No cover image available

The Oxford Handbook of Music and Corpus Studies

(In Progress)

Daniel Shanahan (ed.) et al.

https://doi.org/10.1093/oxfordhb/9780190945442.001.0001

Published: 2022 **Online ISBN:** 9780190945473 **Print ISBN:** 9780190945442

Search in this book

CHAPTER

Computational Modeling of Rhythm Perception and the Role of Enculturation

Bastiaan van der Weij, Marcus T. Pearce, Henkjan Honing

https://doi.org/10.1093/oxfordhb/9780190945442.013.13

Published: 21 March 2024

Abstract

This chapter compares a variety of computational models of rhythm perception and discusses them in three sections, each focusing on one of various different theoretical perspectives that exist in cognitive modeling, namely cognitivism, embodied cognition, and predictive processing. The different perspectives suggest different computational modeling techniques, which this chapter uses to differentiate models of rhythm perception. Cognitivism most naturally accommodates rule-based models, coupled oscillation models use mathematical tools associated with embodied cognition, and probabilistic generative models are consistent with predictive processing theories of cognition. Each section provides a short description of a theoretical perspective, followed by a discussion of rhythm perception models consistent with that perspective. Furthermore, the chapter draws attention to the influence that Western music theory may have had on models and theories of rhythm perception. This potential influence is of interest because rhythm perception is thought to be shaped by the history of experiences and activities of listeners, enabling the culture in which a listener is embedded to influence their perception. The chapter briefly reviews what effects this influence may have on rhythm perception, suggesting the need for modeling enculturated rather than "universal" listeners. Throughout, the chapter notes that rule-based models do not take previous experiences and activities of listeners into account, while some coupled oscillation models and probabilistic generative models, computational paradigms that gained popularity more recently, do, albeit to varying degrees. Additionally, probabilistic generative models, consistent with predictive processing, suggest a normative explanation of how previous experiences and activities shape perception.

Keywords: rhythm perception, meter perception, computational modeling, enculturation, predictive processing

Subject: Musicology and Music History, Music Theory and Analysis, Music

Series: Oxford Handbooks

Collection: Oxford Handbooks Online

Introduction

In the music cognition literature, a conceptual distinction is often drawn between rhythm and meter. Rhythm refers to a temporal pattern of sounds, while meter refers to a subjective phenomenon (Honing and Bouwer 2019). Listening to rhythms tends to induce a sense of pulsation in listeners (Povel and Essens 1985). This pulsation, known as *beat* or *tactus* (Lerdahl and Jackendoff 1983), provides a temporal reference with which movements can be coordinated (Repp 2005; Repp and Su 2013). We speak of *meter* when some beats appear as more accented than others and these accented beats recur more or less regularly (Cooper and Meyer 1960). Pulse and meter form the basis of temporally coordinated musical activities such as clapping, dancing, singing, or playing an instrument. While these characteristics of meter are generally regarded as uncontroversial among music cognition scholars, two aspects that elude consensus are the precise nature of the cognitive phenomenon known as meter and the degree to which it is shaped by a listener's history of prior musical experiences and activities.

Regarding the nature of the cognitive mechanisms, multiple approaches have been proposed. Some of these highlight abstract hierarchical structures (Longuet-Higgins 1978; Longuet-Higgins and Lee 1984; Lerdahl and Jackendoff 1983), others entrainment of attention (Jones and Boltz 1989; Large and Jones 1999), neural resonance (Large and Snyder 2009), or embodied and ecological aspects of rhythm perception (Shove and Repp 1995; Iyer 1998; Clarke 1987; Todd and Lee 2015). More recently, approaches based on predictive processing have been proposed (Vuust and Witek, 2014; Van der Weij, Pearce, and Honing 2017).

Computational cognitive models of rhythm and meter perception (for brevity, we refer to such models collectively as rhythm perception models) are the focus of this chapter. By computational models, we mean models that are described—ideally in a formal language—with a level of precision allowing them to be implemented as a computer program, without the need to fill in many details (see also Temperley 2013). Such models may be distinguished from verbal-conceptual models, which are expressed in prose or as conceptual diagrams, and may be consistent with multiple computational models.

We discuss rhythm perception models in the context of three broad theoretical perspectives, namely cognitivism (cf. Anderson 2003), embodied cognition (Brooks 1991; Van Gelder 1995; Anderson 2003; Chemero 2009), and predictive processing (Clark 2013). Each of the above approaches can be associated with one of these perspectives. In turn, these perspectives can be associated with different computational modeling principles that underlie the models discussed in this chapter.

Briefly, cognitivism views cognition as being primarily involved in rule-governed information processing. This perspective is associated strongly with classical artificial intelligence approaches (e.g., see Newell and Simon 1976). Among rhythm perception models, classic rule-based models (e.g., Longuet-Higgins and Steedman 1971; Longuet-Higgins and Lee 1982) and preference-rule models (Temperley and Sleator 1999; Temperley 2001) may be associated with this perspective (see the section "The Cognitive Perspective"). Embodied cognition may be characterized by a rejection of the idea that information processing and abstract representation provide the most appropriate explanation of many behaviors. It instead emphasizes the role of continuous dynamic interaction between brain, body, and environment. Many characteristics of adaptive oscillator (e.g., McAuley 1995; Large and Palmer 2002), and neural resonance (e.g., Large, Herrera, and Velasco 2015) models harmonize well with this perspective (see "The Embodied Perspective"). Finally, the term predictive processing (introduced by Clark 2013) covers a class of theories that build on the Bayesian brain hypothesis (Knill and Pouget 2004) and predictive coding (Rao and Ballard 1999). These theories propose that perception and cognition can be understood as prediction-error minimization in a probabilistic generative model. Probabilistic generative approaches to rhythm perception (Temperley, 2007; Van der Weij, Pearce, and Honing 2017) are consistent with this perspective (see "The Predictive Processing Perspective").

The question of the degree to which rhythm perception of individual listeners is shaped by their history of prior experiences and activities is often considered in the context of cultural background. Cultural background is one predictor of stable tendencies in histories of prior musical experiences and activities of individual listeners. If these stable tendencies have the power to influence rhythm perception, cultural background may predict certain individual characteristics of rhythm perception in listeners from different cultural backgrounds. We use the term *enculturation* to refer to the acquisition of implicit cultural knowledge by exposure to, and participation in, cultural activities. The predictive processing perspective most prominently draws attention to the role that enculturation might play in the shaping of perception and cognition. Although neither cognitivism nor embodied cognition are explicitly incompatible with this role, predictive processing accounts for it normatively, namely as a consequence of a domain-independent prediction-error minimization mechanism.

While there is considerable evidence, some of which is discussed in the next chapter, suggesting that enculturation shapes rhythm perception, enculturation plays little to no role in the majority of existing rhythm perception models. Some of these models take inspiration from Western music theory and have been evaluated only on Western tonal music. A similar lack of diversity can be observed in the stimulus materials and participants used in empirical and experimental music cognition research (Huron 2008; Jacoby et al. 2020). If rhythm perception is indeed shaped by enculturation, studying it predominantly in the context of Western music is problematic since the resulting knowledge may depend on familiarity with Western music.

Among models that do not account for effects of enculturation, some explicitly limit their scope to Western tonal music (e.g., Longuet-Higgins 1979). Others aim to reflect universal constraints on perception and cognition (Povel and Essens 1985; Large 2010b). Parameters of such models are typically determined by musical intuition (e.g., Longuet-Higgins 1976; Povel and Essens 1985; Temperley and Sleator 1999), or by optimal fit to experimental data (e.g., Shmulevich and Povel 2000). On the other hand, some models aim to simulate the effects of prior *exposure* (one aspect of enculturation) to certain kinds of music on rhythm perception (Van der Weij et al. 2017; Tichko and Large 2019). Parameters of these models are derived from empirical samples of rhythms that are intended to represent previous exposure to rhythms (see also Patel and Demorest 2013; Pearce 2018; Morrison, Demorest, and Pearce 2019). Some of these models are probabilistic generative models, which are consistent with the mechanisms posited by the predictive processing perspective.

In summary, this chapter discusses computational cognitive models of rhythm perception and aligns them with three broad theoretical perspectives on cognition. It furthermore considers the role that enculturation may play in rhythm and meter perception and the degree to which models take this role into account. The next section reviews research related to the role of enculturation in shaping rhythm perception. The remaining sections of this chapter are dedicated to each of the three broad theoretical perspectives on cognition: cognitivism, embodied cognition, and predictive processing. Each section opens with a brief discussion of the theoretical perspective before turning to the rhythm perception models that are consistent with it.

Finally, we note that evaluating the performance of the discussed models and the connection between model predictions and empirical observations receives less attention in this chapter than the reader might have expected. This is partly because the emphasis lies on theoretical differences between cognitive models, and partly because extensive comparisons between computational models, especially on culturally diverse datasets, are simply not available (but for an exception, see Desain and Honing 1999).

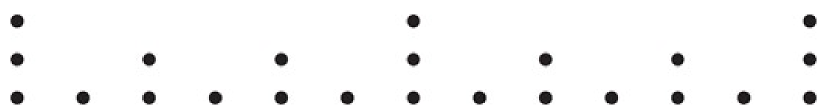
Cross-Cultural Perspective on Rhythm Perception

Perhaps in part due to its high level of pervasiveness,[3] Western tonal music has, explicitly or implicitly, played a significant role in the formation of theories and models in music cognition (see also Jacoby et al. 2020). However, this musical tradition represents only a small slice of the variety in musical cultures that exists around the world (Trehub et al. 2015; Savage et al. 2015; Mehr et al. 2019). If rhythm perception is shaped by enculturation, it can be expected to vary for individual listeners depending on the kind of music they are familiar with. Below we discuss studies that suggest that this is the case. The discussion considers three aspects of rhythm perception: the relation between metrical hierarchies and the likelihood of events, constraints for tactus beats to be isochronous, and the shape of perceptual categories for temporal intervals.

Metrical Hierarchies and Event Likelihood

Metrical hierarchy plays a prominent role in theories of rhythm perception. The way such hierarchies are commonly conceptualized can be attributed in part to the influential work of Lerdahl and Jackendoff (1983), Longuet-Higgins (1978), and Longuet-Higgins and Lee (1984). These authors describe meter as a hierarchy of *metrical levels*, popularly depicted as *metrical grids* by Lerdahl and Jackendoff (see Figure 1). Each metrical level consists of *isochronous* (spaced evenly in time) *beats*. Beats are described as duration-less points in time, represented abstractly in the mind of the listener. The resulting representation imposes a pattern of alternating strong and weak beats onto a perceived rhythm, where the metrical strength (sometimes called metrical accent, or metrical salience) of a beat is determined by the highest metrical level in which it occurs. Lerdahl and Jackendoff (1983) distinguish between *phenomenal accents*, which are due to the way a piece of music is performed, and *metrical accents*, which are due to the metrical interpretation of the music by a listener and occur on metrically strong beats. This distinction highlights the conceptual difference between rhythm and its metrical interpretation by a listener.

Figure 1:



A metrical grid visualizing the putative hierarchical organization of two bars of a ternary time signature (such as 3/4 time). The hierarchy contains three metrical levels. The dots represent beats, the horizontal dimension represents time (which flows from left to right), and the vertical dimension represents metrical salience (towards the top of the figure is more salient). The top level usually indicates the bar-level periodicity. Note that between each pair of dots on a higher level, two or three dots occur at a lower level. For each top-level beat, there are three middle-level beats, and for each middle-level beat there are two beats on the lowest level, indicating ternary subdivision of the top level and binary subdivision of the middle level. Also, note that beats at each level are equidistant in time.

