities (i.e., rhythmic stimuli with no prominent acoustic cues to metric pulses). This approach is not new, as ambiguous and degraded stimuli are ubiquitously used to study perception across domains, revealing important insights into the higher-level nature of the perceptual system that cannot always be obtained using simple and unambiguous sensory inputs (Dolan et al., 1997; Burton et al., 1999; Esterman and Yantis, 2010; Hsieh et al., 2010; Sohoglu et al., 2012; Mesgarani et al., 2014; Holdgraf et al., 2016). Thus, it is surprising to observe that rhythmic stimuli with low contrast at meter periodicities have been often neglected, or considered to induce less stable percept, or no percept at all (e.g. Povel and Essens, 1985; Grahn and Brett, 2007). I have hopefully provided good arguments against this view throughout this thesis. Indeed, while such inputs might be more difficult to perceptually organize, this does not mean the organization does not take place (Grootswagers et al., 2019). Instead, I believe that using such rhythms can provide crucial insights into the nature of transformation taking place during meter perception while preventing from low-level confound, as exemplified in the empirical part of this thesis. It is also important to acknowledge that there likely exists a minimum threshold in terms of the amount of sensory cues the system needs to map an acoustic input onto a stable meter. However, as shown in Study 3, it might not be possible to precisely quantify such threshold, as it likely differs across individuals. For skilled listeners, it may be enough if a single periodicity is weakly cued in the modulations of the acoustic input and these individuals may be able to bootstrap additional metric periodicities as discussed in the paragraphs above (see also Nozaradan et al., 2011; Tal et al., 2017). Furthermore, besides this general ability to map the input onto *a* stable meter, there might be additional processes that support fast and flexible contextual mapping of *specific* acoustic inputs onto *specific* meters based on enculturation and learning (London et al., 2017; van der Weij et al., 2017). This ability would allow individuals to share the same internal representation of time (i.e. meter) when moving to music that lacks unambiguous acoustic cues to a specific meter (e.g. music based on African time-line patterns), or even music that prominently cues meters with different parameters than the ones consistently perceived by enculturated listeners (e.g. ska, reggae). These

processes remain to be thoroughly investigated in future research, as discussed in section 5.1.2.

The current results also corroborate several advantages of the frequency-tagging approach in comparison to other methods that have been previously used to investigate meter perception. Firstly, the approach allows to directly assess the phenomenon, instead of relying on indirect measures that may be subject to a range of confounds. Moreover, the method is sensitive to all properties of the signal that are crucial to estimate contrast relevant for metric pulses (i.e. generalization, differentiation, time-locking). This represents a considerable advantage over other direct methods that may be sensitive only to a subset of these relevant properties (as reviewed in section 1.2.2). The approach also is free from overt behavioral response, that is, can be used to assess meter processing directly from the neural response. This is advantageous, as behavioral responses are often (i) confounded with decisional and cognitive factors, (ii) often limited to measure only one pulse at a time, and (iii) can significantly affect the percept (making it somewhat similar to the infamous Schrödinger's cat) (Su and Pöppel, 2012; Manning and Schutz, 2013). This also allows for generalization of the approach towards infants (Cirelli et al., 2016), as well as nonhuman animals that may show rudiments of meter processing, yet are not able (or motivated) to move along with the pulses (Merchant and Honing, 2014). At the same time, the approach is informed by behavior, as the main goal in cognitive neuroscience is to describe the links between sensory input, brain activity, and overt behavior. In fact, the approach offers an excellent opportunity to provide such links, and this is due to the fact that it can be used to assess contrast at meter periodicities in a wide range of different signals (e.g. sound envelopes, output of cochlear models, spiking neurons, local field potentials, high-gamma power fluctuations, discrete tapping data, continuous movement data from accelerometers, etc.). When used carefully, the frequency-tagging approach allows for direct comparison between these different signals, thus providing critical insights into the transformation relevant to meter perception.

At the same time, the approach is not without a few limitations, which I have discussed throughout the thesis, particularly in section 1.2.2 (e.g. sensitivity to certain non-specific signal changes that do not directly affect periodic contrast, or sensitivity to infrequent shifts in response phase within the analysis windows). Along with several critical points addressed

in section 5.1, these provide future directions for a comprehensive research program incorporating modeling and empirical work that will hopefully offer further insights into the strengths and potential pitfalls of frequency-tagging.

It is important to emphasize that the view of meter as transformation is distinct from approaches that quantify how well the elicited neural activity tracks changes in a stimulus feature, such as amplitude envelope (Harding et al., 2019; Fiveash et al., 2020). These approaches can only provide limited insights into the perceptual transformation towards higher-level categories. While recent developments in linear modeling offer interesting opportunities to estimate contrasts in brain activity induced by more abstract perceptual variables, i.e. going beyond the physical features of the sensory input (Brodbeck et al., 2018; Broderick et al., 2018; Di Liberto et al., 2020a), yet I expect serious limitations of such methods when applied to rhythmic phenomena in musical contexts due to the high autocorrelation inherently present in the relevant signals.

To describe mapping between sensory input and neural response (but also behavioral response) the researcher must decide on the way to represent the two signals. In other words, certain features of the stimulus and certain features of the brain activity must be chosen to describe how the first is mapped and transformed within second. This choice is not trivial and can lead to important insights into brain function (Nourski et al., 2015; Brodbeck et al., 2018; Wang, 2018; Batista and Kording, 2019; Daube et al., 2019). Nevertheless, even if the chosen representation of input and response is not optimal, comprehensively describing the mapping using a large range of inputs can still lead to valid insights into the system (Nozaradan et al., 2017a). Indeed, if we knew the optimal representation in advance, there would be no work left to do, the system would be described (Kriegeskorte and Douglas, 2019). Accordingly, contrast in audio signals can be created in many ways, from simple changes in basic parameters (e.g. amplitude or frequency modulation) to changes in abstract properties (Escera and Malmierca, 2014; Nelken, 2014). Consequently, all these parameters can generate periodic contrast in time, and may be potentially relevant for meter processing. In the current thesis, I chose to focus on simple amplitude modulations, and kept all other sound parameters fixed. Nevertheless, an exciting avenue for future research is to explore how other features contribute to contrast at meter periodicities, and whether they are relevant for mapping the rhythmic

input onto internal metric pulses. An important tool to facilitate the next steps of this enquiry may be models of subcortical auditory processing, whereby the mapping between contrast in different input features and contrast in neural responses has been already well described (Nelson and Carney, 2004; Zilany et al., 2014; Verhulst et al., 2018). Therefore, these models may be used to preprocess more complex auditory inputs, to estimate the contrasts relevant for the system at lower processing levels. This would help researchers to focus on exploring those parts of the input space where the neural transformation cannot be trivially explained by low-level mechanisms, i.e. to focus on what is yet to be discovered about higher-level functions of the brain.

Similarly, in the current thesis, I chose to focus on slow voltage changes in the brain response, as captured with surface EEG. However, there are many features of brain activity that vary over time, and thus could represent contrast at meter periodicities. These include changes in power of band-limited activity, such as beta (Iversen et al., 2009; Fujioka et al., 2015; Merchant and Bartolo, 2018) or high-gamma (Herff et al., 2020), but also multiunit activity from invasive recordings (Chang, 2015). As discussed above, frequency-tagging is not limited to either choice when representing the sensory input or the elicited brain response and thus opens a number of possibilities to explore how brain transforms rhythmic sensory input to build an internal representation of meter.

5.1 Future perspectives

Despite the fact that meter perception has been studied for a couple of decades, we still know very little about the phenomenon. This can be related to multiple factors, for instance predominant focus on Western classical and popular music (as discussed in London et al., 2017; Polak et al., 2018; Jacoby et al., 2020, see also section 1.1.2.1), the lack of definition in music theory and resulting terminological chaos (as discussed in Cohn, 2014, 2015, 2020), or forceful linking of the phenomenon with other domains, such as language (e.g. Fitch, 2013; Kotz et al., 2018). At the same time, progress in understanding has been impeded by the lack of powerful methods to measure the phenomenon in behavior and brain (as discussed in Tranchant and Vuvan, 2015; Lenc et al., 2019; Rajendran and Schnupp, 2019). Hopefully, the

ideas and methods developed in the current thesis will help to overcome some of these issues.

Considering the above, it is surprising to observe recent efforts to create mechanistic models of the phenomenon, particularly at the neural level (Large and Palmer, 2002; Large et al., 2015; Rajendran et al., 2017; Cannon and Patel, 2020). While these models may turn out being good approximations of reality, the main issue, in my opinion, is that the plausibility of these models with respect to the phenomenon cannot be easily assessed simply because we do not know enough about the phenomenon itself. For this reason, I do not explain my results in terms of any particular model. While the empirical results of the current thesis are not sufficient to build a comprehensive theory or model (and this was not the aim of the thesis), they can be used to constrain and test existing models of processes underlying meter perception. The current empirical results indicate that in humans, these processes (i) do not passively track the input but selectively enhance metric periodicities even when little attentional resources are available and even when little periodic cues are present in the acoustic input, (ii) are enhanced by non-temporal sound features such as bass sounds, (iii) can be boosted by recent stimulus history, (iv) can be enhanced by long-term training.

While the frequency-tagging method used in the current thesis does not directly measure dynamic attention or prediction, the empirical results further highlight some logical shortcomings of DAT and predictive coding approaches to meter perception (discussed in section 1.1.2.1). Namely, across the three studies, I have shown that human listeners can perceive meter when stimulated with acoustic inputs where a large number of salient acoustic events are misaligned from the perceived metric pulses. Moreover, neural responses elicited by these inputs show selective enhancement of metric periodicities. DAT would typically explain this by enhancement/suppression of responses evoked by acoustic events that are aligned/misaligned from the internally represented metric pulses (Fitzroy and Sanders, 2015; Bouwer et al., 2020). This mechanism could explain enhancement of meter periodicities in the response to inputs where majority of acoustic events are aligned with the metric pulses. However, the mechanism cannot be easily applied to inputs where most events are misaligned from the metric pulses (but note that a large number of participants in Study 2 perceived the “syncopated” rhythm in such a way, as indicated by

their tapping). Hence, to explain the current results, the DAT model would need to include a source of endogenously generated activity alongside with the attentional modulation of sound-evoked responses, which would decrease its parsimony.

Similarly, explaining the current results in terms of predictive coding is not quite straightforward. Specifically, it is not clear how the selective enhancement of meter periodicities in the neural responses could be explained by sharpening or prediction errors (two representational schemes proposed to implement Bayesian inference in the brain, see Sohoglu and Davis, 2020), if the internal representation of metric pulses directly corresponds to the prediction of the sensory input (Vuust et al., 2018).

Instead of reflecting responses evoked by the acoustic stimulus that are dynamically modulated by attentional fluctuations (DAT), prediction errors or posteriors (predictive coding), I submit a more parsimonious view, where the EEG responses partly reflect a gradual (Nozaradan et al., 2016) transformation of the auditory input towards periodic neural activity that directly represents the perceived metric pulses (Gámez et al., 2019). This activity can be used as a timing reference signal to flexibly time (i) movement (Repp et al., 2008), (ii) allocation of dynamic attention (Breska and Deouell, 2016), and (iii) temporal expectations of specific features in the sensory input (van der Weij et al., 2017). Hence, the most fundamental part of the research on meter perception is to characterize the nature of processes that are involved in this transformation (the aim of the current thesis). Directly following this endeavor is a line of research aiming to clarify whether and how additional cognitive processes (e.g. attention or expectations) utilize this internal timing signal to support adaptive behavior.

To sum up, I believe that the time to develop comprehensive and powerful models is yet to come, but extensive empirical work must be done beforehand. Therefore, I propose a number of critical points I believe should be addressed in empirical studies to move the field a step forward.

5.1.1 Functional anatomy of the transformation

A number of authors have proposed that meter perception emerges from reverberant information flow within a large network of regions involving conventionally auditory and

motor-control structures. While some authors emphasize cortico-cortical connectivity within the dorsal auditory stream (Patel and Iversen, 2014), others have argued that cortico-subcortical connections are crucial for the phenomenon (Merchant et al., 2015a). While these frameworks are not mutually exclusive, it is important to note that neither of them has been extensively tested using direct measures of meter processing. Indeed, the evidence for brain networks involved in meter perception is typically based on invasive recordings from monkeys that do not show qualitatively similar expertise in meter processing as humans (see below), or indirect methods such as fMRI (Grahn and Brett, 2007; Chen et al., 2008a; Bengtsson et al., 2009; Chapin et al., 2010; Kung et al., 2013; Li et al., 2019; Toiviainen et al., 2019; Matthews et al., 2020).