It is commonly assumed in computational and verbal-conceptual theories of rhythm perception that the metrical strength of a beat represents the strength of prediction or expectation that an event will occur (Temperley 2007; Large 2008). The metrical phenomena of "loud rests" (London 1993) and syncopation are commonly related to this assumption. A *loud rest* occurs when an event unexpectedly does not occur at a metrically salient beat. *Syncopation* occurs when a metrically salient beat passes silently or unaccented and is preceded by an onset or accent at a metrically weaker beat (Longuet-Higgins and Lee 1984). These phenomena are generally described as deviations from the norm, or as violations of expectation (Fitch and Rosenfeld 2007; Bouwer et al. 2018). Consequently, measures of syncopation are sometimes used to estimate the perceptual complexity of rhythms based on the idea that rhythms whose constituent events have unpredictable timing will be experienced as more complex (e.g., see Witek et al. 2014). However, recent studies have suggested that the presumed correlation between metrical strength and degree of event expectation may not apply to all listeners.

Palmer and Krumhansl (1990) hypothesized that the “frequency with which musical events in a piece occur in a given metrical context may provide important perceptual cues to meter.” Using a set of Western classical music compositions by four different composers, Palmer and Krumhansl constructed *event-frequency distributions* based on the relative frequency of onsets at different positions in a bar. Such distributions were constructed separately for different meters and composers. In support of their hypothesis, Palmer and Krumhansl found that metrical salience more or less predicts the relative frequency of events and that this effect is stable for different composers. Palmer and Krumhansl furthermore conducted a pair of behavioral experiments, the results of which indicated that expectations of listeners (especially if they are musicians) for notes to occur in different metrical contexts correlated with metrical salience of those contexts.

While Palmer and Krumhansl highlight the role of statistical regularities in music, they interpret this role in the context of multileveled representations of metrical hierarchies. They suggest that observed frequency distributions of musical events in different metrical contexts result from the presence of such metrical hierarchies in the minds of composers and listeners, rather from stylistic constraints in the music. However, Palmer and Krumhansl qualify this finding by noting that their observations are limited to Western classical music. Indeed, subsequent corpus studies applying the same methodology to rhythms from different musical idioms suggest a more prominent role for stylistic constraints in the shaping of frequency distributions of event timing.

Holzapfel (2015) analyzed event-frequency distributions derived from a corpus of Turkish makam music (Karaosmanoğlu 2012). Turkish makam music is a style of both classical and folk music in which rhythmic organization is centered around the notion of an *usul* (Marcus 2001). Usuls are rhythmic modes, characterized by a pattern of drum strokes. Holzapfel derived these distributions for Turkish makam music by collapsing over usul cycles, which are annotated in the corpus. The results show that, compared to Western music, onsets in Turkish makam music were more spread out over different positions in the metrical cycle and that usul patterns could be used to classify the usul underlying makam compositions.

London, Polak, and Jacoby (2017) examined a set of Malian djembe ensemble recordings using event-frequency distributions. This music does not make use of music notation, so London and colleagues relied on onset annotations of the recordings. After correcting for tempo changes, the observed onsets were collapsed over metrical cycles. Results show that the relative frequencies of events at different positions in the metrical cycle do not appear to be structured by metrical salience patterns, even though the rhythms are metrically structured and the consistent timing of subdivisons suggests the presence of a regular beat.

The above findings are consistent with the idea that the distribution of events over positions in the metrical cycle provides a cue for meter. However, metrical hierarchy, as predicted by theories based on Western music theory (Longuet-Higgins 1978; Longuet-Higgins and Lee 1984; Lerdahl and Jackendoff 1983), appears not to be the only predictor of those patterns. Based on their findings in Malian djembe ensemble recordings, London, Polak, and Jacoby (2017) claim that “the shared presumption that onset frequency is correlated with metrical accent holds only contingently, that is, for the corpora of Western classical and popular music that were used in these studies, and for which these models were developed” (p. 478). This also calls into question the view that syncopations necessarily reflect violations of expectation. Iyer (1998) anticipated this, suggesting that one “should not regard the global musical preponderance of “syncopation” (off-beat accents) as a vast set of exceptions to the “normal” accentual rules of meter, but rather as convincing counterexamples to such proposed accentual rules. (p. 44).

Meters with Non-isochronous Tactus Beats

Theories of metrical structure often include constraints for the beats at the tactus level to be evenly spaced (isochronous). Longuet-Higgins (1978) and Longuet-Higgins and Lee (1984) described meter generatively as the recursive subdivision of intervals into two or three evenly spaced beats. Similarly, Lerdahl and Jackendoff (1983) suggested that beats in well-formed metrical hierarchies must be more or less evenly spaced. The authors cited here have indicated that their theories apply primarily to Western music, but their ideas have nevertheless shaped subsequent research, which does not always acknowledge this qualification.

Meters with uneven (non-isochronous) intervals between tactus beats, while relatively uncommon in Western classical music, are prevalent in many musical styles (e.g., see London 1995; Polak et al. 2018). Cross-cultural studies have suggested that these meters are readily processed by listeners familiar with these structures. London (1995) calls such non-isochronous meters “complex,” and argues that a non-isochronous tactus beat needs to be anchored in a faster, and isochronous, underlying pulse, such that tactus beats are measured by either two or three of these faster pulses. This suggestion has been challenged by observations that non-isochronous tactus beats do not always adhere to an underlying isochronous grid (Kvifte 2007). Furthermore, it has been suggested that the purported complexity of non-isochronous meters is overridden by familiarity: adults and infants with exposure to non-isochronous meters can detect violations of the meter while adults with limited exposure to such meters can only do this for isochronous meters (Hannon and Trehub 2005b; Soley and Hannon 2010; Hannon et al. 2012). For such listeners, rhythms in a non-isochronous meter are no more complex than those in an isochronous meter.

Perceptual Categories for Temporal Duration Ratios

It is thought that ratios between continuous time intervals in rhythms are perceived as a small number of discrete perceptual categories (Clarke 1987; Desain and Honing 2003). These categories appear to be centered around small-integer ratios (such as 1:2:1 and 1:2:3), but their size and shape vary, resulting in perceptual biases (Desain and Honing 2003; Jacoby and McDermott 2017). It has been hypothesized that small-integer ratios are a universal constraint on perceptual categories for duration intervals (Mehr et al. 2019). In support of this hypothesis, Mehr et al. found that simple-integer duration ratios are prevalent in a culturally diverse sample of music recordings. Furthermore, Savage et al. (2015) found evidence for the widespread occurrence of binary and ternary subdivision as well as isochronous beats across a range of musical cultures worldwide.

Jacoby and McDermott (2017) found evidence that biases in categorical perception of temporal intervals—that is, the size and shape of perceptual categories—may be attributable in part to enculturation. In a cross-cultural study involving adult members of a native Amazonian society (the Tsimané) and North American adults, they found that the size and shape of perceptual categories for temporal intervals differed significantly between these two groups, but also that, in both groups, perceptual categories appeared to be centered around small-integer ratios. Since the musical practices of the Tsimané and North American participants are different enough to cause the observed differences in perceptual biases, this commonality is especially remarkable and consistent with the potential universality of small-integer ratio categories for ratios between temporal intervals in rhythms.

Polak, London, and Jacoby (2016) and Polak et al. (2018), however, present work that challenges the hypothesized universality of (perceptual categories for) small-integer duration ratios in rhythms. Malian djembe ensemble performances commonly contain “swung” subdivisions—intervals subdivided into intervals related by a complex ratio. Polak and colleagues show that these non-isochronous subdivisions afford the production of precise and consistent timing patterns in an ensemble context. This seems to suggest that these Malian musicians can entrain to complex-ratio beat subdivisions. Polak et al. (2018)

suggest that the production of such non-isochronous subdivisions may be supported by non-isochronous perceptual categories that depend on experience and training and, in a cross-cultural study, found evidence for the presence of such a category in expert musicians from Mali, but not in expert musicians from Germany or Bulgaria.

Musical features that are surprisingly prevalent across a range of musical cultures worldwide, known as statistical universals (Savage et al. 2015), are sometimes seen as evidence for the presence of innate cognitive constraints on music perception and production. While the presence of such constraints may plausibly give rise to statistical universals, there seem to be other ways in which such universals could arise. For example, individuals from different cultures share more than genes. Bodily constraints and stable properties of natural environments that are independent of geographic location might also give rise to universal tendencies in music. Furthermore, it has been argued based on simulations of the dynamics of cultural transmission using Bayesian models that statistical universals may emerge readily from weak and defeasible cognitive biases (Thompson et al. 2016).

Reconciling Enculturation with Universal Tendencies in Rhythm Perception

Some authors, such as Temperley (2000) and Agawu (1995), warn that there exists a tendency, primarily in the ethnomusicological literature, to overstate differences in rhythmic practices and rhythm perception between cultures. Others, such as Iyer (1998), Huron (2008), and, recently, Jacoby et al. (2020) lament the sparsity of cross-cultural work in music cognition and caution against interpreting the idiosyncrasies of a familiar musical (usually Western) culture as the norm, or of culture-specific perceptual constraints as universal.

A theoretical account of rhythm perception that can potentially reconcile these views has been proposed by London (2004). In accord with ideas of Jones and Boltz (1989) and Large and Jones (1999), London suggests that meter perception is a form of entrainment behavior, which serves to guide our attention over time in synchrony with musical rhythm. However, its openness to shaping—by experience, practice, music education, and other forces of influence that an individual's embeddedness in a cultural environment entails—makes metrical entrainment a *skilled behavior*. Thus, meter perception is regarded more than a passive response to music or a bottom-up analysis of sensations. However, London also argues that although the structure of metrical entrainment behavior is plastic, it simultaneously is constrained by a set of well-formedness conditions which he sets forth (echoing Lerdahl and Jackendoff's [1983] approach). London explicitly avoids proposing universal preference rules since, he argues, the relation between rhythm and meter is malleable and ambiguous.

This theoretical account thus argues that certain aspects of rhythm perception can be shaped by enculturation while other aspects are less adaptable and can be captured by well-formedness constraints; it therefore occupies a middle-ground between work emphasizing cultural differences in rhythmic practices and work emphasizing universal constraints on perception. Specifically, constraints on metrical entrainment behavior are argued to arise universally, while an individual's capacity for metrical entrainment depends on their previous experiences and activities and is therefore subject to the influence of enculturation.

Enculturation and Embodiment

A more radical reading of some of the above literature suggests that the enculturation of rhythm perception may involve information that cannot be gleaned from music corpora alone. London, Polak, and Jacoby (2017) suggest that

> [...] while the frequency of onset occurrence of events doubtless plays a role in our acquisition of rhythmic and metrical knowledge, those frequencies occur in holistic contexts that include timing, timbre, and other auditory, visual, and sensorimotor channels of perception. Combinations of these cues forge associations between statistically common rhythms and their characteristic metrical orientations. (p. 479)

Some information about such holistic contexts may be encoded in music corpora, but these annotations provide no substitute for tightly coupled sensing and acting involved in participation in music-related cultural practices such as dancing, attending a concert, or singing in a group. These experiences involve coordinated movements and sensations in, most prominently, the auditory, visual and proprioceptive modalities.

An embodied view of rhythm perception acknowledges the role that these aspects might play in musical experiences. An action-oriented interpretation of predictive processing (Clark 2013), however, in addition to emphasizing the role of embodiment in perception, suggests how embodied experience *shapes* perception (see also Clark 2016). A theory of rhythm perception based on action-oriented predictive processing might therefore be consistent with the theoretical account described by London (2012), in that it may describe the role of practice and training in shaping rhythm perception.

The above considerations, if true, appear troublesome for computational models of rhythm perception that aim to simulate enculturation using samples from music corpora. Nevertheless, stable probabilistic properties of the music to which enculturated individuals are exposed may still play a significant role in shaping rhythm perception. Samples drawn from music corpora are likely to reflect these probabilistic properties. Models that use these samples to simulate the effects of enculturation on rhythm perception may therefore successfully capture some of these effects.

To conclude, the research reviewed above suggests that the experience of meter depends to a large extent on being situated in a cultural environment. If so, it seems that rhythm perception models that aim to accommodate music from different cultures cannot rely on a bottom-up analysis of a rhythm based on hypothetically universal mechanisms for rhythm perception. They must also account, in some way, for the effects of being intimately familiar with certain musical idioms. For computational modeling, empirical samples from music corpora may go some way toward simulating the musical exposure of enculturated individuals. However, such corpora do not capture the holistic context in which exposure to music occurs, possibly leaving some aspects of enculturated rhythm perception unaddressed.

The Cognitivist Perspective

We now turn to rhythm perception models associated with the first of the three different broad theoretical perspectives discussed in this chapter, namely cognitivism. The cognitivist perspective is characterized by the view that cognition is most appropriately explained as pure *information processing*, involving rule-based computation performed on symbolic representations. These representations are derived from sensory input through bottom-up perceptual processes (e.g., Newell and Simon, 1976; Marr, 1982, cf. Anderson, 2003). This emphasis on information processing is typically reflected in the terminology used to motivate and describe cognitivist models. Perceptual and cognitive phenomena are described as involving “problems” or “tasks” that cognition must “solve” or “decide.” Modeling a cognitive process entails identifying the task it performs, identifying the appropriate representations of input and output, and finding an algorithm that generates the appropriate output given an input.

One motivation for the epistemological value of such models is that designing an algorithm to solve a specific cognitive task may provide insight into the cognitive process itself. Commonly, the process of designing such algorithms reveals unanticipated intricacies and complexities of the task itself that were overlooked by verbal-conceptual theories. The cognitivist approach was especially popular in the early days of cognitive science, and its methodology was significantly influenced by contemporary developments in artificial intelligence. These influences have been noted by authors like Longuet-Higgins (1978), Newell and Simon (1976), and Bundy (1990).

Rule-Based Models of Rhythm Perception

The sections below describe a set of cognitivist rhythm perception models proposed by Longuet-Higgins and colleagues, who pioneered computational modeling of music cognition in the 1970s and 1980s. These models propose specific mechanisms for various computational problems that are hypothesized to be involved in rhythm perception. Longuet-Higgins and colleagues point out at several occasions (Longuet-Higgins and Steedman 1971; Longuet-Higgins 1978, 1979) that their work aims to account for perception of Western tonal music by listeners familiar with such music. Therefore, the models described below are not intended as universal accounts of rhythm perception, but rather as reflections of the perception of enculturated listeners. Clarke (1999) more extensively discusses these models and many that followed in this early period of music cognition modeling. Some of these models are still actively used in empirical studies (e.g., Fitch and Rosenfeld 2007; Grahn and Brett 2007; Song et al. 2013; Witek et al. 2014; Bouwer et al. 2018).

Longuet-Higgins and colleagues described several key issues that still inspire modelers of music perception to this day. The central issue is to understand listeners' ability to reconstruct the rhythmic and tonal relations, intended by the composer, between sounds from a performance of Western classical music. Western tonal music notation contains considerable information about the tonal and temporal relations in music. Trained musicians can reconstruct this information from "even a mediocre performance" (Longuet-Higgins and Steedman 1971, p. 221). Therefore, it is argued, scores are likely to provide strong clues toward the kind of rhythmic and tonal relations that listeners infer from a performance. This ability is furthermore argued to be available to anyone "familiar with the composer's language" (Longuet-Higgins 1978, p. 149). In the models described below, the inference of rhythmic relations (meter) and tonal relations (key) is treated independently. The discussion below considers only the parts of these models relevant to the interpretation of rhythm.

Longuet-Higgins and colleagues divided the central issue into a set of sub-problems, which were addressed individually by the computational models that we describe below. These models describe the inference problem from the perspective of the listener who is processing a piece of music note by note. This listener must, from the first few notes, infer the phase and period of the beat. This problem is addressed by Longuet-Higgins and Lee (1982). Then, the established beat must be subdivided recursively until each note initiates a metrical unit at some level of beat subdivision. Sometimes, notes are played slightly earlier or later than expected. In these cases, the listener must figure out whether these deviations represent a change in tempo, a subdivision of the beat, or expressive timing of the performer. Tracking and subdivision of a beat are addressed by Longuet-Higgins (1976). Finally, to find the correct time signature, beats must be grouped into higher-level metrical units such as bars. This problem is addressed by Longuet-Higgins and Steedman (1971). The sections below discuss these models in chronological order.

Grouping Metrical Units

Longuet-Higgins and Steedman (1971) propose an algorithm that addresses how, based on a pattern of note durations in a deadpan performance, a listener may identify metrical units and group them into bars.[4] Motivated by "the progressive character of musical comprehension" (p. 223), Longuet-Higgins and Steedman propose a fundamental principle, which they call the *rule of congruence*, by which the other rules of the model are motivated. The intuitive motivation for this principle is described eloquently by the observation that "music would be a dull affair if all notes had to be in the key and all accents on the beat, but it would be incomprehensible if the key and meter were called into question before they were established" (p. 224). The rule of congruence stresses the important role played by the temporal order of musical events. This emphasis on the *temporal incrementality* of music listening sets this early approach apart from later approaches, which ignore the temporal order of events (e.g., Povel and Essens 1985; Palmer and Krumhansl 1990).

The rules of the model contain many subtleties but can be summarized approximately as follows: The duration of the first or second note (whichever is shorter) establishes the smallest *metrical unit*. By the rule of congruence, an established metrical unit is never abandoned. The metrical hierarchy is progressively constructed from this smallest unit by means of grouping. Such grouping is prompted by one of three cues: (1) the occurrence of a long note beginning on an already established metrical unit; (2) a *dactyl*, a pattern consisting of two long followed by one short interval; or (3) a long note followed by a short note. If any of these is encountered, the current metrical unit is multiplied in length by two or three (depending on the length of the cue-pattern) to form a metrical unit at the next level of the hierarchy.

Beat Tracking and Subdivision

Longuet-Higgins (1976) proposes an algorithm addressing a different issue: How to track and subdivide a beat in a performance with a changing tempo? While Longuet-Higgins and Steedman's (1971) model and its extension (Steedman 1977) assume deadpan performances, listeners can follow along with a beat despite tempo changes and expressive timing. Another way of stating the problem that this work aims to address is as follows: When an onset occurs close to where a beat is expected, how does a listener decide whether the onset marks a subdivision of the beat, a change of tempo, or an expressive deviation?

Longuet-Higgins proposes the following procedure:[5] Assuming a given beat interval, it can be determined where, assuming deadpan timing, the next beat is expected. Based on this, the amount of time by which the next note deviates from this expected beat can be determined. A temporal window around the expected next beat location is created by a parameter called *tolerance*. If the next note falls within the tolerance window, the next beat interval is increased or decreased by half the amount of deviation of the note from deadpan timing. If the next note instead occurs before the tolerance window, the beat interval is subdivided by two or three. The upshot is that, once processing is complete, each note occurs at the beginning of a metrical unit.

This mechanism for adapting beat duration based on deviation from deadpan timing bears some resemblance to beat perception models based on coupled oscillation (Large and Kolen 1994) that are discussed in the next section. The resemblance is notable because coupled oscillation models align with a theoretical perspective that is rather different from cognitivism.

Beat Finding

Longuet-Higgins and Lee (1982) address another puzzle: Assuming that a beat can be tracked and subdivided once established, how is the tactus beat established to begin with? How do we know whether the rhythm begins with an anacrusis? If it does, where does the first beat occur? How do we know the interval between the first and the second beat? Borrowing terminology used by Desain and Honing (1999), Longuet-Higgins and Lee's model maintains a *current beat hypothesis*. This hypothesis is specified by two variables representing virtual points in time: a "first beat", t_1, and a "second beat", t_2. These variables are initialized by the onset times of the first and second note. Subsequent notes revise and update the current beat hypothesis by subjecting it to two types of transformations: lengthening (stretching) it or shifting its position. These transformations are triggered by a set of rules that, in the interest of brevity, we will not attempt to summarize here. The algorithm aims to output values of t_1 and t_2 that encode the position and duration of the first tactus beat interval in a performance, thereby providing an answer to the questions at the beginning of this paragraph.

Desain and Honing (1999) note that this model and some of its successors were evaluated only qualitatively on small toy domains. Furthermore, it is difficult to derive general properties of their functioning since the rules in these models interact with input in complex and unpredictable ways. Desain and Honing propose a unified framework in which the models can be expressed, and they systematically analyze the behavior of the models using an empirical dataset and Monte-Carlo samples from the models' *input space*: the set of all possible input rhythms of up to thirty-five grid points. Desain and Honing's results show that these simple models perform surprisingly well. Their work represents one of the few existing systematic comparisons between computational models on the same dataset.

Optimization and Preference-Rule Models

We describe optimization models here as cognitivist models but differ significantly from the rule-based models described above. Instead of what are sometimes called "hard and fast" rules, optimization models employ soft constraints that can be satisfied to various degrees. In some ways, which we return to below, these models are similar to probabilistic generative models (described in "The Predictive Processing Perspective"). Unlike these models, however, optimization models provide no theoretically motivated interpretation of the metric that is optimized.

The well-known clock model of Povel and Essens (1985) is an optimization model that also contains some rule-based aspects. This model operates by generating a combinatorically exhaustive set of "clocks," defined by a unit (period) and location (the phase of the first event in the rhythm relative to the clock's period), calculating a score for each clock given a rhythm. Input rhythms are first preprocessed to mark events that, according to a set of rules, are predicted to be perceived as accented. The score that is calculated for each clock is based on how well the clock's ticks align with events marked as accented. The model selects the clock that optimizes this score, and its corresponding score is used to predict the degree to which the rhythm induces the clock.

Temperley and Sleator (1999) introduce another optimization approach, which they call a preference-rule model. Preference rule models are intended to be a computational implementation of the system of preference rules proposed by Lerdahl and Jackendoff (1983). Lerdahl and Jackendoff's ideas were influential but lacking in formal rigor (see Hansen 2010, for an extensive discussion), and preference-rule models are an attempt to address this. Temperley and Sleator propose independent models for meter and harmony. As before, our discussion considers only the meter model.

Preference rule models operate by generating an exhaustive set of analyses of a rhythm, specified by a set of well-formedness constraints. Each of these analyses receives a score based on a set of preference rules.