As discussed in section 1.2.1.3, the BOLD response cannot be used to directly estimate contrast at meter periodicities due to its low temporal resolution, and therefore provides limited insights into the transformation fundamental for meter perception. Instead, a promising line of research constitutes direct recordings of neural activity using electrodes implanted in the neural tissue, thus providing excellent temporal, as well as spatial resolution. This method has been widely used in animal models, particularly monkeys, yielding crucial insights into the functional network involved in sensory-motor timing (Merchant et al., 2014; Merchant and Bartolo, 2018) and mapping between sensory input and internal representation of single time intervals (Mendoza et al., 2018). While work with monkeys has been an invaluable source of information about rudimentary mechanisms critically involved in meter perception (Merchant and Honing, 2014), non-human primates show limited capability to transform a non-isochronous rhythmic input towards invariant, internal representation of metric pulses (Merchant and Honing, 2014; Honing et al., 2018). In other words, while these animals may be capable of internalizing the timing of a rhythmic input in a one-to-one fashion and internally generating such timing in an absence of rhythmic input (Merchant et al., 2015b; García-Garibay et al., 2016; Cadena-Valencia et al., 2018; Yc et al., 2018; Gámez et al., 2019), they seem incapable of going beyond the physical structure of the stimulus. In fact, it is far from trivial to train nonhuman primates to synchronize movement even with an isochronous metronome (Yc et al., 2018). In contrast to non-human primates, mapping of complex rhythmic inputs onto periodic metric pulses seems to be a spontaneous human ability widespread across cultures and musical traditions (Nettl, 2000; Savage et al., 2015). This indicates that studying neural responses in monkeys

may not provide a complete insight into the transformations relevant for meter as a higher-level perceptual phenomenon. Recent evidence has suggested that perhaps other species may serve as better models of meter processing. In particular, Patel et al. (2009) reported a case of a cockatoo exhibiting the ability to map a range of naturalistic musical stimuli onto a series of repetitive movement patterns that could be characterized by periodic contrast locked onto a metric pulse (beat). Moreover, the bird had never been trained via operant conditioning, and showed spontaneous generalization towards a large repertoire of movements (Joanne et al., 2019). However, to what extent these behaviors could occur in response to rhythmic stimuli that lack prominent contrast at meter periodicities (and therefore require significant transformation) has not been systematically investigated. Nevertheless, while non-human primates or certain bird species may help to provide important insights into the neurophysiology of meter processing, *a fundamental part of the work needs to be done in humans first*. Without thoroughly characterizing the nature of the perceptual phenomenon in humans, we cannot easily design experiments to search for correlates of the phenomenon in non-human species.

For these reasons, combining frequency-tagging with data from human patients implanted with intracranial electrodes for clinical purposes may offer an invaluable opportunity to gain insights into the functional brain networks involved in meter processing (Nozaradan et al., 2016a; Herff et al., 2020). Firstly, this approach could be used to map the transformation of a rhythmic input across the network of brain regions implicated in meter processing, including the dorsal auditory stream and supplementary motor areas (Patel and Iversen, 2014; Merchant et al., 2015a). Moreover, observations from intracerebral recordings could be complemented with studies combining surface EEG and noninvasive brain stimulation methods such as TMS and tDCS, thus assessing the causal contribution of individual regions in the network to the transformation relevant for meter perception (see e.g. Ross et al., 2018b, 2018a). On the other hand, non-invasive stimulation of the vestibular system could be used to address questions relevant to the results of Study 2 (see e.g. Trainor et al., 2009). Finally, using intracerebral recordings, as well as surface EEG combined with methods similar to Study 1 (see also Nozaradan et al., 2018), the contribution of active movement to the transformation across the network could be assessed.

5.1.2 Role of learning and training

While it is widely assumed that meter perception is a spontaneously developed human ability, this claim needs further empirical support. The issue is that “meter perception” is often taken as a unitary phenomenon, yet multiple processes may be involved in the mapping of rhythmic inputs onto internal metric pulses, as indicated by the set of studies presented in the current thesis. A particularly interesting may be the distinction between processing of rhythmic inputs that already contain high contrast at particular metric periodicities in their acoustic structure (i.e. requiring little transformation), and inputs that lack such prominent contrast (i.e. where metric pulses must be internally enhanced). Indeed, as observed e.g. in Study 3, most human listeners (perhaps except of beat-deaf individuals, see e.g. Phillips-Silver et al., 2011) have no issue with tapping the pulse in high meter contrast rhythms, yet some individuals show great difficulties to do so with low meter contrast rhythms (and this cannot be explained by lack of understanding). However, this distinction has not been addressed at all in previous studies of individual differences in meter processing (Tranchant et al., 2016), nor is part of current testing batteries (Fujii and Schlaug, 2013; Dalla Bella et al., 2017; Vuvan et al., 2018). If the transformation of low meter contrast inputs towards high meter contrast outputs can be understood as a general skill, this could be revealed in a longitudinal training study where the skill is directly enhanced through targeted exercises (developed with experienced music teachers). High experimental control of the training parameters, as opposed to ad-hoc comparisons of musicians and non-musicians, could hopefully provide important insights into the different processes involved in meter processing (Paton and Buonomano, 2018). Similar training paradigms may be used to investigate whether particular metric interpretations can be associated with specific parameters of the auditory input through learning, as suggested by recent analyses of music corpora (London et al., 2017; van der Weij et al., 2017). Such experiments would be highly relevant for the question of long-term flexibility in the mapping between sound input and internal meter with specific parameters (i.e. periods and phase), as further discussed in section 1.3. Similarly, testing whether such associations can be formed specifically with bass sounds would be informative with respect to the results of Study 2.

5.1.3 Further examining short-term context effects

In meter processing, the effect of recent context can be understood in two ways. One was explored in Study 3 of the current thesis, and the contextual variable could be characterized as input degradation. In other words, the sensory input was manipulated such that at some instances it delivered prominent cues to a specific perceptual organization (in our case acoustic contrast emphasizing particular metric pulses), whereas at other instances the input was akin to noise (containing little periodic contrast at all). Hence, two states of the system were of interest: perceptual organization (a meter internally represented) vs. no perceptual organization (no meter internally represented). Note that the conclusions hold even if these two states are not exactly discrete. This design is similar to a visual experiment where stimulus features important to elicit a higher-level percept (e.g. face or object) are gradually distorted (e.g. by scrambling, or adding noise) (Kleinschmidt et al., 2002; Esterman and Yantis, 2010; Melloni et al., 2011; Liu-Shuang et al., 2015). Yet, a complementary way to understand context effects is in terms of switching between two different perceptual organizations, and research using bi-stable inputs has yielded key insights into context effects in visual and auditory perception (Gepshtein and Kubovy, 2005; Snyder et al., 2008, 2015; Kogo et al., 2015; Silva et al., 2016; Chambers et al., 2017; Brascamp et al., 2018). Indeed, for any rhythmic input, there are always multiple plausible meters, i.e. different ways of perceptual organization (Repp, 2007; Repp et al., 2008; Chemin et al., 2014). This is similar to perceptual organization of the auditory scene into objects and streams (Pressnitzer et al., 2011; Denham et al., 2013; Sussman, 2017). Thus, the follow-up of Study 3 should explore the effect of recent context on switching between two states of the system, corresponding to perception of two different meters (e.g. a {2,2,3} and {3,2,2} meter, see Figure 1.1). To this end, the stimulus could start from creating high periodic contrast cueing one particular meter that gradually decreases as periodic contrast cueing another meter increases. In another condition, the sweep could be presented in an opposite direction, thus obtaining brain and tapping responses to physically identical segments of the input that are preceded by different contexts. The study of short-term context effects could be further extended beyond sweep designs, inspired by the large amount of prior studies on contextual effects in other perceptual domains (Snyder et al., 2009; Snyder and Weintraub, 2011; Klampfl et al., 2012; Schwiedrzik et al., 2014; St. John-Saaltink et al., 2016). Such non-

sweep designs may be better suited to assess whether the context effect translates over various parameters of the input, such as spatial location, pitch, or timbre and tempo. This could elucidate the processing level at which the context effect takes place (Snyder et al., 2015; Fritsche et al., 2017). Replicating Study 3 under different attentional conditions may be also informative in this regard.

A considerable issue when measuring effects of context on meter processing using EEG is that short segments of the signal must be analyzed, thus often leading to low signal-to-noise ratio. One way to mitigate this is by using spatial filtering optimized to suppress noise in the recordings (Cohen and Gulbinaite, 2017; Kaneshiro et al., 2020). However, one needs to be careful when optimizing spatial filters to prevent over-fitting and various biases that could be introduced to the data. For this reason, data from intracerebral EEG could be an invaluable source of insight due to their extraordinarily high signal-to-noise ratio without the need to apply complex preprocessing pipelines.

5.1.4 Meter phase

Throughout this thesis, I focused mainly on meter periods, while meter phase (i.e. the alignment of the perceived pulses with the rhythmic input) was addressed to a lesser extent. There are two reasons for this: (i) meter phase is extremely difficult to estimate from either movement or brain responses and (ii) estimating phase is not critical to quantify a periodic contrast that defines a metric pulse (see section 1.2 for discussion). Therefore, the fact that phase was not central to the current approach does not weaken the conclusions of this thesis. At the same time, I do not claim that meter phase is irrelevant to the perceptual phenomenon. Instead, phase is integral to perception of meter, and it contributes to the unique experience of genres such as ska, reggae, swing etc. In saying that, frequency-tagging is not insensitive to phase (Appelbaum et al., 2008; Cottureau et al., 2011; Rossion et al., 2012), especially when compared across conditions or brain regions. Therefore, insights into the internal representation of meter phase could be achieved using frequency-tagging. However, multiple caveats must be kept in mind. First, even though magnitude and phase are independent mathematically, in noisy signals the two are critically linked (van Diepen and Mazaheri, 2018). Therefore, when comparing phases across two conditions, amplitude differences may confound the results. Second, phase analysis may be less

powerful, as phase estimates cannot be easily pooled across harmonics in the same way as magnitudes. As a good starting point, flexibility of meter perception, i.e. one-to-many mapping, could be used to explore the phase of the neural response across conditions with physically identical sensory input, but different phase of the perceived meter. In such design, the phase of the perceived meter could be manipulated either through top-down intention (similarly to the design of Nozaradan et al., 2011, previously used to investigate meter period), or through recent context (similar to Study 3 in the current thesis).

5.2 Conclusions

This thesis builds on recent advances in music theory (Cohn, 2020) and psychology (London et al., 2017) to provide a coherent framework where meter is defined as a set of pulses that can be directly measured as periodic *contrast* in a range of signals. This definition is used as a basis to get insights into meter perception, a process of *transformation* from auditory rhythmic input to internally represented metric pulses that temporally organize perception and behavior (Agmon, 1990). The thesis presents a method based on frequency-tagging, which (i) allows appropriate measurement of signal properties relevant for periodic contrast in sound input, brain activity, and movement, and (ii) captures the mapping across these different signals. This work paves the way for future research investigating the nature of the phenomenon fundamental to our experience of listening, making, and moving to music (Honing, 2012). Furthermore, this work constitutes a unique approach to gain insight into more general higher-level perceptual processes that support temporal coordination of an individual with a complex dynamic environment. Further clarifying these processes is timely given the growing use of rhythmic auditory stimulation for the clinical rehabilitation of cognitive and motor neurological disorders (Hove and Keller, 2015)

References

- Agmon E (1990) Music Theory as Cognitive Science: Some Conceptual and Methodological Issues. *Music Percept* 7:285–308.
- Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci* 8:457–464.
- Aiken SJ, Picton TW (2008) Human cortical responses to the speech envelope. *Ear Hear* 29:139–157.
- Alain C, Izenberg A (2003) Effects of visual attentional load on low-level auditory scene analysis. *J Cogn Neurosci* 15:1063–1073.
- Alho K, Woods DL, Algazi A (1994) Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology* 31:469–479.
- Alonso-Prieto E, Van Belle G, Liu-Shuang J, Norcia AM, Rossion B (2013) The 6 Hz fundamental stimulation frequency rate for individual face discrimination in the right occipito-temporal cortex. *Neuropsychologia* 51:2863–2875.
- Amiot E (2016) *Music through Fourier space*. Cham, Switzerland: Springer International Publishing.
- Appelbaum LG, Wade AR, Pettet MW, Vildavski VY, Norcia AM (2008) Figure–ground interaction in the human visual cortex. *J Vis* 8:1–19.
- Arnal LH, Kleinschmidt A, Spinelli L, Giraud A-L, Mégevand P (2019) The rough sound of salience enhances aversion through neural synchronisation. *Nat Commun* 10:3671.
- Arzounian D, de Kerangal M, de Cheveigné A (2017) Sequential dependencies in pitch judgments. *J Acoust Soc Am* 142:3047–3057.
- Assaneo MF, Poeppel D (2018) The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Sci Adv* 4.
- Auksztulewicz R, Friston K (2015) Attentional enhancement of auditory mismatch responses: A DCM/MEG study. *Cereb Cortex* 25:4273–4283.
- Barnes R, Jones MR (2000) Expectancy, Attention, and Time. *Cogn Psychol* 41:254–311.
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* 67:1–48.
- Bates D, Maechler M, Bolker B, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4. R Packag version 11-17.