Given an analysis and a piece of music, a preference rule yields a score representing the degree to which the analysis is preferred for the piece of music. The total score of the analysis is calculated as a weighted linear combination of the scores of the individual preference rules. The analysis with the highest total score is the analysis that the model predicts to be correct.

For their meter model, Temperley and Sleator formalize three preference rules: the regularity rule, which prefers analyses in which beats are equally spaced; the event rule, which prefers analyses in which beats are aligned with events; and the length rule, which prefers analyses that align strong beats with the onsets of longer durations. Well-formed metrical hierarchies are constrained to contain exactly five metrical levels. Generalizing from Lerdahl and Jackendoff (1983), who based their theory primarily on music as notated in scores, Temperley and Sleator allow beats to be irregularly spaced. Analyses in which beats are regularly spaced are nevertheless preferred by the regularity rule.

The regularity rule is the only preference rule that depends only on the analysis and not on its relation to a piece of music. This type of rule resembles the concept of a *prior probability* in probabilistic generative models. This probability represents the *a priori* probability of an analysis which is independent of the rhythm that is analyzed. Preference-rule as well as generative approaches entail a trade-off between the *a priori* preferability of an analysis and its congruence with a piece of music: the more unlikely an analysis is *a priori*, the more strongly it needs to be supported by the piece of music.

Preference rule models have some advantages compared to rule-based models. Because preference rules represent soft constraints, they naturally allow for a certain degree of deviation from the norm: decreased congruence in one aspect (e.g., the degree to which beats are spaced evenly) may be compensated for by increased congruence in another (e.g., alignment of strong beats with notes). Furthermore, some rule-based approaches have been criticized for being opaque: it is difficult to describe regularities in their behavior based on the formulation of their rules because the rules interact in complex ways (Desain and Honing 1999). Preference-rule models, by contrast, have the benefit of being easy to interpret because preference rules represent aspects of the relation between music and interpretation that are to be preferred.

A limitation of the optimization model of Povel and Essens (1985), but not necessarily of preference rule models (see Temperley 2001, ch. 8), is that it does not consider the dynamic interplay between the unfolding music and the listener's perception and expectations. It is commonly emphasized (Longuet-Higgins and Steedman 1971; Longuet-Higgins 1978; Lee 1991; Large and Kolen 1994) that this interplay should be central to any account of rhythm perception.

The Embodied Perspective

The term "embodied cognition" carries a variety of connotations (Wilson and Golonka 2013). Here we interpret it as emphasizing continuous dynamic interaction between brain, body, and environment, from which various behavioral and cognitive phenomena are *emergent* (Brooks 1991; Van Gelder 1995; Chemero 2009). This poses a contrast with the emphasis that cognitivist approaches place on strict information processing, which downplays the role of an agent's physical interaction with its environment. The emphasis on dynamic interaction is reflected in the type of models typically associated with this perspective, namely dynamical systems models (Chemero 2009).

There is a class of cognitive models of rhythm perception that proposes that pulse and meter perception is based on *coupled oscillation* (Large and Kolen 1994; McAuley 1995). These models posit that constraints on meter and how it is induced emerge jointly from the dynamics of coupled oscillation. While cognitivist approaches describe meter perception as a cognitive mechanism that infers an abstract representation (meter), often in a bottom-up fashion, from perceptual input (rhythm), coupled oscillation models pose no sharp distinction between representations and cognitive mechanisms that infer representations from sensory input.

Although coupled oscillation models do not necessarily emphasize a role for the body and environment in rhythm perception, they are compatible with two central tenets of embodied cognition: a rejection or downplay of the importance of cognitive representations (Anderson 2003; Wilson and Golonka 2013) and a rejection of the idea that cognition is most appropriately described in terms of computation and symbol manipulation (see Van Gelder 1995). This sentiment is reflected strongly in the fragment below, which appears in an introduction to neural resonance models (a type of coupled oscillation models) of music perception (Large 2010b).

> The brain does not "solve" problems of missing fundamentals, it does not "compute" keys of melodic sequences, and it does not "infer" meters of rhythmic input. Rather, it *resonates* to music. (p. 201; emphasis in original)

Coupled oscillation models can account for a remarkable number of phenomena in rhythm perception without resorting to domain-specific constraints.[6] Among these phenomena are aspects that are argued to pose challenges for other approaches, namely tracking a beat in rhythms with tempo changes (Large and Jones 1999) or expressive timing (Large and Palmer 2002) and entraining to syncopated rhythms in which the pulse frequency is absent from the Fourier spectrum of the rhythm (Velasco and Large 2011). Coupled oscillation models have therefore emerged as popular models of rhythm perception. Below, two types of coupled oscillation models are discussed: *adaptive oscillator models* and *neural resonance models*.

Adaptive Oscillator Models

McAuley (1994) proposed the term *adaptive oscillator* for a class of oscillators that adapt their period in response to exogenous rhythms. McAuley (1994, 1995) and Large and Kolen (1994) independently (McAuley 1995, p. 67) proposed oscillators of this type as models for rhythm perception. McAuley (1995) described the theoretical status of these models as somewhere in between a "single-neuron model and that of a psychological theory." Large and Kolen (1994) describe their model as representing "a single abstract processing unit, amenable to connectionist implementation." Thus, both proposals describe these models as abstract, rather than mechanistic, accounts of rhythm perception (in contrast to neural resonance models).

Different aspects of the behavior of coupled oscillators may be connected to different aspects of rhythm perception. Large and Kolen (1994) describe an oscillator that synchronizes with a periodic component of a rhythmic pattern as "embodying the notion of musical pulse, or beat." Similarly, McAuley (1995, p. 12) writes that oscillators model "global dynamics of perceptual mechanisms involved in the processing of rhythmic patterns." Metrical hierarchy is proposed to emerge from two or more endogenous oscillators entraining to each other, as well as to an exogenous rhythm (Large and Kolen 1994; McAuley 1995). McAuley (1995) also raises dimensionality reduction, or efficient memory encoding, as a motivation for the approach: an oscillator may be seen as an efficient memory representation of where the pulse is. Finally, it is often stated that the oscillators encode a *prediction* or *expectation* of when events are expected.

An oscillator produces periodic behavior that is described by two state variables, period, p, and phase, ϕ. Period represents the amount of time required for an oscillator to complete its cycle. Phase represents the relative position of the oscillator within its cycle and evolves from zero to one, at which point it wraps back to zero. An oscillator is sometimes said to "fire" when its phase reaches zero.

The paragraphs below describe a general mathematical framework for adaptive oscillator models. This framework is limited to describing how one oscillator is influenced by another. Extensions to two endogenous oscillators that entrain to different periodicities in a rhythm are described by Large and Jones (1999) and Large and Palmer (2002). The oscillator that is being influenced is the *endogenous* oscillator: a source of endogenous oscillations. The other is an *exogenous* "oscillator," which, in adaptive oscillator models, is not an actual oscillator, but an exogenous rhythm. Causation is unidirectional: the exogenous oscillator (the rhythm) exerts influence on the endogenous oscillator, but the endogenous oscillator does not influence the behavior of the exogenous oscillator. This influence is called *coupling* and causes the phase and period of the endogenous oscillator to be perturbed by the activity of the exogenous oscillator.

Adaptive oscillator models can be evaluated in a sequence of discrete time steps. Because the phase and period are only perturbed by the firing of the exogenous oscillator (or the presence of an onset, since the exogenous oscillator is a rhythm), the dynamic behavior of the system can be described entirely by considering only the *relative phase* of the two oscillators at moments when the exogenous oscillator fires. The relative phase is the (circular) difference between the endogenous and exogenous oscillator's phase. The relative phase at the $(n+1)$ th firing of the exogenous oscillator, given the relative phase at the n th firing and the periods of the endogenous and exogenous oscillators, is described by

$$\phi_{n+1} = \left(\phi_n + \tfrac{q}{p}\right) \mod 1, \tag{1}$$

where q represents the period of the exogenous oscillator and p represents the period of the endogenous oscillator. This equation is known as a *circle map*. In the above form, it describes the relative phase of two uncoupled oscillators (e.g., two metronomes ticking away independently at their own tempos).

Coupling is introduced by allowing the exogenous oscillator to perturb the phase of the endogenous oscillator. Such *phase coupling* is incorporated by adding a coupling term to the circle map.

$$\phi_{n+1} = \left(\phi_n + \tfrac{q}{p} + \eta_\phi F_\phi\left(\phi_n\right)\right) \mod 1. \tag{2}$$

Here, η_ϕ is a parameter that controls the coupling strength. The term $F_\phi\left(\phi_n\right)$ is the coupling function, which, given the relative phase, calculates the amount by which the phase is perturbed.

To model the relative phase of an endogenous oscillator and a rhythm, the "period," q of the exogenous oscillator is replaced by the n th inter-onset interval, i_n, in a rhythm. The equation below illustrates this:

$$\phi_{n+1} = \left(\phi_n + \tfrac{i_n}{p} + \eta_\phi F_\phi\left(\phi_n\right)\right) \mod 1. \tag{3}$$

If t_m is the m th onset in a rhythm, then the n th inter-onset interval is $i_n = t_{n+1} - t_n$.

Musical rhythms tend to fluctuate in tempo in a way that listeners can track (Repp 2005). To account for this, adaptive oscillator models implement *period coupling*. The ability of oscillators to adapt their period to a rhythm motivates McAuley (1994) to call these oscillators *adaptive*.

The function below calculates a new period at each onset t_n:

$$p_{n+1} = p_n + p_n \eta_p F_p(\phi_n). \tag{4}$$

The function is parameterized by a coupling-strength parameter, η_p, and a period coupling function, $F_p(\phi_n)$, which calculates the change in the period given a relative phase ϕ_n.

Given suitable coupling functions, the endogenous oscillator will *entrain* (be driven to fire in synchrony) to an approximately periodic rhythm. The degree to which this happens depends on how close the period p of the oscillator is to a (sub)harmonic of the period of the rhythm, q.

The dynamic behavior of the system can be visualized by *regime diagrams*. Such illustrations (e.g., see Large and Kolen 1994) visualize the time it takes for an oscillator to settle into a mode-locked state as a function of the coupling strength and the ratio between the endogenous oscillator's period and a driving pulse.[7] Regime diagrams reveal regions centered around p/q values, where p and q s are small integers, in which stable phase-locked (entrained) states emerge readily. These *entrainment regions* are wider around points where the ratio between p and q can be expressed by small integers (e.g., 1:1, 1:2, 2:3) and increase in width as coupling strength increases. Entrainment regions describe the constraints on pulse and meter perception predicted by adaptive oscillator models.

The oscillator described so far is easily disturbed by rhythms that contain onsets far from where the oscillator "expects" the onset. To allow an oscillator to entrain to a single periodic component in a rhythm that contains more onsets apart from periodic ones, Large and Kolen (1994) propose period and phase coupling functions for which the strength of their effect depends on how close the onset occurs to where the endogenous oscillator predicts it to occur. The further an onset deviates from the prediction, the smaller the influence it exerts on the endogenous oscillator's phase and period. This endows the oscillator with what Large and Kolen call a *temporal receptive field*. The width and sharpness of the temporal receptive field are parameterized. In Large and Kolen's model, these parameters remain fixed throughout a simulation, but Large and Palmer (2002) propose a model in which the temporal receptive field sharpens as onsets occur closer to where they are predicted.