- Batista AP, Kording KP (2019) A Deep Dive to Illuminate V4 Neurons. *Trends Neurosci* 42:563–564.
- Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* 7:1129–1159.
- Bengtsson SL, Ullén F, Henrik Ehrsson H, Hashimoto T, Kito T, Naito E, Forssberg H, Sadato N (2009) Listening to rhythms activates motor and premotor cortices. *Cortex* 45:62–71.
- Bergeson TR, Trehub SE (2006) Infants Perception of Rhythmic Patterns. *Music Percept* 23:345–360.
- Bidet-Caulet A, Fischer C, Besle J, Aguera P-E, Giard M-H, Bertrand O (2007) Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J Neurosci* 27:9252–9261.
- Billig AJ, Carlyon RP (2016) Automaticity and primacy of auditory streaming: Concurrent subjective and objective measures. *J Exp Psychol Hum Percept Perform* 42:339–353.
- Bolger D, Coull JT, Schön D (2014) Metrical Rhythm Implicitly Orients Attention in Time as Indexed by Improved Target Detection and Left Inferior Parietal Activation. *J Cogn Neurosci* 26:593–605.
- Bolger D, Trost W, Schön D (2013) Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychol (Amst)* 142:238–244.
- Bouwer FL, Burgoyne JA, Odijk D, Honing H, Grahn JA (2018) What makes a rhythm complex? The influence of musical training and accent type on beat perception. *PLoS One* 13:e0190322.
- Bouwer FL, Fahrenfort JJ, Millard SK, Slagter HA (2020a) A silent disco: Persistent entrainment of low-frequency neural oscillations underlies beat-based, but not memory-based temporal expectations. *bioRxiv* Available at: <https://doi.org/10.1101/2020.01.08.899278> [Accessed January 16, 2020].
- Bouwer FL, Honing H (2015) Temporal attending and prediction influence the perception of metrical rhythm: evidence from reaction times and ERPs. *Front Psychol* 6:1094.
- Bouwer FL, Honing H, Slagter HA (2020b) Beat-based and Memory-based Temporal Expectations in Rhythm: Similar Perceptual Effects, Different Underlying Mechanisms. *J Cogn Neurosci*:1–24.
- Bouwer FL, Van Zuijen TL, Honing H (2014) Beat processing is pre-attentive for metrically simple rhythms with clear accents: An ERP study. *PLoS One* 9:e97467.

- Bouwer FL, Werner CM, Knetemann M, Honing H (2016) Disentangling beat perception from sequential learning and examining the influence of attention and musical abilities on ERP responses to rhythm. *Neuropsychologia* 85:80–90.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Brascamp J, Sterzer P, Blake R, Knapen T (2018) Multistable Perception and the Role of the Frontoparietal Cortex in Perceptual Inference. *Annu Rev Psychol* 69:77–103.
- Breska A, Deouell LY (2016) When Synchronizing to Rhythms Is Not a Good Thing: Modulations of Preparatory and Post-Target Neural Activity When Shifting Attention Away from On-Beat Times of a Distracting Rhythm. *J Neurosci* 36:7154–7166.
- Breska A, Ivry RB (2018) Double dissociation of single-interval and rhythmic temporal prediction in cerebellar degeneration and Parkinson’s disease. *Proc Natl Acad Sci* 115:12283–12288.
- 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. *Psychol Sci* 14:362–366.
- Brodbeck C, Hong LE, Simon JZ (2018) Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. *Curr Biol* 28:3976–3983.e5.
- Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Lalor EC (2018) Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. *Curr Biol* 28:803–809.e3.
- Broderick MP, Anderson AJ, Lalor EC (2019) Semantic Context Enhances the Early Auditory Encoding of Natural Speech. *J Neurosci* 39:7564–7575.
- Bruce IC, Erfani Y, Zilany MSA (2018) A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. *Hear Res* 360:40–54.
- Bruner JS, Potter MC (1964) Interference in visual recognition. *Science* 144:424–425.
- Burger B, London J, Thompson MR, Toiviainen P (2018) Synchronization to metrical levels in music depends on low-frequency spectral components and tempo. *Psychol Res* 82:1195–1211.
- Burger B, Thompson MR, Luck G, Saarikallio S, Toiviainen P (2012) Music moves us: Beat-related musical features influence regularity of music-induced movement. In: *Proceedings of the ICMP-ESCOM 2012 Joint Conference: 12th Biennial International*

- Conference for Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (Cambouropoulos E, Tsougras C, Mavromatis P, Pasiadis K, eds), pp 183–187. Thessaloniki: School of Music Studies, Aristotle University of Thessaloniki.
- Burger B, Thompson MR, Luck G, Saarikallio S, Toiviainen P (2013) Influences of rhythm- and timbre-related musical features on characteristics of music-induced movement. *Front Psychol* 4:183.
- Burger B, Thompson MR, Luck G, Saarikallio SH, Toiviainen P (2014) Hunting for the beat in the body: on period and phase locking in music-induced movement. *Front Hum Neurosci* 8:903.
- Burton AM, Wilson S, Cowan M, Bruce V (1999) Face recognition in poor-quality video: Evidence from Security Surveillance. *Psychol Sci* 10:243–248.
- Burunat I, Tsatsishvili V, Brattico E, Toiviainen P (2017) Coupling of Action-Perception Brain Networks during Musical Pulse Processing: Evidence from Region-of-Interest-Based Independent Component Analysis. *Front Hum Neurosci* 11:230.
- Butler MJ (2006) *Unlocking the groove: Rhythm, meter, and musical design in electronic dance music*. Bloomington: Indiana University Press.
- Cadena-Valencia J, García-Garibay O, Merchant H, Jazayeri M, Lafuente V de (2018) Entrainment and maintenance of an internal metronome in premotor cortex. *Elife* 7:1–23.
- Câmara GS, Danielsen A (2018) Groove. In: *Oxford Handbook of Critical Concepts in Music Theory* (Rehding A, Rings S, eds), pp 271–294. New York: Oxford University Press.
- Cameron DJ, Bentley J, Grahm JA (2015) Cross-cultural influences on rhythm processing: Reproduction, discrimination, and beat tapping. *Front Psychol* 6:366.
- Cameron DJ, Grahm JA (2014) Enhanced timing abilities in percussionists generalize to rhythms without a musical beat. *Front Hum Neurosci* 8:1003.
- Cameron DJ, Grahm JA (2016) The Neuroscience of Rhythm. In: *The Oxford Handbook of Music Psychology* (Hallam S, Cross I, Thaut M, eds), pp 357–368. Oxford, UK: Oxford University Press.
- Cameron DJ, Zioga I, Lindsen JP, Pearce MT, Wiggins GA, Potter K, Bhattacharya J (2019) Neural entrainment is associated with subjective groove and complexity for performed but not mechanical musical rhythms. *Exp Brain Res* 237:1981–1991.

- Cannon J, Patel A (2020) How beat perception coopts motor neurophysiology: a proposal. bioRxiv Available at: <https://doi.org/10.1101/805838> [Accessed September 1, 2020].
- Carney LH, Li T, McDonough JM (2015) Speech Coding in the Brain: Representation of Vowel Formants by Midbrain Neurons Tuned to Sound Fluctuations. *eNeuro* 2:4–15.
- Celma-Miralles A, de Menezes RF, Toro JM (2016) Look at the Beat, Feel the Meter: Top–Down Effects of Meter Induction on Auditory and Visual Modalities. *Front Hum Neurosci* 10:108.
- Celma-Miralles A, Toro JM (2019) Ternary meter from spatial sounds: Differences in neural entrainment between musicians and non-musicians. *Brain Cogn* 136:103594.
- Chait M, Ruff CC, Griffiths TD, McAlpine D (2012) Cortical responses to changes in acoustic regularity are differentially modulated by attentional load. *Neuroimage* 59:1932–1941.
- Chambers C, Akram S, Adam V, Pelofi C, Sahani M, Shamma S, Pressnitzer D (2017) Prior context in audition informs binding and shapes simple features. *Nat Commun* 8:15027.
- Chandrasekaran B, Kraus N (2010) The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology* 47:236–246.
- Chang EF (2015) Towards large-scale, human-based, mesoscopic neurotechnologies. *Neuron* 86:68–78.
- Chapin HL, Zanto T, Jantzen KJ, Kelso SJA, Steinberg F, Large EW (2010) Neural responses to complex auditory rhythms: The role of attending. *Front Psychol* 1:224.
- Chemin B, Mouraux A, Nozaradan S (2014) Body Movement Selectively Shapes the Neural Representation of Musical Rhythms. *Psychol Sci* 25:2147–2159.
- Chen JL, Penhune VB, Zatorre RJ (2008a) Listening to musical rhythms recruits motor regions of the brain. *Cereb Cortex* 18:2844–2854.
- Chen JL, Penhune VB, Zatorre RJ (2008b) Moving on Time: Brain Network for Auditory–Motor Synchronization is Modulated by Rhythm Complexity and Musical Training. *J Cogn Neurosci* 20:226–239.
- Chiu MG (2018) Form as meter: metric forms through Fourier space. Available at: <https://search.proquest.com/openview/778fe8e5c0a7282bd1782d453cf7ae83/1?pq-origsite=gscholar&cbl=18750&diss=y> [Accessed January 13, 2020].
- Cicchini GM, Anobile G, Burr DC (2014) Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proc Natl Acad Sci* 111:7867–7872.

- Cicchini GM, Arrighi R, Cecchetti L, Giusti M, Burr DC (2012) Optimal Encoding of Interval Timing in Expert Percussionists. *J Neurosci* 32:1056–1060.
- Cicchini GM, Burr DC (2018) Serial effects are optimal. *Behav Brain Sci* 41:e229.
- Cicchini GM, Mikellidou K, Burr DC (2018) The functional role of serial dependence. *Proc R Soc B Biol Sci* 285:20181722.
- Cirelli LK, Spinelli C, Nozaradan S, Trainor LJ (2016) Measuring neural entrainment to beat and meter in infants: Effects of music background. *Front Neurosci* 10:229.
- Coffey EBJ, Herholz SC, Chepesiuk AMP, Baillet S, Zatorre RJ (2016) Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat Commun* 7.
- Coffey EBJ, Nicol T, White-Schwoch T, Chandrasekaran B, Krizman J, Skoe E, Zatorre RJ, Kraus N (2019) Evolving perspectives on the sources of the frequency-following response. *Nat Commun* 10.
- Cohen MX, Gulbinaite R (2017) Rhythmic entrainment source separation: Optimizing analyses of neural responses to rhythmic sensory stimulation. *Neuroimage* 147:43–56.
- Cohn R (2014) Meter Without Tactus. In: *Music Theory Annual Meeting* (Lochhead J, Will R, eds). Brunswick, ME: American Musicological Society.
- Cohn R (2015) Why We Don't Teach Meter, and Why We Should. *J Music Theory Pedagogy* 29:1–19.
- Cohn R (2016) A Platonic Model of Funky Rhythms. *Music Theory Online* 22:1–17.
- Cohn R (2020) Meter. In: *Oxford Handbook of Critical Concepts in Music Theory* (Rehding A, Rings S, eds), pp 207–233. New York, NY: Oxford University Press.
- Cooper GW, Meyer LB (1963) *The Rhythmic Structure of Music*. Chicago, IL: University of Chicago Press.
- Cope TE, Grube M, Singh B, Burn DJ, Griffiths TD (2014) The basal ganglia in perceptual timing: Timing performance in Multiple System Atrophy and Huntington's disease. *Neuropsychologia* 52:73–81.
- Costa-Faidella J, Sussman ES, Escera C (2017) Selective entrainment of brain oscillations drives auditory perceptual organization. *Neuroimage* 159:195–206.
- Cottareau BR, McKee SP, Ales JM, Norcia AM (2011) Disparity-tuned population responses from human visual cortex. *J Neurosci* 31:954–965.
- Dalla Bella S, Farrugia N, Benoit CE, Begel V, Verga L, Harding E, Kotz SA (2017) BAASTA: Battery for the Assessment of Auditory Sensorimotor and Timing Abilities. *Behav Res*

Methods 49:1128–1145.

Daube C, Ince RAA, Gross J (2019) Simple Acoustic Features Can Explain Phoneme-Based Predictions of Cortical Responses to Speech. *Curr Biol* 29:1924–1937.

de Lange FP, Heilbron M, Kok P (2018) How Do Expectations Shape Perception? *Trends Cogn Sci* 22:764–779.

Dean I, Harper NS, McAlpine D (2005) Neural population coding of sound level adapts to stimulus statistics. *Nat Neurosci* 8:1684–1689.

Demarchi G, Sanchez G, Weisz N (2019) Automatic and feature-specific prediction-related neural activity in the human auditory system. *Nat Commun* 10:1–11.

Denham SL, Gyimesi K, Stefanics G, Winkler I (2013) Perceptual bistability in auditory streaming: How much do stimulus features matter? *Learn Percept* 5:73–100.

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

Di Liberto GM, O’Sullivan JA, Lalor EC (2015) Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr Biol* 25:2457–2465.

Di Liberto GM, Pelofi C, Bianco R, Patel P, Mehta AD, Herrero JL, de Cheveigné A, Shamma S, Mesgarani N (2020a) Cortical encoding of melodic expectations in human temporal cortex. *Elife* 9:26–74.

Di Liberto GM, Pelofi C, Shamma S, de Cheveigné A (2020b) Musical expertise enhances the cortical tracking of the acoustic envelope during naturalistic music listening. *Acoust Sci Technol* 41:361–364.

Di Liberto GM, Peter V, Kalashnikova M, Goswami U, Burnham D, Lalor EC (2018) Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *Neuroimage* 175:70–79.

Di Liberto GM, Wong D, Melnik GA, de Cheveigné A (2019) Low-frequency cortical responses to natural speech reflect probabilistic phonotactics. *Neuroimage* 196:237–247.

Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc Natl Acad Sci* 109:11854–11859.

Doelling KB, Assaneo MF, Bevilacqua D, Pesaran B, Poeppel D (2019) An oscillator model better predicts cortical entrainment to music. *Proc Natl Acad Sci*:201816414.

Doelling KB, Poeppel D (2015) Cortical entrainment to music and its modulation by expertise. *Proc Natl Acad Sci* 112:201508431.

- Dolan RJ, Fink GR, Rolls ET, Booth M, Holmes A, Frackowiak RSJ, Friston KJ (1997) How the brain learns to see objects and faces in an impoverished context. *Nature* 389:596–599.
- Draguhn A, Buzsaki G (2004) Neuronal Oscillations in Cortical Networks. *Science* 304:1926–1929.
- Drake C, Botte MC (1993) Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Percept Psychophys* 54:277–286.
- Drake C, El Heni J Ben (2003) Synchronizing with Music: Intercultural Differences. *Ann N Y Acad Sci* 999:429–437.
- Drake C, Jones MR, Baruch C (2000) The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition* 77:251–288.
- Dyson BJ, Lain C, He Y (2005) Effects of visual attentional load on low-level auditory scene analysis. *Cogn Affect Behav Neurosci* 5:319–338.
- Eck D (2003) A Positive-Evidence Model for Rhythmical Beat Induction. *J New Music Res* 30:187–200.
- Eger E, Henson RN, Driver J, Dolan RJ (2007) Mechanisms of top-down facilitation in perception of visual objects studied by fMRI. *Cereb Cortex* 17:2123–2133.
- Escera C, Malmierca MS (2014) The auditory novelty system: An attempt to integrate human and animal research. *Psychophysiology* 51:111–123.
- Escoffier N, Sheng DYJ, Schirmer A (2010) Unattended musical beats enhance visual processing. *Acta Psychol (Amst)* 135:12–16.
- Essens PJ, Povel D-J (1985) Metrical and nonmetrical representations of temporal patterns. *Percept Psychophys* 37:1–7.
- Esterman M, Yantis S (2010) Perceptual Expectation Evokes Category-Selective Cortical Activity. *Cereb Cortex* 20:1245–1253.
- Etard O, Reichenbach T (2019) Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension. *J Neurosci* 39:5750–5759.
- Fischer J, Whitney D (2014) Serial dependence in visual perception. *Nat Neurosci* 17:738–743.
- Fitch WT (2013) Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Front Syst Neurosci* 7:68.
- Fitch WT, Rosenfeld AJ (2007) Perception and production of syncopated rhythms. *Music*

Percept 25:43–58.

- Fitzroy AB, Sanders LD (2015) Musical Meter Modulates the Allocation of Attention across Time. *J Cogn Neurosci* 27:2339–2351.
- Fiveash A, Schön D, Canette LH, Morillon B, Bedoin N, Tillmann B (2020) A stimulus-brain coupling analysis of regular and irregular rhythms in adults with dyslexia and controls. *Brain Cogn* 140:105531.
- Friston K (2005) A theory of cortical responses. *Philos Trans R Soc B Biol Sci* 360:815–836.
- Fritsche M, Mostert P, de Lange FP (2017) Opposite Effects of Recent History on Perception and Decision. *Curr Biol* 27:590–595.
- Fu KMG, Foxe JJ, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2001) Attention-dependent suppression of distracter visual input can be cross-modally cued as indexed by anticipatory parieto-occipital alpha-band oscillations. *Cogn Brain Res* 12:145–152.
- Fujii S, Schlaug G (2013) The Harvard Beat Assessment Test (H-BAT): A battery for assessing beat perception and production and their dissociation. *Front Hum Neurosci* 7:771.
- Fujioka T, Ross B, Trainor LJ (2015) Beta-Band Oscillations Represent Auditory Beat and Its Metrical Hierarchy in Perception and Imagery. *J Neurosci* 35:15187–15198.
- Fujioka T, Trainor LJ, Large EW, Ross B (2012) Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations. *J Neurosci* 32:1791–1802.
- Fujioka T, Zendel BR, Ross B (2010) Endogenous Neuromagnetic Activity for Mental Hierarchy of Timing. *J Neurosci* 30:3458–3466.
- Gabrielsson A (1993) The Complexities of Rhythm. In: *Psychology and Music: The Understanding of Melody and Rhythm* (Tighe T, Dowling W, eds), pp 93–120. New York, NY: Psychology Press.
- Gabrielsson ALF (2003) Music Performance Research at the Millennium. *Psychol Music* 31:221–272.
- Galambos R, Makeig S, Talmachoff P (1981) A 40-Hz auditory potential recorded from the human scalp. *Proc Natl Acad Sci U S A* 78:2643–2647.
- Gámez J, Mendoza G, Prado L, Betancourt A, Merchant H (2019) The amplitude in periodic neural state trajectories underlies the tempo of rhythmic tapping. *PLoS Biol* 17:e3000054.
- Gao X, Gentile F, Rossion B (2018) Fast periodic stimulation (FPS): a highly effective approach in fMRI brain mapping. *Brain Struct Funct* 223:2433–2454.

- García-Garibay O, Cadena-Valencia J, Merchant H, de Lafuente V (2016) Monkeys share the human ability to internally maintain a temporal rhythm. *Front Psychol* 7:1971.
- Garner WR, Gottwald RL (1968) The perception and learning of temporal patterns. *Q J Exp Psychol* 20:97–109.
- Geiser E, Sandmann P, Jäncke L, Meyer M (2010) Refinement of metre perception - training increases hierarchical metre processing. *Eur J Neurosci* 32:1979–1985.
- Geiser E, Ziegler E, Jancke L, Meyer M (2009) Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex* 45:93–102.
- Gepshtein S, Kubovy M (2005) Stability and change in perception: Spatial organization in temporal context. *Exp Brain Res* 160:487–495.
- Gescheider GA (1997) *Psychophysics : The Fundamentals*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.
- Ghinst M Vander, Bourguignon M, Niesen M, Wens V, Hassid S, Choufani G, Jousmäki V, Hari R, Goldman S, De Tiège X (2019) Cortical tracking of speech-in-noise develops from childhood to adulthood. *J Neurosci* 39:2938–2950.
- Goldstone RL, Hendrickson AT (2009) Categorical perception. *Wiley Interdiscip Rev Cogn Sci* 1:69–78.
- Grahn JA (2012) Neural Mechanisms of Rhythm Perception: Current Findings and Future Perspectives. *Top Cogn Sci* 4:585–606.
- Grahn JA, Brett M (2007) Rhythm and Beat Perception in Motor Areas of the Brain. *J Cogn Neurosci* 19:893–906.
- Grahn JA, Rowe JB (2009) Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J Neurosci* 29:7540–7548.
- Grahn JA, Rowe JB (2013) Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity. *Cereb Cortex* 23:913–921.
- Greb B (2017) *The Language of Drumming*. Milwaukee: Hal Leonard Corporation.
- Grootswagers T, Robinson AK, Shatek SM, Carlson TA (2019) Untangling featural and conceptual object representations. *Neuroimage* 202:116083.
- Grootswagers T, Robinson AK, Shatek SM, Carlson TA (2020) The neural dynamics underlying prioritisation of task-relevant information. *bioRxiv* Available at: <https://doi.org/10.1101/2020.06.25.172643> [Accessed July 3, 2020].
- Grube M, Cooper FE, Chinnery PF, Griffiths TD (2010) Dissociation of duration-based and

- beat-based auditory timing in cerebellar degeneration. *Proc Natl Acad Sci U S A* 107:11597–11601.
- Grube M, Griffiths TD (2009) Metricality-enhanced temporal encoding and the subjective perception of rhythmic sequences. *Cortex* 45:72–79.
- Guiliana M (2018) *Exploring Your Creativity on the Drumset*. Milwaukee: Hal Leonard Corporation.
- Gutschalk A, Rupp A, Dykstra AR (2015) Interaction of streaming and attention in human auditory cortex. *PLoS One* 10:e0118962.
- Halpern A, Darwin C (1982) Duration discrimination in a series of rhythmic events. *Percept Psychophys* 31:86–89.
- Hamilton LS, Edwards E, Chang EF (2018) A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. *Curr Biol* 28:1860-1871.e4.
- Handel S, Oshinsky JS (1981) The meter of syncopated auditory polyrhythms. *Percept Psychophys* 30:1–9.
- Hannon EE, Snyder JS, Eerola T, Krumhansl CL (2004) The Role of Melodic and Temporal Cues in Perceiving Musical Meter. *J Exp Psychol Hum Percept Perform* 30:956–974.
- Hannon EE, Soley G, Levine RS (2011) Constraints on infants’ musical rhythm perception: Effects of interval ratio complexity and enculturation. *Dev Sci* 14:865–872.
- Hannon EE, Soley G, Ullal S (2012a) Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *J Exp Psychol Hum Percept Perform* 38:543–548.
- Hannon EE, Trehub SE (2005a) Tuning in to musical rhythms: infants learn more readily than adults. *Proc Natl Acad Sci U S A* 102:12639–12643.
- Hannon EE, Trehub SE (2005b) Metrical categories in infancy and adulthood. *Psychol Sci* 16:48–55.
- Hannon EE, Vanden Bosch der Nederlanden CM, Tichko P (2012b) Effects of perceptual experience on children’s and adults’ perception of unfamiliar rhythms. *Ann N Y Acad Sci* 1252:92–99.
- Harding EE, Sammler D, Henry MJ, Large EW, Kotz SA (2019) Cortical tracking of rhythm in music and speech. *Neuroimage* 185:96–101.
- Haroush K, Hochstein S, Deouell LY (2010) Momentary fluctuations in allocation of attention: Cross-modal effects of visual task load on auditory discrimination. *J Cogn*

- Neurosci 22:1440–1451.
- Harrison PMC, Müllensiefen D (2018) Development and Validation of the Computerised Adaptive Beat Alignment Test (CA-BAT). *Sci Rep* 8.
- Hartmann WM (1997) *Signals, sound, and sensation*. New York, NY: AIP Press.
- Haumann NT, Vuust P, Bertelsen F, Garza-Villarreal EA (2018) Influence of musical enculturation on brain responses to metric deviants. *Front Neurosci* 12:218.
- Hegd  J, Kersten D (2010) A link between visual disambiguation and visual memory. *J Neurosci* 30:15124–15133.
- Heggli OA, Konvalinka I, Kringelbach ML, Vuust P (2019) Musical interaction is influenced by underlying predictive models and musical expertise. *Sci Rep* 9.
- Henry MJ, Herrmann B, Grahn JA (2017) What can we learn about beat perception by comparing brain signals and stimulus envelopes? *PLoS One* 12:e0172454.
- Herff SA, Herff C, Milne AJ, Johnson GD, Shih JJ, Krusienski DJ (2020) Prefrontal High Gamma in ECoG tags periodicity of musical rhythms in perception and imagination. *eNeuro* 7:1–11.
- Hervais-Adelman A, Davis MH, Johnsrude IS, Carlyon RP (2008) Perceptual Learning of Noise Vcoded Words: Effects of Feedback and Lexicality. *J Exp Psychol Hum Percept Perform* 34:460–474.
- Hickey P, Merseal H, Patel AD, Race E (2020) Memory in time: Neural tracking of low-frequency rhythm dynamically modulates memory formation. *Neuroimage* 213:116693.
- Holdgraf CR, De Heer W, Pasley B, Rieger J, Crone N, Lin JJ, Knight RT, Theunissen FE (2016) Rapid tuning shifts in human auditory cortex enhance speech intelligibility. *Nat Commun* 7:13654.
- Holmes E, Purcell DW, Carlyon RP, Gockel HE, Johnsrude IS (2018) Attentional Modulation of Envelope-Following Responses at Lower (93–109 Hz) but Not Higher (217–233 Hz) Modulation Rates. *J Assoc Res Otolaryngol* 19:83–97.
- Honing H (2012) Without it no music: Beat induction as a fundamental musical trait. *Ann N Y Acad Sci* 1252:85–91.
- Honing H, Bouwer FL (2018) Rhythm. In: *Foundations in Music Psychology: Theory and Research* (Rentfrow PJ, Levitin DJ, eds), pp 33–70. MIT Press.
- Honing H, Bouwer FL, H den GP (2014) Perceiving temporal regularity in music: The role of