In Large and Kolen's model, another pair of parameters specify a lower and upper bound on the oscillator's period. The oscillator's initial period, called its resting period, lies halfway between the lower and upper bound. When no onsets are encountered within its temporal receptive field, the oscillator maintains its current period. Large and Kolen associate this behavior with the tendency of a pulse percept to be sustained in the absence of rhythmic events (Cooper and Meyer 1960).

A set of adaptive oscillator models proposed by McAuley (1993, 1994, 1995) are similar to Large and Kolen's adaptive oscillator. Instead of gradual phase adaptation, McAuley's models reset their phase based on the relative phase. Furthermore, these oscillators have a resting period to which they gradually return in absence of inputs.

Both Large and Kolen (1994) and McAuley (1995) have associated their models with dynamic attending theory (Jones and Boltz 1989). Large and Jones (1999) present an adaptive oscillator model where self-sustained oscillations take on the role of *attending rhythms* (Jones and Boltz 1989). An attending rhythm consists of periodic pulses of attention, defined as periods of sharp perceptual acuity during which an event is anticipated. In Large and Jones' model, an attentional pulse is implemented by a bell-shaped probability density function centered around phase zero. The width of this distribution is governed by a concentration parameter, reflecting attentional focus. As synchronization (due to entrainment implemented by sinusoidal phase and period coupling) increases, the pulse of attention sharpens, focusing attention in time.

The adaptive oscillator of Large and Jones is further developed by Large and Palmer (2002). In this extension, phase and period coupling depend on the strength of attention (as indexed by the attentional pulse), creating a temporal receptive field. However, since the width of the attentional pulse depends on the degree of synchrony between the oscillator and the rhythm, this temporal receptive field narrows as synchrony increases. Increased synchrony thus leads the oscillator to become less sensitive to events deviating far from where the beat is expected and more sensitive to events close to where the beat is expected.

Large and Jones report simulation results of a model in which two adaptive oscillators, both driven by a rhythm, are bidirectionally coupled to each other. Inter-oscillator coupling is defined such that the oscillators are driven toward either a 2:1 or 3:1 period ratio to ensure metrical entrainment between them. Simulations carried out by Large and Palmer (2002) show that such inter-oscillator coupling can improve entrainment stability in tracking expressive piano performances and that the models can be used to detect phrase-boundaries based on phrase-final lengthening and also to detect melody notes in chords which are accentuated by being timed slightly early (melody leads). Furthermore, Large and Palmer show that in performances performed with strong rubato, detection of melody leads in certain situations to improved beat tracking performance of the adaptive oscillator model.

Large and Palmer (2002) also describe some of their adaptive oscillator model's limitations. First, the model is not suitable for finding the initial beat and has to be provided with this information. Furthermore, for inter-oscillator coupling, functions that actively drive the oscillators to the desired metrical ratios are required, introducing an asymmetry between rhythm-to-oscillator and oscillator-to-oscillator coupling.

Neural Resonance Models

Interactions between excitatory and inhibitory populations of neurons can give rise to neural oscillations (Large et al. 2015). It has been hypothesized that pulse and meter perception are emergent phenomena of such oscillations (Large 2008; Large and Snyder 2009). The neural resonance theory of rhythm and meter perception proposes an explanation based on a mathematical description of a biological mechanism rather than an abstract model like an adaptive oscillator.

At a high level of mathematical abstraction, the dynamics of neural oscillations may be described by a *canonical model* (Large 2008), which describes dynamical behavior that is shared by a large class of more detailed models. *Neural resonance models* of pulse and meter perception are based on a canonical model of neural oscillation.

These models share some characteristics with adaptive oscillator models. Both approaches propose explanations of pulse and meter perception based on coupled oscillation. The phase dynamics (but not amplitude dynamics) of neural resonance models can also be described by a circle map (Large 2008). Unlike adaptive oscillator models, however, neural resonance models posit a specific neural mechanism from which oscillations arise. Furthermore, neural oscillators do not exhibit period coupling: they oscillate near their natural frequency, which does not adapt to tempo fluctuations. Instead neural resonance models posit *gradient-frequency networks* of neural oscillators, in which oscillators with natural frequencies close to (harmonics of) periodicities in the rhythm resonate to the rhythm (Velasco and Large 2011; Large, Herrera, and Velasco 2015).

Neural resonance models exhibit three behaviors associated with different aspects of pulse and meter perception (Large 2010b). *Spontaneous oscillation* relates to the perception of pulse, and, in particular, the tendency for pulse to persist in the absence of exogenous events. *Entrainment* of neural oscillations to exogenous rhythms reflects the perception of a periodic pulse in rhythms that are not strictly periodic. *Higher-order resonance*—the capacity of neural oscillation to resonate at harmonics or sub-harmonics of periodicities in a rhythm—is posited to account for meter induction and for the perception of pulse at frequencies that are absent from the Fourier spectrum of a rhythm (Velasco and Large 2011).

Coupled Oscillation Models and Enculturation

Large (2010a) describes how neural resonance models can be extended with plastic inter-oscillator connections that adapt their strength based on the principles of Hebbian learning. Large, Herrera, and Velasco (2015) note that this makes it possible to simulate the effect of enculturation on neural resonance models. Recently, Tichko and Large (2019) proposed a simulation of effects of exposure to music with non-isochronous meters in infants observed by Hannon and Trehub (2005a,2005b), using a gradient-frequency network of neural oscillators inter-connected by plastic connections. This model resembles a model proposed by Large, Herrera, and Velasco (2015), which consists of two networks connected to each other. One represents a sensory network that receives input from a rhythm, the other represents a motor network that is connected via bidirectional coupling to the sensory network. Citing findings of limited development of movement-to-rhythm synchronization in infants, Tichko and Large only use a sensory network.

To represent exposure to non-isochronous meters in Balkan music and isochronous meters in Western tonal music, Tichko and Large expose two instantiations of their network to a different rhythm. One network is exposed to a 4/4 rhythm, intended to represent exposure to Western tonal music, the other to a 7/8 rhythm, intended to represent exposure to Balkan music in a non-isochronous meter. Both networks, and another network without prior exposure, are exposed to six rhythms: the two training rhythms and two modified versions of each training rhythm, one that preserves the meter and one that violates it. Analyzing the response of the networks to the different rhythms, Tichko and Large show an effect of training that resembles the results obtained by Hannon and Trehub (2005a,2005b). However, because these results are based on simulations involving a single training and a single test rhythm per condition, it remains unclear how robust they are to variation in the specific rhythms used for training and testing. Furthermore, the results appear to be influenced by the frequencies of periodicities in the training rhythm. It could be periodicity, rather than the type of meter (isochronous or non-isochronous), that primarily influenced the results.

Large (2010b) suggests that innate constraints on music perception may emerge from the intrinsic dynamics of the brain. An important question for these models is therefore whether the dynamic behavior of neural resonance models sufficiently explains empirically observed variance in metrical entrainment behavior. Both adaptive oscillators and neural resonance models predict that perceived pulses are constrained to be isochronous. Coupled oscillation models can entrain to periodic components related by simple integer ratios, which may be polyrhythmic (such as 3:2 or 4:3), but they cannot entrain to a non-isochronous beat. For example, while the networks described by Tichko and Large (2019) resonate to rhythms in (isochronous) 4/4 and (non-isochronous) 7/8 meters, they do not entrain to the non-isochronous tactus level of the 7/8 meter. It could be that, as Large (2008, 221) suggests, non-isochronous meters are “compelling specifically because they thwart an intrinsic expectation of periodicity.” However, as discussed earlier in “A Cross-Cultural Perspective,” there is evidence suggesting that non-isochronous meters are readily processed given familiarity with music in which they are prevalent (Hannon and Trehub 2005b; Soley and Hannon 2010; Hannon, Soley, and Ullal et al. 2012). There also is evidence that, given the right kind of training, metrical entrainment to non-isochronous beat-subdivisions is possible (Polak, London, and Jacoby 2016; Polak et al. 2018). Whether rhythm perception shows more flexibility than the constraints of coupled oscillation models allow remains an active topic of discussion.

The Predictive Processing Perspective

The predictive processing perspective (described elaborately by Clark 2013) builds on a class of theories that notably include predictive coding (Rao and Ballard 1999) and the Bayesian brain hypothesis (Knill and Pouget 2004). It has recently come to be associated with several different theoretical perspectives that range from cognitivist to radically embodied (for discussions, see Allen and Friston 2018; Wiese and Metzinger 2017). Nevertheless, these perspectives share a commitment to the idea that perception is based on minimizing prediction error, which in turn is based on Bayesian inference. Predictive processing theories propose a domain-general mechanism that underlies both perception and perceptual learning.

More specifically, predictive processing posits that perception and cognition involve an internal, multi-layered generative model of sensations. This model can be represented as a Bayesian network: a directed acyclic graph that may be interpreted to reflect causal dependencies between random variables (Pearl 2000). Sensations are considered to be the result of a stochastic generative process (the environment) that is predicted by the outcomes generated by leaf nodes of an (internal) generative model. The better the generative model resembles the generative process underlying sensations, which involves the underlying environmental causes of sensations, the more accurately sensations can be predicted. Prediction error, which is continuously generated by the discrepancy between observed and predicted outcomes, revises the generative model to better predict future sensations. These changes, which are driven by prediction error, are hypothesized to underlie both perception and perceptual learning.

Note, however, that this process, and its implied outcome (convergence to a generative model of the environment, limited in accuracy only by physiology), is argued to be significantly altered when the role of *action*—the ability of an organism to influence the flow of sensory stimulation and to shape its environment—is considered (see Clark 2016). This role can be integrated into predictive processing to create what Clark (2013) calls *action-oriented* predictive processing. In any case, sensitivity to the statistical structure of the environment plays a significant role in all predictive processing accounts (action-oriented or not). The probabilistic generative models discussed in this chapter are passive models that do not account for the effects that action may have on their input. Such effects remain an important topic for future research, together with the question of how they interact with the effects of passive exposure simulated by probabilistic generative models.

Prediction-error minimization in predictive processing is equivalent to probabilistic inference in a probabilistic generative model. In such a model, random variables upon which outcomes are conditioned are called *latent* variables. Latent variables cannot directly be observed, but their probability distribution may be inferred through probabilistic inference. The marginal distribution of observed variables corresponds to the generative model's predictions of stochastic outcomes. This distribution assigns a probability to every possible outcome of the generative process. The probability of a given observation is known as the *model evidence* for that observation.

Prediction error is operationalized by the negative logarithm of the model evidence, such that minimizing prediction error corresponds to maximizing model evidence. This quantity corresponds to a measure of *information* defined in information theory (Shannon 1948). A system that minimizes prediction error thus also minimizes information transmission. The intuition behind this is that only parts of the sensory signal that have not already been predicted by the generative model need to be considered.

The approaches discussed below estimate their parameters directly from samples of empirical data, annotated with underlying structure (meter), using a maximum likelihood approach. These samples are called the *training data* of the model. The maximum likelihood approach ensures that the estimated parameters cause the model to assign the maximum possible probability to the training data. This training process has been used to simulate the effects of prior exposure to music on rhythm perception. A later section describes in more detail how probabilistic generative models can be used to simulate enculturation.

It is worth noting that the maximum likelihood approach is different from a so-called fully Bayesian approach, in which a model describes its own parameters as random variables. This allows models to infer their own parameters from data using probabilistic inference and eliminates the need for annotated (also known as labeled) training data, allowing models to "bootstrap" themselves off mere observations. The distinction between a training phase in which the model is parameterized and a testing phase in which the model is evaluated thereby also disappears. Furthermore, the fully Bayesian approach accounts in a principled way for uncertainty that the model has about its own parameters. Such uncertainty plays an important role in predictive processing (see Clark 2016) but is beyond the scope of this chapter.

Probabilistic Generative Modeling of Rhythm Perception

Temperley (2007) makes a strong case for probabilistic approaches to music perception based on the hypothesis that knowledge of musical style is probabilistic in nature and inferred by listeners from regularities in the music they have been exposed to. The basic framework outlined by Temperley applies to all models described in this chapter and also corresponds to the predictive processing framework described above. In this framework, observed variables represent the *musical surface* and latent variables represent its *underlying structure*. For rhythm models, the musical surface corresponds to a pattern of event times, and its structure corresponds to some conceptualization of meter. In predictive processing terms, the musical surface is the outcome of a generative process involving latent variables that represent perceptual concepts like meter.

A generative model of rhythm perception may represent rhythms and meter by multiple random variables, but to obtain a compact representation, these variables can be merged into two variables, namely R (for rhythm) and M (for meter). According to the product rule of probability, the joint distribution of these variables, $p\left(R, M\right)$, can be written in one of the following ways:

$$p\left(R, M\right) = p\left(M|R\right)p\left(R\right) = p\left(R|M\right)p\left(M\right). \tag{5}$$

Since the goal is to describe the *generative* process underlying rhythms, generative models of rhythm perception aim to estimate the factors $p\left(M\right)$—the *a priori* probability of a meter—and $p\left(R|M\right)$—the probability of a rhythm given a meter. When these factors appear in Bayes' theorem, as shown below, they are known as the *prior distribution* and the *likelihood function*. It follows from Equation 5 that the probability of a meter given a rhythm, $p\left(M|R\right)$, can be expressed in terms of the generative model as follows:

$$\overbrace{p(M|R)}^{\text{posterior}} = \frac{\overbrace{p(R|M)}^{\text{likelihood}}\overbrace{p(M)}^{\text{prior}}}{\underbrace{p(R)}_{\text{model evidence}}}. \tag{6}$$

This equation is known as Bayes' theorem and forms the basis of *probabilistic inference* in generative models. Since it enables inferring the distribution of latent variables from an observed outcome, inference is sometimes called the *inversion* of a generative model (MacKay 2003).

Equation 6 reveals some similarities between probabilistic generative models and preference-rule models (described earlier). Since model evidence is independent of latent variables, the posterior probability of a meter is influenced only by the two factors in the numerator of the fraction on the right-hand side of the equation. Meters that are probable *a posteriori* strike a balance between *a priori* probability and the probability of the rhythm given the meter. Meters that are *a priori* improbable require strong bottom-up evidence to be *a posteriori* probable, compared to meters that are *a priori* probable. A similar dynamic interaction occurs in preference-rule models, which postulate rules that apply only to a given metrical analysis (e.g., the regularity rule), comparable to a prior distribution, and rules that measure the fit between an analysis and a rhythm (e.g., the event and length rules), comparable to a likelihood function.

Model evidence, recall, is the probability that a generative model assigns to an outcome. In (Equation 6), it is given by the denominator of the fraction, which may also be written as

$$p\,(R) = \sum^{M} p\,(R|M)p\,(M). \tag{7}$$

Differences between generative rhythm perception models, which are all compatible with this general framework, reside in how the prior distribution and likelihood function are implemented. The sections below describe different possibilities that have been explored in the literature. Where applicable, we describe how these possibilities are applied in different generative models of rhythm proposed by Temperley (2007, 2009) and Van der Weij, Pearce, and Honing (2017).

Rhythmic Outcomes: Grids, Intervals, and Phases

How a generative model represents a rhythm corresponds to how the stochastic outcomes of the model should be interpreted. All of the models we discuss below represent rhythms as temporally ordered sequences of outcomes. These sequences depend only on *note-onset times*: the times at which note events begin (i.e., when they are played, struck, plucked, or sung). Temporal intervals are always represented as integer multiples of some atomic temporal unit, which may either be an absolute duration (e.g., 50 milliseconds) or a symbolic score-duration (e.g., a sixteenth note). However, the models differ in whether they represent rhythms by grids of temporal bins, sequences of temporal intervals, or more abstractly in terms of the metrical functions of notes. Below, we introduce a distinction between four types of models: grid, interval, phase, and metrical salience models.

In *grid* models, stochastic outcomes represent temporally adjacent grid cells, each of which represents an atomic temporal interval. Outcomes in such models are binary variables representing *whether* an onset occurs within (or at) the current grid cell or whether it remains silent. Grid models, in other words, predict *what* happens at the current moment.

Interval models, by contrast, predict *when* an onset occurs relative to the last onset. In interval models, outcomes represent the time interval between two note-onsets: the inter-onset interval. When a model is temporally discrete, this interval is often one out of a prespecified set of possibilities that may occur with non-zero probability. This set of possibilities is sometimes called an alphabet (Conklin and Witten 1995).

Phase and *metrical salience* models predict a more abstract property of the next event, namely its metrical function. By the *phase* of an onset, we mean its position in a metrical cycle denoted by bars notated in a score. By *metrical salience*, we mean the highest metrical level in which a beat associated with the current onset occurs. Phase and metrical salience representations are *variant* to meter: how a given note-onset event is represented depends on its metrical interpretation. Given a meter and the position of bar lines, predictions of metrical salience or phase do not correspond to a unique point in time but constrain the possible points in time at which an event may occur that are in agreement with the prediction. Like interval models, phase and metrical salience models predict *when* an onset occurs but do so more abstractly.

A final important aspect of representation, which is of relevance to generative rhythm models, is that the *granularity* of a representation affects prediction error. Predictions that have a low temporal granularity are more likely to be correct since they are consistent with a large number of events. When the granularity of a representation depends on the value of a latent variable, as is the case for phase and salience models, this introduces a (possibly unintended) bias into the model. Since in phase and salience models, the temporal granularity of a prediction depends on the period of the metrical cycle, such models are susceptible to biases favoring meters with shorter metrical cycles.

The Prior Probability of Meters

The prior distribution of meters, $p(M)$, describes the probability of meters independently, that is, without considering observations (which represent bottom-up sensory input). Van der Weij, Pearce, and Honing (2017) employ a categorical distribution that reflects the relative frequency of meters derived from notated time signatures in empirical training data. This approach makes no assumptions about the internal structure of meter, assuming that this structure may be culture-specific. On the other hand, it has no way of estimating the probability of meters that do not occur in its training data.

Models proposed by Temperley (2007, 2009) employ prior distributions based on a hierarchical view of meter consistent with the ideas of Lerdahl and Jackendoff (1983). In these prior distributions, a meter is generated by a set of stochastic outcomes represented by different random variables, such as the duration of a tactus interval, whether tactus beats are grouped by two or three, and whether tactus beats are subdivided into two or three sub-tactus beats. Compared to the approach of Van der Weij, Pearce, and Honing, this prior requires fewer parameters and can, due to its compositional nature, estimate the probability of meters not occurring in training data. On the other hand, it makes assumptions about the structure of meter that may be specific to the Western musical idiom.

Priors may also be based on abstract theoretical measures. Studying the production and categorical perception of interval ratios, Sadakata, Desain, and Honing (2006) assign prior probabilities to interval ratios that are proportional to a theoretical quantification of the ratio complexity. Such priors are consistent with the hypothesis that, due to cognitive constraints, some meters may be generated more readily than others.

Likelihood Functions: Generating Rhythms from Meters

To illustrate how the design of the likelihood function affects which cues for meter a generative model is sensitive to, we discuss six models described by Temperley (2010) in a model comparison study investigating the probabilistic principles underlying what the study calls "common practice rhythm." Unlike the multilayered generative models of Temperley (2007, 2009) and Van der Weij, Pearce, and Honing (2017), these models assume that the meter is known and fixed.[8] The six models can be distinguished by two aspects of their design: the representation of rhythms and metrical structure, and the probabilistic independence assumptions they make. These aspects are summarized in Table 1 for the six models that Temperley (2010) presents, and the paragraphs below discuss them in more detail.

Table 1: An overview of six likelihood functions discussed by Temperley (2010). The middle column indicates whether each model uses a grid, interval, or phase representation. The right-most column indicates what outcomes of each model are conditioned on. In the right-most column, "N/A" indicates that outcomes are modeled independently.

Model	Representation	Metrical context
Uniform Position Model	Grid	N/A
Zeroth-Order Duration Model	Interval	N/A
Metrical Position Model	Grid	Salience
Fine-grained Position Model	Grid	Phase
Hierarchical Position Model	Grid	Salience, Metrical anchoring
First-Order Metrical Duration Model	Phase	Previous phase

Regarding representation, Temperley distinguishes between "position models" and "duration models." As Table 1 shows, four different position models, and two duration models are discussed. The four position models correspond to what we call grid models. Grid cells, in this case, correspond to eighth-notes. Of the two duration models discussed, one corresponds to what we call an interval model, while the other is a phase model.[9]

The number of probabilistic independence assumptions made by a model must fall somewhere in between two extremes. At one extreme, each random variable depends on all other random variables, which corresponds to a fully connected Bayesian network. At another extreme, the outcome of each variable is assumed to be independent of all other outcomes, which corresponds to an unconnected Bayesian network. The Uniform Position Model, which models the independent probability of an onset at a grid cell, and the Zeroth-Order Duration Model, which models the independent probability of inter-onset intervals, posit only a single variable at each time step. In the search for a model that balances prediction performance with complexity, these models may be seen as baselines against which the effect of progressively removing independence assumptions from the models may be compared.

Some generative models can be evaluated incrementally over a sequence of time steps. This is possible only when variables are conditioned on no other variables than those occurring in the same or the preceding time steps. Models in which variables in each time step are conditioned on variables in the n immediately preceding time steps are called n th-order *Markov models*. For example, in zeroth-order Markov models, outcomes are independent of preceding outcomes, while in first-order Markov models, outcomes depend on variables in the preceding time step. Except for the First-Order Metrical Duration Model and the Hierarchical Position Model, all models compared by Temperley are zeroth-order. The Hierarchical Position Model is not a Markov model: it conditions outcomes at a given metrical level on outcomes at higher metrical levels. Rhythms are generated hierarchically, rather than in temporal order, by this model.

Within a time step, the presence or absence of independence assumptions may incorporate sensitivity to metrical structure into a model. The Uniform Position model and Zeroth-Order Duration Model are not sensitive to metrical structure; that is, the probability of their outcomes is independent of meter. The Metrical Position Model, Fine-Grained Position Model, and Hierarchical Position Model, however, condition outcomes on the metrical status of a grid cell. Of these, the Fine-Grained Position Model differs from the other two in the representation of metrical status: outcomes are conditioned on the phase of a grid cell, while in other two models they depend on the metrical salience of a grid cell. In the Hierarchical Position Model outcomes depend on the metrical salience of the current grid cell and on whether the surrounding metrically stronger beats contain onsets. In Table 1, this situation is referred to as *metrical anchoring* (Temperley 2009).