- auditory Event-Related Potentials (ERPs) in probing beat perception. In: *Advances in Experimental Medicine and Biology* (Merchant H, de Lafuente V, eds), pp 305–323. New York, NY: Springer New York.
- Honing H, Bouwer FL, Prado L, Merchant H (2018) Rhesus Monkeys (*Macaca mulatta*) sense isochrony in rhythm, but not the beat: Additional support for the gradual audiomotor evolution hypothesis. *Front Neurosci* 12:475.
- Honing H, Ladinig O (2009) Exposure Influences Expressive Timing Judgments in Music. *J Exp Psychol Hum Percept Perform* 35:281–288.
- Hove MJ, Keller PE (2015) Impaired movement timing in neurological disorders: Rehabilitation and treatment strategies. *Ann N Y Acad Sci* 1337:111–117.
- Hove MJ, Keller PE, Krumhansl CL (2007) Sensorimotor synchronization with chords containing tone-onset asynchronies. *Percept Psychophys* 69:699–708.
- Hove MJ, Marie C, Bruce IC, Trainor LJ (2014) Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proc Natl Acad Sci* 111:10383–10388.
- Hsieh PJ, Vul E, Kanwisher N (2010) Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex. *J Neurophysiol* 103:1501–1507.
- Hurley BK, Fink LK, Janata P (2018) Mapping the dynamic allocation of temporal attention in musical patterns. *J Exp Psychol Hum Percept Perform* 44:1694–1711.
- Iemi L, Chaumon M, Crouzet SM, Busch NA (2017) Spontaneous Neural Oscillations Bias Perception by Modulating Baseline Excitability. *J Neurosci* 37:807–819.
- Istók E, Friberg A, Huotilainen M, Tervaniemi M (2013) Expressive Timing Facilitates the Neural Processing of Phrase Boundaries in Music: Evidence from Event-Related Potentials. *PLoS One* 8:e55150.
- Iversen JR, Patel AD (2008) The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. In: *Proceedings of the 10th International Conference on Music Perception and Cognition* (Miyazaki K, Adachi M, Hiraga Y, Nakajima Y, Tsuzaki M, eds), pp 465–468. Sapporo, Japan: ICMPC10.
- Iversen JR, Repp BH, Patel AD (2009) Top-down control of rhythm perception modulates early auditory responses. *Ann N Y Acad Sci* 1169:58–73.
- Ivry RB, Hazeltine RE (1995) Perception and Production of Temporal Intervals Across a Range of Durations: Evidence for a Common Timing Mechanism. *J Exp Psychol Hum*

Percept Perform 21:3–18.

Jacoby N et al. (2020) Cross-Cultural Work in Music Cognition: Challenges, Insights, and Recommendations. *Music Percept An Interdiscip J* 37:185–195.

Jacoby N, McDermott JH (2017) Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Curr Biol* 27:359–370.

Jacques C, Retter TL, Rossion B (2016) A single glance at natural face images generate larger and qualitatively different category-selective spatio-temporal signatures than other ecologically-relevant categories in the human brain. *Neuroimage* 137:21–33.

Janata P, Tomic ST, Haberman JM (2012) Sensorimotor coupling in music and the psychology of the groove. *J Exp Psychol Gen* 141:54–75.

Jazayeri M, Shadlen MN (2010) Temporal context calibrates interval timing. *Nat Neurosci* 13:1020–1026.

Jeffreys H (1998) *The theory of probability*. Oxford: Oxford University Press.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front Hum Neurosci* 4:186.

Joanne R, Keehn J, Iversen JR, Schulz I, Patel AD (2019) Spontaneity and diversity of movement to music are not uniquely human. *Curr Biol* 29:R621–R622.

Johndro H, Jacobs L, Patel AD, Race E (2019) Temporal predictions provided by musical rhythm influence visual memory encoding. *Acta Psychol (Amst)* 200.

Johnson JA, Zatorre RJ (2005) Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates. *Cereb Cortex* 15:1609–1620.

Jonas J, Jacques C, Liu-Shuang J, Brissart H, Colnat-Coulbois S, Maillard L, Rossion B (2016) A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proc Natl Acad Sci* 113:E4088–E4097.

Jones MR, Boltz M (1989) Dynamic Attending and Responses to Time. *Psychol Rev* 96:459–491.

Jones MR, Moynihan H, Mackenzie N, Puente J (2002) Temporal Aspects of Stimulus-Driven Attending in Dynamic Arrays. *Psychol Sci* 13:313–319.

Jones MR, Yee W (1997) Sensitivity to Time Change: The Role of Context and Skill. *J Exp Psychol Hum Percept Perform* 23:693–709.

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2000) Removal of eye activity artifacts from visual event-related potentials in normal and clinical

- subjects. *Clin Neurophysiol* 111:1745–1758.
- Kalender B, Trehub SE, Schellenberg EG (2013) Cross-cultural differences in meter perception. *Psychol Res* 77:196–203.
- Kaneshiro B, Nguyen DT, Norcia AM, Dmochowski JP, Berger J (2020) Natural music evokes correlated EEG responses reflecting temporal structure and beat. *Neuroimage* 214:116559.
- Keele SW, Nicoletti R, Lvrly RI, Pokorny RA (1989) Mechanisms of perceptual timing: Beat-based or interval-based judgements? *Psychol Res* 50:251–256.
- Keitel C, Maess B, Schröger E, Müller MM (2013) Early visual and auditory processing rely on modality-specific attentional resources. *Neuroimage* 70:240–249.
- Keitel C, Schröger E, Saupe K, Müller MM (2011) Sustained selective intermodal attention modulates processing of language-like stimuli. *Exp Brain Res* 213:321–327.
- Keller P (1999) Attending in complex musical interactions: The adaptive dual role of meter. *Aust J Psychol* 51:166–175.
- Keller P, Burnham D (2005) Musical Meter in Attention to Multipart Rhythm. *Music Percept* 22:629–661.
- Keller PE (2012) What movement force reveals about cognitive processes in music performance. *Art motion II*:115–153.
- Keller PE, Novembre G, Hove MJ (2014) Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philos Trans R Soc Lond B Biol Sci* 369:20130394.
- Keller PE, Repp BH (2005) Staying offbeat: Sensorimotor syncopation with structured and unstructured auditory sequences. *Psychol Res* 69:292–309.
- Keshishian M, Akbari H, Khalighinejad B, Herrero JL, Mehta AD, Mesgarani N (2020) Estimating and interpreting nonlinear receptive field of sensory neural responses with deep neural network models. *Elife* 9:1–24.
- Klompfl S, David S V, Yin P, Shamma SA, Maass W (2012) A quantitative analysis of information about past and present stimuli encoded by spikes of A1 neurons. *J Neurophysiol* 108:1366–1380.
- Kleinschmidt A, Büchel C, Hutton C, Friston KJ, Frackowiak RS. (2002) The Neural Structures Expressing Perceptual Hysteresis in Visual Letter Recognition. *Neuron* 34:659–666.
- Koelsch S, Vuust P, Friston K (2019) Predictive Processes and the Peculiar Case of Music.

- Trends Cogn Sci 23:63–77.
- Kogo N, Hermans L, Stuer D, van Ee R, Wagemans J (2015) Temporal dynamics of different cases of bi-stable figure-ground perception. *Vision Res* 106:7–19.
- Kotz SA, Ravignani A, Fitch WT (2018) The Evolution of Rhythm Processing. *Trends Cogn Sci* 22:896–910.
- Kriegeskorte N, Douglas PK (2019) Interpreting encoding and decoding models. *Curr Opin Neurobiol* 55:167–179.
- Kuchibhotla K, Bathellier B (2018) Neural encoding of sensory and behavioral complexity in the auditory cortex. *Curr Opin Neurobiol* 52:65–71.
- Kunert R, Jongman SR (2017) Entrainment to an Auditory Signal: Is Attention Involved? *J Exp Psychol Gen* 146:77–88.
- Kung S-J, Chen JL, Zatorre RJ, Penhune VB (2013) Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *J Cogn Neurosci* 25:401–420.
- Ladinig O, Honing H, Háden G, Winkler I (2009) Probing Attentive and Preattentive Emergent Meter in Adult Listeners without Extensive Music Training. *Music Percept* 26:377–386.
- Lahav A, Saltzman E, Schlaug G (2007) Action Representation of Sound: Audiomotor Recognition Network While Listening to Newly Acquired Actions. *J Neurosci* 27:308–314.
- Lakatos P, Gross J, Thut G (2019) A new unifying account of the roles of neuronal entrainment. *Curr Biol* 29:1–16.
- Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *J Neurophysiol* 94:1904–1911.
- Lappe C, Trainor LJ, Herholz SC, Pantev C (2011) Cortical Plasticity Induced by Short-Term Multimodal Musical Rhythm Training. *PLoS One* 6:21493.
- Large EW (2000a) On synchronizing body movement to music. *Hum Mov Sci* 19:527–566.
- Large EW (2000b) Rhythm categorization in context. In: *Proceedings of the International Conference on Music Perception and Cognition* (Woods C, Luck G, Brochard R, O'Neill S, Sloboda J, eds). Keele, UK: Keele University, Psychology Department.
- Large EW, Herrera JA, Velasco MJ (2015) Neural Networks for Beat Perception in Musical

- Rhythm. *Front Syst Neurosci* 9:159.
- Large EW, Jones MR (1999) The dynamic of attending: How People Track Time-Varying Events. *Psychol Res* 106:119–159.
- Large EW, Palmer C (2002) Perceiving temporal regularity in music. *Cogn Sci* 26:1–37.
- Large EW, Snyder JS (2009) Pulse and meter as neural resonance. *Ann N Y Acad Sci* 1169:46–57.
- Lavie N, Dalton P (2014) Load Theory of Attention and Cognitive Control. In: *The Oxford Handbook of Attention* (Nobre AC, Kastner S, eds), pp 56–75. Oxford: Oxford University Press.
- Lee KM, Skoe E, Kraus N, Ashley R (2009) Selective Subcortical Enhancement of Musical Intervals in Musicians. *J Neurosci* 29:5832–5840.
- Lee MD, Wagenmakers E-J (2014) *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge university press.
- Leek MR (2001) Adaptive procedures in psychophysical research. *Percept Psychophys* 63:1279–1292.
- Lenc T, Keller PE, Varlet M, Nozaradan S (2018) Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proc Natl Acad Sci* 115:8221–8226.
- Lenc T, Keller PE, Varlet M, Nozaradan S (2019) Reply to Rajendran and Schnupp: Frequency tagging is sensitive to the temporal structure of signals. *Proc Natl Acad Sci* 116:2781–2782.
- Lenc T, Keller PE, Varlet M, Nozaradan S (2020) Neural and behavioral evidence for frequency-selective context effects in rhythm processing in humans. *Cereb Cortex Commun*.
- Lenth R (2018) *Emmeans: Estimated marginal means, aka least-squares means*. R Packag version 11 Available at: <https://cran.r-project.org/package=emmeans> [Accessed January 11, 2018].
- Leonard MK, Baud MO, Sjerps MJ, Chang EF (2016) Perceptual restoration of masked speech in human cortex. *Nat Commun* 7:13619.
- Lerdahl F, Jackendoff R (1983) *A Generative Theory of Tonal Music*. Cambridge (MA): MIT Press.
- Lewis A (2018) Clapping on 2 & 4? Available at: <https://youtu.be/hm37TiQQUpk> [Accessed September 3, 2020].