Another means of introducing sensitivity to meter into a model is by choosing a representation of outcomes that is itself sensitive to meter. This strategy is employed in the First-Order Duration Model, which is a phase model: it predicts the *phase* of an outcome. Since this is a first-order Markov model, the probability of a phase is additionally conditioned on the previous outcome.

The Metrical Position Model is a grid model that most faithfully embodies the theory that the frequent occurrence of onsets on metrically strong beats is a strong cue for meter (Palmer and Krumhansl 1990). The model conditions the probability of an onset at a grid cell on the metrical salience of that grid cell. The metrical salience representation has a lower temporal granularity than the phase representation: for example, the second and fourth beat of a 4/4 bar have different phases but are indistinguishable by metrical salience. If the assumption that metrical salience, rather than phase, most strongly predicts onset likelihood is true, then models based on metrical salience would more compactly capture statistical patterns in common-practice rhythms than models based on phase, and the Metrical Position Model should perform as well as the Fine-Grained Position model, despite having fewer parameters.

A phase representation, on the other hand, assumes periodicity of meter, but otherwise makes few theoretical commitments to its organization. For example, the phase representation of a rhythm does not depend on whether the underlying meter is 3/4 or 6/8. A phase model may be able to distinguish between these meters, but differences between them must be encoded in a probability distribution of phases that is conditioned on meter. These differences may be learned during model training, where the parameters of the model are estimated from empirical training data. In any case, this aspect is irrelevant in Temperley's model comparison study where all considered rhythms have a 4/4 meter.

Temperley evaluates the performance of these six models in terms of the per-rhythm cross-entropy (the negative logarithm of model evidence).[10] The definition of cross-entropy is identical to that of prediction error. Results can therefore be interpreted as representing how well the models predict rhythms in the style of the chosen samples. The objective is to investigate which general principles underlie the composition of what Temperley calls "common-practice rhythm." Accordingly, the models are trained and evaluated on empirical samples of European folksongs and first-violin parts of string quartets by Mozart and Haydn.

The results show that, in general, the four models sensitive to metrical structure achieve better prediction performance than those not sensitive to such structure. Overall, the First-Order Metrical Duration Model achieves the best performance, and the Fine-Grained Position Model outperforms the Metrical Position model. Both of these models are based on a phase representation, suggesting that, even in common-practice rhythm, phase may provide greater predictive power for the timing of notes in the empirical samples of Western music than metrical salience. However, the model comparison does not include a first-order salience model, which would allow for a more elaborate comparison of phase and salience representations.

While the best performance is achieved by the First-Order Metrical Duration Model, the Hierarchical Position Model achieves comparable performance using significantly fewer parameters. Taking this into account, Temperley concludes that the Hierarchical Position Model most accurately captures statistical properties of common-practice rhythms. Findings of Holzapfel (2015), and London, Polak, and Jacoby (2017), however, suggest that the relatively strong performance of this model might not generalize well to non-Western musical idioms, in which metrical salience sometimes is less predictive of onset probability.

Some of the likelihood functions described above are used in multilayered generative rhythm models. In particular, Temperley (2007, 2009) proposes grid models, in which the grid cells (or *pips*) represent not symbolic score-units but real absolute durations. The model described by Temperley (2007) uses the

Metrical Position Model as its likelihood function. The Hierarchical Position Model is used as the likelihood function in the model described by Temperley (2009). However, this model is not a Markov model and violates temporal incrementality. Accordingly, the model is presented primarily as a music analysis model, rather than a music perception model. Both models contain several variables that accommodate a certain degree of freedom in tempo and timing, but these aspects are beyond the scope of this chapter. Another multilayered generative model of rhythms described recently by Van der Weij, Pearce, and Honing (2017) uses a different representation and a different likelihood function. The next section describes this model in more detail.

Modeling Sequential Structure in Rhythms

The models described so far are based on zeroth- or first-order Markov models, or on hierarchical models (Temperley 2009, 2007, 2010). Van der Weij, Pearce, and Honing (2017) instead propose a probabilistic generative model using a *variable-order* Markov model. In this model, events (outcomes) are conditioned on all preceding events in a sequence that represents a rhythm. This is achieved using a modeling technique called prediction by partial match (PPM), proposed originally as a data compression method (Cleary and Witten 1984).

Instead of a grid or phase representation, Van der Weij, Pearce, and Honing (2017) propose a representation of outcomes that is sensitive to meter but maps one-to-one onto inter-onset intervals. The representation combines the phase of an onset with the number of metrical cycles elapsed since the last onset: it encodes the temporal interval between the current event and the bar-level downbeat preceding the last event. The representation is referred to here as the *downbeat distance*. This representation ensures that the temporal granularity of predictions is independent of meter.

Compared to zeroth-order models, a variable-order Markov model of events represented by downbeat distances widens the range of cues for meter that Van der Weij, Pearce, and Honing's model is sensitive to. The probability of an event given a meter depends not only on its metrical context, but also on the downbeat distances of previous events. This changes the role of meter from a periodic template of onset probabilities (Palmer and Krumhansl 1990; Temperley 2007) into a periodic temporal reference with respect to which patterns of events are interpreted and remembered. It allows a model to learn rhythmic patterns that occur predictably in particular metrical contexts. For example, it may be the case that, in a hypothetical musical sample, syncopations occur predictably in certain contexts, even though in the same style, notes generally begin on metrically strong beats. Such predictable deviations from the norm would be undetectable in event-frequency distributions (Palmer and Krumhansl 1990; Holzapfel 2015; London et al. 2017), which are sensitive only to zeroth-order statistical properties of rhythms.

Simulations performed by Van der Weij, Pearce, and Honing suggest that variable-order Markov modeling improves prediction performance of rhythms derived from German folksongs. Applying different variants of their model, in which the maximum order of the variable-order Markov model (the order bound) varies between zero and four, they find that the prediction of rhythms derived from German folksongs improves when the order bound is increased. The performance gain is most pronounced between zeroth-order and first-order modeling, but small improvements occur beyond first-order models.

The increased complexity of the relation between rhythm and meter supported by Van der Weij, Pearce, and Honing's model may improve the model's applicability to music from different cultures. In music from, for example, regions in western Africa (Locke 1982) and the African diaspora (Iyer 1998), it is common for onsets to occur consistently on beats that, according to a Western theoretical understanding of meter (Longuet-Higgins 1978; Lerdahl and Jackendoff 1983), are metrically weak instead of on metrically strong beats. Findings presented by London, Polak, and Jacoby (2017) illustrate this quantitatively for Malian djembe music, and Holzapfel (2015) shows that rhythms in Turkish makam music also deviate from norms based on patterns of metrical salience. It nevertheless remains an open question whether these observations warrant the level of flexibility in the relation between rhythm and meter afforded by Van der Weij, Pearce, and Honing's model. Comparing the performance of different probabilistic generative rhythm models on culturally diverse samples of rhythms may provide more insight into this matter.

The applicability of Van der Weij, Pearce, and Honing's model to rhythms from diverse musical cultures is somewhat hampered by its reliance on Western music notation. Music notation plays little or no role in many musical traditions around the world, and transcribing music from these traditions in Western music notation may not be appropriate. For example, Western music notation's emphasis on temporal intervals related by small-integer ratios cannot naturally express so-called swung beat subdivisions, as found, for example, in jazz music (Honing and De Haas 2008) and Malian djembe music (Polak et al. 2016; Polak et al. 2018).

It appears, in any case, that the model partially fulfills a set of requirements that Iyer (1998) proposes for rhythm perception models; namely that

> [...] any model of rhythm perception and cognition must include stages at which incoming rhythms are compared to known rhythms, matched against known meters, and situated among broader expectations about musical events. It also must involve some degree of what may be called active perception, by which is meant the assessment of various alternative readings of the musical signal, and the switching among them, all carried out *in time* and continually revised and updated. (p. 55; emphasis in original)

Simulating Enculturation with Probabilistic Generative Models

Probabilistic generative models suggest a principled method for simulating the effects of prior exposure on perception. This method can be understood in terms of the exhaustive set of outcomes (event-timing patterns) that a probabilistic generative rhythm perception model can generate. For each item in this set, there is an unknown probability of encountering it as a musical rhythm. For some items, this probability is low because they are unlikely rhythms; for others, it is high because they, for example, correspond to stereotypical rhythms. There is, in other words, an unknown probability distribution of musical rhythms. To minimize prediction error, a generative rhythm perception model aims to approximate this distribution as closely as possible.

The approximation is performed by estimating the model's parameters from a (relatively) small sample drawn from the target distribution. How well the parameterized model approximates the unknown target distribution (its generalization performance) is usually evaluated by testing the model on another small sample from this distribution. The generalization performance depends on whether the model's design enables it to capture the relevant statistical properties from the training sample. Evaluating which aspects of a model's design improve the model's generalization performance may provide valuable insights into the statistical constraints that underlie a corpus of rhythms.

However, the target distribution of relevance to enculturated (situated) individuals depends on the cultural environment of those individuals. A generative model aiming to simulate the perception of such individuals should derive its parameters from a sample that represents music that an enculturated individual is likely to have been exposed to. Music corpora, such as the Essen folksong collection (Schaffrath and Huron 1995), may be used for this purpose. Parameterizations that result from training a generative model on such a sample can be seen as a simulation of an enculturated listener (Van der Weij, Pearce, and Honing 2017). The success of the enculturated model in predicting perceptual idiosyncrasies resulting from such biased sampling may provide evidence as to whether learning mechanisms of listeners resemble those posited by predictive processing. This approach is entirely compatible with the cultural distance hypothesis of Demorest and Morrison (2016) and Morrison, Demorest, and Pearce (2019), according to which the degree of overlap in statistical structure between the music of two cultures predicts the ability of listeners from those cultures to process music from the other culture.

Conclusion

A great variety of rhythm perception models exists in the music cognition literature. Some of these models propose incremental changes to other models, but others propose radically different principles. This chapter reviews a selection of previously proposed rhythm perception models and associates them with three broad perspectives—cognitivism, embodied cognition, and predictive processing—that entail different views on the nature of perception and cognition.

The cognitivist perspective describes cognition as information processing involving the rule-based manipulation of symbolic representations. Rule-based models of rhythm perception, such as those proposed by Longuet-Higgins and Steedman (1971), Longuet-Higgins (1976), and Longuet-Higgins and Lee (1982), can be associated with this perspective. Embodied cognition instead emphasizes the role of continuous dynamic interaction between brain, body, and environment. Coupled oscillation models (McAuley 1994; Large and Kolen 1994; Large and Snyder 2009), although they do not always emphasize an explicit role of embodiment, are consistent with this view. Finally, predictive processing views perception and perceptual learning as the result of a single underlying mechanism, namely prediction error minimization based on Bayesian inference. Probabilistic generative models, such as those proposed by Temperley (2007, 2009), are consistent with this perspective.

Additionally, this chapter reviewed literature that studies the role of enculturation in shaping rhythm perception. Although the extent to which rhythm perception is constrained by enculturation and by universal principles remains a topic of debate, it seems uncontroversial that experience, training, and practice play a role. Despite this consensus, few models of rhythm perception account for the possible effects of enculturation. Instead, some models aim to represent universal aspects of perception (such as Povel and Essens 1985; Large 2010b), while others aim to model the perceptual processes of listeners enculturated in a musical idiom (such as Longuet-Higgins 1979).