- Ley A, Vroomen J, Formisano E (2014) How learning to abstract shapes neural sound representations. *Front Neurosci* 8:132.
- Li Q, Liu G, Wei D, Liu Y, Yuan G, Wang G (2019) Distinct neuronal entrainment to beat and meter: Revealed by simultaneous EEG-fMRI. *Neuroimage* 194:128–135.
- Lieberman A, Fischer J, Whitney D (2014) Serial dependence in the perception of faces. *Curr Biol* 24:2569–2574.
- Lieberman A, Zhang K, Whitney D (2016) Serial dependence promotes object stability during occlusion. *J Vis* 16:16.
- Lieberman MC (1978) Auditory-nerve response from cats raised in a low-noise chamber. *J Acoust Soc Am* 63:442–455.
- Liu-Shuang J, Ales J, Rossion B, Norcia AM, H. TRB, M. C, J.-M. G (2015) Separable effects of inversion and contrast-reversal on face detection thresholds and response functions: A sweep VEP study. *J Vis* 15:11–11.
- Lochy A, Jacques C, Maillard L, Colnat-Coulbois S, Rossion B, Jonas J (2018) Selective visual representation of letters and words in the left ventral occipito-temporal cortex with intracerebral recordings. *Proc Natl Acad Sci U S A* 115:E7595–E7604.
- Lohse M, Bajo VM, King AJ, Willmore BDB (2020) Neural circuits underlying auditory contrast gain control and their perceptual implications. *Nat Commun* 11.
- London J (2004) *Hearing in Time*. New York, United States: Oxford University Press.
- London J, Polak R, Jacoby N (2017) Rhythm histograms and musical meter: A corpus study of Malian percussion music. *Psychon Bull Rev* 24:474–480.
- Longuet-Higgins HC, Lee CS (1984) The rhythmic interpretation of monophonic music. *Music Percept* 1:424–440.
- Longuet Higgins HC, Lee CS (1982) The perception of musical rhythms. *Perception* 11:115–128.
- Lozano-Soldevilla D, VanRullen R (2019) The Hidden Spatial Dimension of Alpha: 10-Hz Perceptual Echoes Propagate as Periodic Traveling Waves in the Human Brain. *Cell Rep* 26:374–380.e4.
- Luck SJ (2014) *An introduction to the event-related potential technique*. Cambridge, United States: MIT press.
- Macmillan NA, Creelman CD (2005) *Detection theory: A user's guide*, 2nd ed. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.

- Maes P-J, Leman M, Palmer C, Wanderley MM (2014) Action-based effects on music perception. *Front Psychol* 4:1008.
- Makeig S (2002) Response: Event-related brain dynamics – unifying brain electrophysiology. *Trends Neurosci* 25:390.
- Manassi M, Liberman A, Chaney W, Whitney D (2017) The perceived stability of scenes: Serial dependence in ensemble representations. *Sci Rep* 7:1971.
- Manning F, Schutz M (2013) “Moving to the beat” improves timing perception. *Psychon Bull Rev* 20:1133–1139.
- Marois R, Ivanoff J (2005) Capacity limits of information processing in the brain. *Trends Cogn Sci* 9:296–305.
- Martens PA (2011) The ambiguous tactus: Tempo, subdivision benefit, and three listener strategies. *Music Percept* 28:433–448.
- Masutomi K, Barascud N, Kashino M, McDermott JH, Chait M (2016) Sound segregation via embedded repetition is robust to inattention. *J Exp Psychol Hum Percept Perform* 42:386–400.
- Matthews TE, Witek MAG, Lund T, Vuust P, Penhune VB (2020) The sensation of groove engages motor and reward networks. *Neuroimage* 214:116768.
- Mazaheri A, van Schouwenburg MR, Dimitrijevic A, Denys D, Cools R, Jensen O (2014) Region-specific modulations in oscillatory alpha activity serve to facilitate processing in the visual and auditory modalities. *Neuroimage* 87:356–362.
- McAuley JD, Jones MR (2003) Modeling Effects of Rhythmic Context on Perceived Duration: A Comparison of Interval and Entrainment Approaches to Short-Interval Timing. *J Exp Psychol Hum Percept Perform* 29:1102–1125.
- McAuley JD, Kidd GR (1998) Effect of Deviations from Temporal Expectations on Tempo Discrimination of Isochronous Tone Sequences. *J Exp Psychol Hum Percept Perform* 24:1786–1800.
- McAuley JD, Semple P (1999) The effect of tempo and musical experience on perceived beat. *Aust J Psychol* 51:176–187.
- McKinney MF, Moelants D (2006) Ambiguity in Tempo Perception: What Draws Listeners to Different Metrical Levels? *Music Percept* 24:155–166.
- Meddis R (1986) Simulation of mechanical to neural transduction in the auditory receptor. *J Acoust Soc Am* 79:702–711.

- Melloni L, Schwiedrzik CM, Muller N, Rodriguez E, Singer W (2011) Expectations Change the Signatures and Timing of Electrophysiological Correlates of Perceptual Awareness. *J Neurosci* 31:1386–1396.
- Mendoza G, Méndez JC, Pérez O, Prado L, Merchant H (2018) Neural basis for categorical boundaries in the primate pre-SMA during relative categorization of time intervals. *Nat Commun* 9:1–17.
- Merchant H, Bartolo R (2018) Primate beta oscillations and rhythmic behaviors. *J Neural Transm* 125:461–470.
- Merchant H, Bartolo R, Pérez O, Méndez JC, Mendoza G, Gámez J, Yc K, Prado L (2014) Neurophysiology of timing in the hundreds of milliseconds: Multiple layers of neuronal clocks in the medial premotor areas. *Adv Exp Med Biol* 829:143–154.
- Merchant H, Grahn J, Trainor L, Rohrmeier M, Fitch WT (2015a) Finding the beat: a neural perspective across humans and non-human primates. *Philos Trans R Soc B Biol Sci* 370:20140093.
- Merchant H, Honing H (2014) Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front Neurosci* 7:274.
- Merchant H, Pérez O, Bartolo R, Méndez JC, Mendoza G, Gámez J, Yc K, Prado L (2015b) Sensorimotor neural dynamics during isochronous tapping in the medial premotor cortex of the macaque. *Eur J Neurosci* 41:586–602.
- Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485:233–236.
- Mesgarani N, David S V., Fritz JB, Shamma SA (2014) Mechanisms of noise robust representation of speech in primary auditory cortex. *Proc Natl Acad Sci* 111:6792–6797.
- Miller NS, Auley JDMC (2005) Tempo sensitivity in isochronous tone sequences: The multiple-look model revisited. *Percept Psychophys* 67:1150–1160.
- Milne AJ, Bulger D, Herff SA (2017) Exploring the space of perfectly balanced rhythms and scales. *J Math Music* 11:101–133.
- Milne AJ, Herff SA (2020) The perceptual relevance of balance, evenness, and entropy in musical rhythms. *Cognition* 203:104233.
- Mo J, Schroeder CE, Ding M (2011) Attentional modulation of alpha oscillations in macaque inferotemporal cortex. *J Neurosci* 31:878–882.

- Molloy K, Lavie N, Chait M (2019) Auditory figure-ground segregation is impaired by high visual load. *J Neurosci* 39:1699–1708.
- Molloy K, Lavie N, Chait M (2020) Does auditory processing rely on encapsulated, or domain-general computational resources? *Acoust Sci Technol* 41:13–15.
- Moore BC, Peters RW, Glasberg BR (1993) Detection of temporal gaps in sinusoids: effects of frequency and level. *J Acoust Soc Am* 93:1563–1570.
- Moore BCJ, Glasberg BR, Varathanathan A, Schlittenlacher J (2016) A Loudness Model for Time-Varying Sounds Incorporating Binaural Inhibition. *Trends Hear* 20:1–16.
- Morey RD (2008) Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutor Quant Methods Psychol* 4:61–64.
- Morey RD, Rouder JN (2014) BayesFactor: Computation of Bayes factors for common designs. R Packag version 0912-42 9 Available at: <https://cran.r-project.org/package=BayesFactor> [Accessed May 7, 2018].
- Morillon B, Baillet S (2017) Motor origin of temporal predictions in auditory attention. *Proc Natl Acad Sci* 114:E8913–E8921.
- Mouraux A, Iannetti GD, Colon E, Nozaradan S, Legrain V, Plaghki L (2011) Nociceptive Steady-State Evoked Potentials Elicited by Rapid Periodic Thermal Stimulation of Cutaneous Nociceptors. *J Neurosci* 31:6079–6087.
- Murphy S, Spence C, Dalton P (2017) Auditory perceptual load: A review. *Hear Res* 352:40–48.
- Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin Neurophysiol* 118:2544–2590.
- Narain D, Remington ED, Zeeuw CID, Jazayeri M (2018) A cerebellar mechanism for learning prior distributions of time intervals. *Nat Commun* 9:469.
- Nelken I (2014) Stimulus-specific adaptation and deviance detection in the auditory system: experiments and models. *Biol Cybern* 108:655–663.
- Nelson PC, Carney LH (2004) A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *J Acoust Soc Am* 116:2173–2186.
- Nettl B (2000) An ethnomusicologist contemplates universals in musical sound and musical culture. In: *The origins of music* (Wallin NL, Merker B, Brown S, eds), pp 463–472. Massachusetts: MIT Press.
- Norcia AM, Appelbaum LG, Ales JM, Cottareau BR, Rossion B (2015) The steady-state visual

- evoked potential in vision research: A review. *J Vis* 15:4.
- Notter MP, Hanke M, Murray MM, Geiser E (2018) Encoding of Auditory Temporal Gestalt in the Human Brain. *Cereb Cortex*:1–10.
- Nourski K V., Brugge JF (2011) Representation of temporal sound features in the human auditory cortex. *Rev Neurosci* 22:187–203.
- Nourski K V., Steinschneider M, Rhone AE, Oya H, Kawasaki H, Howard MA, McMurray B (2015) Sound identification in human auditory cortex: Differential contribution of local field potentials and high gamma power as revealed by direct intracranial recordings. *Brain Lang* 148:37–50.
- Nozaradan S (2014) Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos Trans R Soc Lond B Biol Sci* 369:1–10.
- Nozaradan S, Keller PE, Rossion B, Mouraux A (2017a) EEG Frequency-Tagging and Input–Output Comparison in Rhythm Perception. *Brain Topogr* 31:153–160.
- Nozaradan S, Mouraux A, Jonas J, Colnat-Coulbois S, Rossion B, Maillard L (2016a) Intracerebral evidence of rhythm transform in the human auditory cortex. *Brain Struct Funct* 222:2389–2404.
- Nozaradan S, Peretz I, Keller PE (2016b) Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. *Sci Rep* 6:20612.
- Nozaradan S, Peretz I, Missal M, Mouraux A (2011) Tagging the neuronal entrainment to beat and meter. *J Neurosci* 31:10234–10240.
- Nozaradan S, Peretz I, Mouraux A (2012) Selective Neuronal Entrainment to the Beat and Meter Embedded in a Musical Rhythm. *J Neurosci* 32:17572–17581.
- Nozaradan S, Schönwiesner M, Caron-Desrochers L, Lehmann A (2016c) Enhanced brainstem and cortical encoding of sound during synchronized movement. *Neuroimage* 142:231–240.
- Nozaradan S, Schönwiesner M, Keller PE, Lenc T, Lehmann A (2018) Neural bases of rhythmic entrainment in humans: critical transformation between cortical and lower-level representations of auditory rhythm. *Eur J Neurosci* 47:321–332.
- Nozaradan S, Schwartze M, Obermeier C, Kotz SA (2017b) Specific contributions of basal ganglia and cerebellum to the neural tracking of rhythm. *Cortex* 95:156–168.
- Nozaradan S, Zerouali Y, Peretz I, Mouraux A (2013) Capturing with EEG the neural

- entrainment and coupling underlying sensorimotor synchronization to the beat. *Cereb Cortex* 25:736–747.
- O’Sullivan J, Herrero J, Smith E, Schevon C, McKhann GM, Sheth SA, Mehta AD, Mesgarani N (2019) Hierarchical Encoding of Attended Auditory Objects in Multi-talker Speech Perception. *Neuron* 104:1195-1209.e3.
- O’Sullivan JA, Shamma SA, Lalor EC (2015) Evidence for neural computations of temporal coherence in an auditory scene and their enhancement during active listening. *J Neurosci* 35:7256–7263.
- Obleser J, Henry MJ, Lakatos P (2017) What do we talk about when we talk about rhythm? *PLoS Biol* 15:128.
- Okamoto H, Stracke H, Bermudez P, Pantev C (2011) Sound Processing Hierarchy within Human Auditory Cortex. *J Cogn Neurosci* 23:1855–1863.
- Okawa H, Suefusa K, Tanaka T (2017) Neural Entrainment to Auditory Imagery of Rhythms. *Front Hum Neurosci* 11:493.
- Oppenheim A V, Schafer RW (2009) *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ: Pearson Education.
- Pablos Martin X, Deltenre P, Hoonhorst I, Markessis E, Rossion B, Colin C (2007) Perceptual biases for rhythm: The Mismatch Negativity latency indexes the privileged status of binary vs non-binary interval ratios. *Clin Neurophysiol* 118:2709–2715.
- Palmer C, Krumhansl CL (1990) Mental representations for musical meter. *J Exp Psychol Hum Percept Perform* 16:728–741.
- Park H, Ince RAA, Schyns PG, Thut G, Gross J (2015) Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Curr Biol* 25:1649–1653.
- Parncutt R (1994) A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms. *Music Percept* 11:409–464.
- Pashler H (2001) Perception and production of brief durations: Beat-based versus interval-based timing. *J Exp Psychol Hum Percept Perform* 27:485–493.
- Patel AD, Iversen JR (2014) The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front Syst Neurosci* 8:57.
- Patel AD, Iversen JR, Bregman MR, Schulz I (2009) Experimental Evidence for Synchronization to a Musical Beat in a Nonhuman Animal. *Curr Biol* 19:827–830.

- Paton JJ, Buonomano D V. (2018) The Neural Basis of Timing: Distributed Mechanisms for Diverse Functions. *Neuron* 98:687–705.
- Patterson RD, Holdsworth J (1996) A functional model of neural activity patterns and auditory images. *Adv Speech, Hear Lang Process* 3:547–563.
- Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb Cortex* 23:1378–1387.
- Penhune VB, Zatorre RJ (2019) Rhythm and time in the premotor cortex. *PLoS Biol* 17:e3000293.
- Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004) Attentional modulation of human auditory cortex. *Nat Neurosci* 7:658–663.
- Phillips-Silver J, Aktipis CA, Bryant GA (2010) The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music Percept* 28:3–14.
- 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* 49:961–969.
- Phillips-Silver J, Trainor LJ (2005) Feeling the beat: Movement influences infant rhythm perception. *Science* 308:1430.
- Phillips-Silver J, Trainor LJ (2008) Vestibular influence on auditory metrical interpretation. *Brain Cogn* 67:94–102.
- Picton, John, Dimitrijevic, Purcell (2003) Human auditory steady-state responses. *Int J Audiol* 42:177–219.
- Pikovsky A, Kurths J, Rosenblum M, Kurths J (2003) Synchronization: a universal concept in nonlinear sciences. Cambridge, United Kingdom: Cambridge university press.
- Poeppel D, Assaneo MF (2020) Speech rhythms and their neural foundations. *Nat Rev Neurosci* 21:322–334.
- Polak R (2018) The lower limit for meter in dance drumming from West Africa. *Empir Musicol Rev* 12:205.
- Polak R, Jacoby N, Fischinger T, Goldberg D, Holzapfel A, London J (2018) Rhythmic prototypes across cultures: A comparative study of tapping synchronization. *Music Percept* 36:1–23.
- Polak R, London J, Jacoby N (2016) Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: A corpus study of malian

- jembe drumming. *Front Neurosci* 10:285.
- Povel D-J, Essens PJ (1985) Perception of temporal patterns. *Music Percept* 2:411–440.
- Povel DJ, Okkerman H (1981) Accents in equitone sequences. *Percept Psychophys* 30:565–572.
- Pressing J (2002) *Black Atlantic Rhythm: Its Computational and Transcultural Foundations*. *Music Percept* 19:285–310.
- Pressnitzer D, Suied C, Shamma S (2011) Auditory scene analysis: the sweet music of ambiguity. *Front Hum Neurosci* 5:158.
- Rabinowitz NC, Willmore BDB, Schnupp JWH, King AJ (2011) Contrast Gain Control in Auditory Cortex. *Neuron* 70:1178–1191.
- Rajendran VG, Harper NS, Garcia-Lazaro JA, Lesica NA, Schnupp JWH (2017) Midbrain adaptation may set the stage for the perception of musical beat. *Proc R Soc London B Biol Sci* 284:1–7.
- Rajendran VG, Harper NS, Schnupp JWH (2020) Auditory cortical representation of music favours the perceived beat. *R Soc Open Sci* 7:191194.
- Rajendran VG, Schnupp JWH (2019) Frequency tagging cannot measure neural tracking of beat or meter. *Proc Natl Acad Sci* 116:2779–2780.
- Raviv O, Ahissar M, Loewenstein Y (2012) How Recent History Affects Perception: The Normative Approach and Its Heuristic Approximation. *PLoS Comput Biol* 8:e1002731.
- Regan D (1989) *Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. New York, NY: Elsevier.
- Repp BH (1992) Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *J Acoust Soc Am* 92:2546–2568.
- Repp BH (2005) Sensorimotor synchronization: a review of the tapping literature. *Psychon Bull Rev* 12:969–992.
- Repp BH (2007) Hearing a melody in different ways: Multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. *Cognition* 102:434–454.
- Repp BH (2008) Multiple temporal references in sensorimotor synchronization with metrical auditory sequences. *Psychol Res* 72:79–98.
- Repp BH (2010a) Self-generated interval subdivision reduces variability of synchronization with a very slow metronome. *Music Percept* 27:389–397.
- Repp BH (2010b) Sensorimotor synchronization and perception of timing: Effects of music

- training and task experience. *Hum Mov Sci* 29:200–213.
- Repp BH, Doggett R (2007) Tapping to a Very Slow Beat: A Comparison of Musicians and Nonmusicians. *Source Music Percept An Interdiscip J* 24:154101:367–376.
- Repp BH, Iversen JR, Patel AD (2008) Tracking an Imposed Beat within a Metrical Grid. *Music Percept* 26:1–18.
- Repp BH, Su Y-H (2013) Sensorimotor synchronization: A review of recent research (2006–2012). *Psychon Bull Rev* 20:403–452.
- Retter TL, Rossion B (2016) Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia* 91:9–28.
- Riecke L, Peters JC, Valente G, Kemper VG, Formisano E, Sorger B (2017) Frequency-Selective Attention in Auditory Scenes Recruits Frequency Representations Throughout Human Superior Temporal Cortex. *Cereb Cortex* 27:3002–3014.
- Riecke L, Scharke W, Valente G, Gutschalk A (2014) Sustained selective attention to competing amplitude-modulations in human auditory cortex. *PLoS One* 9:e108045.
- Rimmele JM, Poeppel D, Ghitza O (2020) Acoustically driven cortical delta oscillations underpin perceptual chunking. *bioRxiv* Available at: <https://doi.org/10.1101/2020.05.16.099432> [Accessed September 5, 2020].
- Robinson BL, McAlpine D (2009) Gain control mechanisms in the auditory pathway. *Curr Opin Neurobiol* 19:402–407.
- Roeske TC, Tchernichovski O, Poeppel D, Jacoby N (2020) Categorical Rhythms Are Shared between Songbirds and Humans. *Curr Biol* 30:1–12.
- Ross B, Borgmann C, Draganova R, Roberts LE, Pantev C (2000) A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *J Acoust Soc Am* 108:679–691.
- Ross B, Draganova R, Picton TW, Pantev C (2003) Frequency specificity of 40-Hz auditory steady-state responses. *Hear Res* 186:57–68.
- Ross J, Iversen J, Balasubramaniam R, Ross JM (2018a) Dorsal Premotor Contributions to Auditory Rhythm Perception: Causal Transcranial Magnetic Stimulation Studies of Interval, Tempo, and Phase. *bioRxiv* Available at: <https://doi.org/10.1101/368597> [Accessed September 1, 2020].
- Ross JM, Iversen JR, Balasubramaniam R (2018b) The Role of Posterior Parietal Cortex in Beat-based Timing Perception: A Continuous Theta Burst Stimulation Study. *J Cogn*

- Neurosci 26:1–10.
- Rossion B (2014) Understanding individual face discrimination by means of fast periodic visual stimulation. *Exp Brain Res* 232:1599–1621.
- Rossion B, Prieto EA, Boremanse A, Kuefner D, Van Belle G (2012) A steady-state visual evoked potential approach to individual face perception: Effect of inversion, contrast-reversal and temporal dynamics. *Neuroimage* 63:1585–1600.
- Rossion B, Retter TL (2020) Face Perception. In: *The Cognitive Neurosciences* (Poeppel D, Mangun GR, Gazzaniga MS, eds), pp 129–140. Cambridge, United States: MIT Press.
- Rossion B, Retter TL, Liu-Shuang J (2020) Understanding human individuation of unfamiliar faces with oddball fast periodic visual stimulation and electroencephalography. *Eur J Neurosci*.
- Rossion B, Torfs K, Jacques C, Liu-Shuang J (2015) Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *J Vis* 15:1–18.
- Rüsseler J, Altenmüller E, Nager W, Kohlmetz C, Münte TF (2002) Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neurosci Lett* 308:33–36.
- Sankaran N, Carlson TA, Thompson WF (2020) The rapid emergence of musical pitch structure in human cortex. *J Neurosci* 40:2108–2118.
- Savage PE, Brown S, Sakai E, Currie TE (2015) Statistical universals reveal the structures and functions of human music. *Proc Natl Acad Sci* 112:8987–8992.
- Schaefer RS, Vlek RJ, Desain P (2011) Decomposing rhythm processing: Electroencephalography of perceived and self-imposed rhythmic patterns. *Psychol Res* 75:95–106.
- Schirmer A, Wijaya M, Chiu MH, Maess B, Gunter TC (2020) Interpersonal Synchrony Special Issue Musical rhythm effects on visual attention are non-rhythmical: evidence against metrical entrainment. *Soc Cogn Affect Neurosci*:1–14.
- Schulze HH (1978) The detectability of local and global displacements in regular rhythmic patterns. *Psychol Res* 40:173–181.
- Schwiedrzik CM, Melloni L, Schurger A (2018) Mooney face stimuli for visual perception research. *PLoS One* 13:e0200106.
- Schwiedrzik CM, Ruff CC, Lazar A, Leitner FC, Singer W, Melloni L (2014) Untangling

- perceptual memory: Hysteresis and adaptation map into separate cortical networks. *Cereb Cortex* 24:1152–1164.
- Shannon R V, Zeng F-G, Kamath V, Wygonski J, Ekelid M (1995) Speech Recognition with Primarily Temporal Cues. *Science* 270:303–304.
- Shera CA, Guinan JJ, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci* 99:3318–3323.
- Shomstein S, Yantis S (2004) Control of attention shifts between vision and audition in human cortex. *J Neurosci* 24:10702–10706.
- Silva LL, Abrahamyan A, Gardner JL, Carandini M, Dakin SC (2016) Adaptable history biases in human perceptual decisions. *Proc Natl Acad Sci* 113:E3548–E3557.
- Sioros G, Miron M, Davies M, Gouyon F, Madison G (2014) Syncopation creates the sensation of groove in synthesized music examples. *Front Psychol* 5:1036.
- Skoe E, Kraus N (2010) Auditory brain stem response to complex sounds: a tutorial. *Ear Hear* 31:302–324.
- Slaney M (1998) Auditory toolbox, version 2. Interval Res Corp Tech Rep 1998–010:1–52.
- Smith JO (2007) Mathematics of the discrete Fourier transform (DFT): with audio applications. Charleston, SC: BookSurge Publishing.
- Snyder JS, Carter OL, Lee SK, Hannon EE, Alain C (2008) Effects of context on auditory stream segregation. *J Exp Psychol Hum Percept Perform* 34:1007–1016.
- Snyder JS, Holder WT, Weintraub DM, Carter OL, Alain C (2009) Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology* 46:1208–1215.
- Snyder JS, Krumhansl CL (2001) Tapping to Ragtime: Cues to Pulse Finding. *Music Percept* 18:455–489.
- Snyder JS, Large EW (2005) Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cogn Brain Res* 24:117–126.
- Snyder JS, Schwiedrzik CM, Vitela AD, Melloni L (2015) How previous experience shapes perception in different sensory modalities. *Front Hum Neurosci* 9:594.
- Snyder JS, Weintraub DM (2011) Pattern Specificity in the Effect of Prior Δf on Auditory Stream Segregation. *J Exp Psychol Hum Percept Perform* 37:1649–1656.
- Snyder JS, Weintraub DM (2013) Loss and persistence of implicit memory for sound: Evidence from auditory stream segregation context effects. *Attention, Perception,*

- Psychophys 75:1059–1074.
- Sohn H, Narain D, Meirhaeghe N, Jazayeri M (2019) Bayesian Computation through Cortical Latent Dynamics. *Neuron* 103:934–947.e5.
- Sohoglu E, Davis MH (2016) Perceptual learning of degraded speech by minimizing prediction error. *Proc Natl Acad Sci U S A* 113:E1747–E1756.
- Sohoglu E, Davis MH (2020) Rapid computations of spectrotemporal prediction error support perception of degraded speech. *Elife* 9:e58077.
- Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive Top-Down Integration of Prior Knowledge during Speech Perception. *J Neurosci* 32:8443–8453.
- Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2014) Top-down influences of written text on perceived clarity of degraded speech. *J Exp Psychol Hum Percept Perform* 40:186–199.
- Song C, Pearce M, Harte C (2015) SYNPHY: a python toolkit for syncopation modelling. In: *Proceedings of the 12th International Conference on Sound and Music Computing (SMC-15)* (Timoney J, Lysaght T, eds), pp 295–300. Maynooth, Ireland: Maynooth University, Department of Computer Science.
- St. John-Saaltink E, Kok P, Lau HC, de Lange FP (2016) Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *J Neurosci* 36:6186–6192.
- Stanislaw H, Todorov N (1999) Calculation of signal detection theory measures. *Behav Res Methods, Instruments, Comput* 31:137–149.
- Stein T, Peelen M V. (2015) Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *J Exp Psychol Gen* 144:1089–1104.
- Stupacher J, Hove MJ, Janata P (2016) Audio Features Underlying Perceived Groove and Sensorimotor Synchronization in Music. *Music Percept An Interdiscip J* 33:571–589.
- Stupacher J, Wood G, Witte M (2017) Neural entrainment to polyrhythms: A comparison of musicians and non-musicians. *Front Neurosci* 11:208.
- Su YH, Pöppel E (2012) Body movement enhances the extraction of temporal structures in auditory sequences. *Psychol Res* 76:373–382.
- Suárez-Pinilla M, Seth AK, Roseboom W (2018) Serial dependence in the perception of visual variance. *J Vis* 18:4.
- Sussman ES (2017) Auditory Scene Analysis: An Attention Perspective. *J Speech Lang Hear Res* 60:2989.

- Sussman ES, Chen S, Sussman-Fort J, Dinces E (2014) The five myths of MMN: Redefining how to Use MMN in basic and clinical research. *Brain Topogr* 27:553–564.
- Sussman ES, Horváth J, Winkler I, Mark O (2007) The role of attention in the formation of auditory streams. *Percept Psychophys* 69:136–152.
- Tal I, Large EW, Rabinovitch E, Wei Y, Schroeder CE, Poeppel D, Zion Golumbic E (2017) Neural Entrainment to the Beat: The “Missing-Pulse” Phenomenon. *J Neurosci* 37:6331–6341.
- Teki S, Barascud N, Picard S, Payne C, Griffiths TD, Chait M (2016) Neural correlates of auditory figure-ground segregation based on temporal coherence. *Cereb Cortex* 26:3669–3680.
- Teki S, Chait M, Kumar S, von Kriegstein K, Griffiths TD (2011) Brain Bases for Auditory Stimulus-Driven Figure-Ground Segregation. *J Neurosci* 31:164–171.
- Teng X, Tian X, Rowland J, Poeppel D (2017) Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. *PLoS Biol* 15:e2000812.
- Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, Mehta PR, Goodyer IM, Fletcher PC (2015) Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proc Natl Acad Sci U S A* 112:13401–13406.
- Thaut MH, McIntosh GC, Hoemberg V (2015) Neurobiological foundations of neurologic music therapy: Rhythmic entrainment and the motor system. *Front Psychol* 5:1185.
- Tierney A, Kraus N (2014) Neural Entrainment to the Rhythmic Structure of Music. *J Cogn Neurosci* 27:400–408.
- Todd NPM, Cody FWJ, Banks JR (2000) A saccular origin of frequency tuning in myogenic vestibular evoked potentials?: implications for human responses to loud sounds. *Hear Res* 141:180–188.
- Todd NPM, Lee CS (2015a) The sensory-motor theory of rhythm and beat induction 20 years on: a new synthesis and future perspectives. *Front Hum Neurosci* 9:444.
- Todd NPM, Lee CS (2015b) Source analysis of electrophysiological correlates of beat induction as sensory-guided action. *Front Psychol* 6:1178.
- Toiviainen P, Burunat I, Brattico E, Vuust P, Alluri V (2019) The chronnectome of musical beat. *Neuroimage* 216:116191.

- Toiviainen P, Luck G, Thompson MR (2010) Embodied Meter: Hierarchical Eigenmodes in Music - Induced Movement. *Music Percept* 28:59–70.
- Toiviainen P, Snyder JS (2003) Tapping to Bach: Resonance-Based Modeling of Pulse. *Music Percept* 21:43–80.
- Tomic ST, Janata P (2008) Beyond the beat: Modeling metric structure in music and performance. *J Acoust Soc Am* 124:4024–4041.
- Torchiano M (2020) effsize: Efficient Effect Size Computation. R Packag version 08 Available at: <https://cran.r-project.org/package=effsize> [Accessed April 9, 2019].
- Trainor LJ, Gao X, Lei J jiang, Lehtovaara K, Harris LR (2009) The primal role of the vestibular system in determining musical rhythm. *Cortex* 45:35–43.
- Trainor LJ, Marie C, Bruce IC, Bidelman GM (2014) Explaining the high voice superiority effect in polyphonic music: Evidence from cortical evoked potentials and peripheral auditory models. *Hear Res* 308:60–70.
- Trainor LJ, Unrau A (2009) Extracting the beat: An experience-dependent complex integration of multisensory information involving multiple levels of the nervous system. *Empir Musicol Rev* 4:32–36.
- Tranchant P, Vuvan DT (2015) Current conceptual challenges in the study of rhythm processing deficits. *Front Neurosci* 9:197.
- Tranchant P, Vuvan DT, Peretz I (2016) Keeping the Beat: A Large Sample Study of Bouncing and Clapping to Music. *PLoS One* 11:e0160178.
- Trehub SE, Hannon EE (2009) Conventional rhythms enhance infants' and adults' perception of musical patterns. *Cortex* 45:110–118.
- van der Weij B, Pearce MT, Honing H (2017) A probabilistic model of meter perception: Simulating enculturation. *Front Psychol* 8:824.
- Van Diepen RM, Foxe JJ, Mazaheri A (2019) The functional role of alpha-band activity in attentional processing: the current zeitgeist and future outlook. *Curr Opin Psychol* 29:229–238.
- van Diepen RM, Mazaheri A (2017) Cross-sensory modulation of alpha oscillatory activity: suppression, idling, and default resource allocation Foxe J, ed. *Eur J Neurosci* 45:1431–1438.
- van Diepen RM, Mazaheri A (2018) The Caveats of observing Inter-Trial Phase-Coherence in Cognitive Neuroscience. *Sci Rep* 8:2990.

- van Noorden L, Moelants D (1999) Resonance in the Perception of Musical Pulse. *J New Music Res* 28:43–66.
- van Rijn H (2016) Accounting for memory mechanisms in interval timing: A review. *Curr Opin Behav Sci* 8:245–249.
- Verhulst S, Altoè A, Vasilkov V (2018) Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss. *Hear Res* 360:55–75.
- Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *J Stat Softw* 36:1–48.
- Vulf (2017) VULFPECK /// Joe Dart Beastly Solo III. Available at: <https://youtu.be/uLu8HV7UKnk> [Accessed September 3, 2020].
- Vuust P, Dietz MJ, Witek MAG, Kringelbach ML (2018) Now you hear it: A predictive coding model for understanding rhythmic incongruity. *Ann N Y Acad Sci* 1423:19–29.
- Vuust P, Gebauer LK, Witek MAG (2014) Neural Underpinnings of Music: The Polyrhythmic Brain. In: *Neurobiology of Interval Timing* (Merchant H, de Lafuente V, eds), pp 339–356. New York, NY: Springer New York.
- Vuust P, Ostergaard L, Pallesen KJ, Bailey C, Roepstorff A (2009) Predictive coding of music - Brain responses to rhythmic incongruity. *Cortex* 45:80–92.
- Vuust P, Pallesen KJ, Bailey C, Van Zuijen TL, Gjedde A, Roepstorff A, Østergaard L (2005) To musicians, the message is in the meter: Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage* 24:560–564.
- Vuust P, Witek MAG (2014) Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Front Psychol* 5:1111.
- Vuvan DT, Paquette S, Mignault Goulet G, Royal I, Felezeu M, Peretz I (2018) The Montreal Protocol for Identification of Amusia. *Behav Res Methods* 50:662–672.
- Wang X (2018) Cortical Coding of Auditory Features. *Annu Rev Neurosci* 41:527–552.
- Wang Y, Ding N, Ahmar N, Xiang J, Poeppel D, Simon JZ (2012) Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. *J Neurophysiol* 107:2033–2041.
- Wen H, Liu Z (2016) Separating Fractal and Oscillatory Components in the Power Spectrum of Neurophysiological Signal. *Brain Topogr* 29:13–26.
- Wesolowski BC, Hofmann A (2016) There's more to groove than bass in electronic dance

- music: Why some people won't dance to techno. *PLoS One* 11:e0163938.
- Windsor WL (1993) Dynamic Accents and the Categorical Perception of Metre. *Psychol Music* 21:127–140.
- Winkler I, Háden GP, Ladinig O, Sziller I, Honing H (2009) Newborn infants detect the beat in music. *Proc Natl Acad Sci U S A* 106:2468–2471.
- Witek MAG (2017) Filling In: Syncopation, Pleasure and Distributed Embodiment in Groove. *Music Anal* 36:138–160.
- Witek MAG, Clarke EF, Kringelbach ML, Vuust P (2014a) Effects of polyphonic context, instrumentation, and metrical location on syncopation in music. *Music Percept* 32:201–217.
- Witek MAG, Clarke EF, Wallentin M, Kringelbach ML, Vuust P (2014b) Syncopation, body-movement and pleasure in groove music. *PLoS One* 9:e94446.
- Witek MAG, Liu J, Kuubertzie J, Yankyera AP, Adzei S, Vuust P (2020) A critical cross-cultural study of sensorimotor and groove responses to syncopation among Ghanaian and American university students and staff. *Music Percept* 37:278–297.
- Wojtczak M, Mehta AH, Oxenham AJ (2017) Rhythm judgments reveal a frequency asymmetry in the perception and neural coding of sound synchrony. *Proc Natl Acad Sci* 114:1201–1206.
- Woldorff MG, Hackley SA, Hillyard SA (1991) The Effects of Channel-Selective Attention on the Mismatch Negativity Wave Elicited by Deviant Tones. *Psychophysiology* 28:30–42.
- Wollman I, Arias P, Aucouturier JJ, Morillon B (2020) Neural entrainment to music is sensitive to melodic spectral complexity. *J Neurophysiol* 123:1063–1071.
- Woods DL, Alho K, Algazi A (1992) Intermodal selective attention. II. Effects of attentional load on processing of auditory and visual stimuli in central space. *Electroencephalogr Clin Neurophysiol* 82:341–355.
- Wunderlich JL, Cone-Wesson BK (2001) Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *J Acoust Soc Am* 109:1526–1537.
- Xia Y, Leib A, Whitney D (2018) Serial dependence in the perception of attractiveness. *J Vis* 16:1–8.
- Xiang J, Simon J, Elhilali M (2010) Competing streams at the cocktail party: exploring the mechanisms of attention and temporal integration. *J Neurosci* 30:12084–12093.

- Xu B, Liu-Shuang J, Rossion B, Tanaka J (2017) Individual differences in face identity processing with fast periodic visual stimulation. *J Cogn Neurosci* 29:1368–1377.
- Yc K, Prado L, Ayala YA, Merchant H, Gámez J, Dotov D (2018) Predictive rhythmic tapping to isochronous and tempo changing metronomes in the nonhuman primate. *Ann N Y Acad Sci* 1423:396–414.
- Yerkes BD, Weintraub DM, Snyder JS (2019) Stimulus-based and task-based attention modulate auditory stream segregation context effects. *J Exp Psychol Hum Percept Perform* 45:53–66.
- Yeston M (1976) The Stratification of Musical Rhythm. *J Aesthet Art Crit* 35:155.
- Yi HG, Leonard MK, Chang EF (2019) The Encoding of Speech Sounds in the Superior Temporal Gyrus. *Neuron* 102:1096–1110.
- Yin P, Strait DL, Radtke-Schuller S, Fritz JB, Shamma SA (2020) Dynamics and Hierarchical Encoding of Non-compact Acoustic Categories in Auditory and Frontal Cortex. *Curr Biol* 30:1649-1663.e5.
- Yon D, Gilbert SJ, de Lange FP, Press C (2018) Action sharpens sensory representations of expected outcomes. *Nat Commun* 9:33–35.
- Zalta A, Petkoski S, Morillon B (2020) Natural rhythms of periodic temporal attention. *Nat Commun* 11:1051.
- Zelic G, Varoqui D, Kim J, Davis C (2018) A flexible and accurate method to estimate the mode and stability of spontaneous coordinated behaviors: The index-of-stability (IS) analysis. *Behav Res Methods* 50:182–194.
- Zeni S, Holmes NP (2018) The Effect of a Regular Auditory Context on Perceived Interval Duration. *Front Psychol* 9:1567.
- Zhou H, Melloni L, Poeppel D, Ding N (2016) Interpretations of Frequency Domain Analyses of Neural Entrainment: Periodicity, Fundamental Frequency, and Harmonics. *Front Hum Neurosci* 10:274.
- Zilany MSA, Bruce IC, Carney LH (2014) Updated parameters and expanded simulation options for a model of the auditory periphery. *J Acoust Soc Am* 135:283–286.
- Zoefel B, ten Oever S, Sack AT (2018) The Involvement of Endogenous Neural Oscillations in the Processing of Rhythmic Input: More Than a Regular Repetition of Evoked Neural Responses. *Front Neurosci* 12:95.
- Zuk NJ, Carney LH, Lalor EC (2018) Preferred Tempo and Low-Audio-Frequency Bias Emerge

From Simulated Sub-cortical Processing of Sounds With a Musical Beat. *Front Neurosci* 12:349.