Probabilistic generative models, which are consistent with the principles of predictive processing, can simulate the effect of previous exposure to rhythms by deriving their parameters from empirical samples of music. Neural resonance models have recently been extended to use plastic connections and Hebbian learning enabling them to adapt based on previous exposure to rhythms (Large 2010a; Large et al. 2015). In general, rhythm perception models have primarily been evaluated on datasets of Western tonal music. Therefore, it seems fruitful for future research to evaluate and compare probabilistic models on culturally diverse samples of rhythms.

References

Agawu, Kofi. 1995. “The Invention of ‘African Rhythm’.” *Journal of the American Musicological Society* 48 (3) 380–395.
Google Scholar WorldCat

Allen, Micah, and Karl J. Friston. 2018. “From Cognitivism to Autopoiesis: Towards a Computational Framework for the Embodied Mind.” *Synthese* 195 (6): 2459–2482.
Google Scholar WorldCat

Anderson, Michael L. 2003. “Embodied Cognition: A Field Guide.” *Artificial Intelligence* 149 (1): 91–130.
Google Scholar WorldCat

Bouwer, Fleur L., J. Ashley Burgoyne, Daan Odijk, Henkjan Honing, and Jessica A. Grahn. 2018. “What Makes a Rhythm Complex? The Influence of Musical Training and Accent Type on Beat Perception.” *PLOS One* 13 (1): 1–26.
Google Scholar WorldCat

Brooks, Rodney A. 1991. “Intelligence Without Representation.” *Artificial Intelligence* 47 (1–3): 139–159.
Google Scholar WorldCat

Bundy, Alan. 1990. “What Kind of Field is AI?” In *The Foundations of Artificial Intelligence: A Sourcebook*, edited by Derek Partridge and Yorick Wilks, 215–222. Cambridge: Cambridge University Press.
Google Scholar Google Preview WorldCat COPAC

Chemero, Anthony. 2009. *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.
Google Scholar Google Preview WorldCat COPAC

Clark, Andy. 2013. “Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science.” *The Behavioral and Brain Sciences* 36 (3): 181–253.
Google Scholar WorldCat

Clark, Andy. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.
Google Scholar Google Preview WorldCat COPAC

Clarke, Eric F. 1987. “Categorical Rhythm Perception: An Ecological Perspective.” In *Action and Perception in Rhythm and Music*, edited by Alf Gabrielsson, 19–33. Stockholm: Royal Swedish Academy of Music.
Google Scholar Google Preview WorldCat COPAC

Clarke, Eric F. 1989. “The Perception of Expressive Timing in Music.” *Psychological Research* 51 (1): 2–9.
Google Scholar WorldCat

Clarke, Eric F. 1999. “Rhythm and Timing in Music.” In *The Psychology of Music*, 2nd ed., edited by Diana Deutsch, 473–500. New York: Academic Press.
Google Scholar Google Preview WorldCat COPAC

Cleary, John G., and Ian H. Witten. 1984. “Data Compression Using Adaptive Coding and Partial String Matching. ” *IEEE Transactions on Communications* 32 (4): 396–402.
Google Scholar WorldCat

Conklin, Darrell, and Ian H. Witten. 1995. “Multiple Viewpoint Systems for Music Prediction.” *Journal of New Music Research* 24 (1), 51–73.
Google Scholar WorldCat

Cooper, Grosvenor and Leonard B. Meyer. 1960. *The Rhythmic Structure of Music*. Chicago: The University of Chicago Press.
Google Scholar Google Preview WorldCat COPAC

Demorest, Steven M., and Steven J. Morrison. 2016. "Quantifying Culture: The Cultural Distance Hypothesis of Melodic Expectancy." In *The Oxford Handbook of Cultural Neuroscience*, edited by Joan Y. Chiao, Shu-Chen Li, Rebecca Seligman, and Robert Turner, 183–194. Oxford: Oxford University Press.
Google Scholar Google Preview WorldCat COPAC

Desain, Peter, and Henkjan Honing. 1992. *Music, Mind and Machine*. Amsterdam: Thesis Publishers.
Google Scholar Google Preview WorldCat COPAC

Desain, Peter, and Henkjan Honing. 1999. "Computational Models of Beat Induction: The Rule-Based Approach." *Journal of New Music Research* 28 (1): 29–42.
Google Scholar WorldCat

Desain, Peter, and Henkjan Honing. 2003. "The Formation of Rhythmic Categories and Metric Priming." *Perception* 32 (3): 341–365.
Google Scholar WorldCat

Fitch, W. Tecumseh, and Andrew J. Rosenfeld. 2007. "Perception and Production of Syncopated Rhythms." *Music Perception* 25 (1): 43–58.
Google Scholar WorldCat

van Gelder, Tim. 1995. "What Might Cognition Be, If Not Computation?" *The Journal of Philosophy* 92 (7): 345–381.
Google Scholar WorldCat

Grahn, Jessica A., and Matthew Brett. 2007. "Rhythm and Beat Perception in Motor Areas of the Brain." *Journal of Cognitive Neuroscience* 19 (5): 893–906.
Google Scholar WorldCat

Hannon, Erin E., Gaye Soley, and Sangeeta Ullal. 2012. "Familiarity Overrides Complexity in Rhythm Perception: A Cross-Cultural Comparison of American and Turkish Listeners." *Journal of Experimental Psychology*: Human Perception and Performance 38 (3): 543–548.
WorldCat

Hannon, Erin E., and Sandra E. Trehub. 2005a. "Metrical Categories in Infancy and Adulthood." *Psychological Science* 16 (1): 48–55.
Google Scholar WorldCat

Hannon, Erin E., and Sandra E. Trehub. 2005b. "Tuning in to Musical Rhythms: Infants Learn More Readily than Adults." *Proceedings of the National Academy of Sciences* 102 (35): 12639–12643.
Google Scholar WorldCat

Hansen, Niels Christian. 2010. "The Legacy of Lerdahl and Jackendoff's A Generative Theory of Tonal Music: Bridging a Significant Event in the History of Music Theory and Recent Developments in Cognitive Music Research." *Danish Yearbook of Musicology* 38:33–55.
Google Scholar WorldCat

Holzapfel, André. 2015. "Relation Between Surface Rhythm and Rhythmic Modes in Turkish Makam Music." *Journal of New Music Research* 44 (1): 25–38.
Google Scholar WorldCat

Honing, Henkjan, and Fleur L. Bouwer. 2019. "Rhythm." In Foundations of Music Psychology: *Theory and Research*, edited by J. Rentfrow and D. Levitin, 33–70. Cambridge, MA: MIT Press.
Google Scholar Google Preview WorldCat COPAC

Honing, Henkjan, and W. Bas de Haas. 2008. "Swing Once More: Relating Timing and Tempo in Expert Jazz Drumming." *Music Perception* 25 (5): 471–476.

Google Scholar WorldCat

Huron, David B. 2008. “Lost in Music.” *Nature* 453 (7194): 456–457.
Google Scholar WorldCat

Iyer, Vijay. 1998. “*Microstructures of Feel, Macrostructures of Sound: Embodied Cognition in West African and African-American Musics.*” PhD diss., University of California, Berkeley.
Google Scholar Google Preview WorldCat COPAC

Jacoby, Nori, Elizabeth Hellmuth Margulis, Martin Clayton, Erin Hannon, Henkjan Honing, John Iversen, Tobias Robert Klein, et al. 2020. “Cross-Cultural Work in Music Cognition: Challenges, Insights, and Recommendations.” *Music Perception* 37 (3): 185–195.
Google Scholar WorldCat

Jacoby, Nori, and Josh H. McDermott. 2017. “Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction.” *Current Biology* 27 (3): 359–370.
Google Scholar WorldCat

Jones, Mari Riess, and Marilyn Boltz. 1989. “Dynamic Attending and Responses to Time.” *Psychological Review* 96 (3): 459–491.
Google Scholar WorldCat

Karaosmanoğlu, M. Kemal. (2012). “*A Turkish Makam Music Symbolic Database for Music Information Retrieval: SymbTr.*” *In Proceedings of the 13th International Society for Music Information Retrieval*, 223–228. Tapei, Taiwan.
Google Scholar Google Preview WorldCat COPAC

Knill, David C., and Alexandre Pouget. 2004. “The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation.” *TRENDS in Neurosciences* 27 (12): 712–719.
Google Scholar WorldCat

Kvifte, Tellef. 2007. “On the Perception of Meter.” *Ethnomusicology* 51 (1): 64–84.
Google Scholar WorldCat

Large, Edward W. 2008. “Resonating to Musical Rhythm: Theory and Experiment.” In *Psychology of Time*, edited by Simon Grondin, 189–231. Bingley, UK: Emerald Publishing Group.
Google Scholar Google Preview WorldCat COPAC

Large, Edward W. 2010a. “A Dynamical Systems Approach to Musical Tonality.” In *Nonlinear Dynamics in Human Behavior*, edited by Raoul Huys and Viktor K. Jirsa, 193–211. Berlin: Springer.
Google Scholar Google Preview WorldCat COPAC

Large, Edward W. 2010b. “Neurodynamics of Music.” In *Music Perception*, edited by Mari Riess Jones, Richard R. Fay, and Arthur N. Popper, 36:201–231. Springer Handbook of Auditory Research. New York: Springer.
Google Scholar Google Preview WorldCat COPAC

Large, Edward W., Jorge A. Herrera, and Marc J. Velasco. 2015. “Neural Networks for Beat Perception in Musical Rhythm.” *Frontiers in Systems Neuroscience* 11:1–14.
Google Scholar WorldCat

Large, Edward W., and Mari Riess Jones. 1999. “The Dynamics of Attending: How People Track Time-Varying Events.” *Psychological Review* 106 (1): 119–159.
Google Scholar WorldCat

Large, Edward W., and John F. Kolen. 1994. “Resonance and the Perception of Musical Meter.” *Connection Science* 6 (1): 177–208.
Google Scholar WorldCat

Large, Edward W., and Caroline Palmer. 2002. “Perceiving Temporal Regularity in Music.” *Cognitive Science* 26 (1): 1–37.
Google Scholar WorldCat

Large, Edward W., and Joel S. Snyder. 2009. “Pulse and Meter as Neural Resonance.” *Annals of the New York Academy of Sciences* 1169 (1): 46–57.
Google Scholar WorldCat

Lee, Christopher S. 1991. “The Perception of Metrical Structure: Experimental Evidence and a Model.” In *Representing Musical Structure*, edited by Peter Howell, Robert West, and Ian Cross, 59–127. London: Academic Press.
Google Scholar Google Preview WorldCat COPAC

Lerdahl, Fred, and Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
Google Scholar Google Preview WorldCat COPAC

Locke, David. 1982. “Principles of Offbeat Timing and Cross-Rhythm in Southern Eve Dance Drumming.” *Ethnomusicology* 26 (2): 217–246.
Google Scholar WorldCat

London, Justin. 1993. “Loud Rests and Other Strange Metric Phenomena or, Meter as Heard.” *Music Theory Online* 0 (2).

London, Justin. 1995. “Some Examples of Complex Meters and Their Implications for Models of Metric Perception.” *Music Perception* 13 (1): 59–77.
Google Scholar WorldCat

London, Justin. 2004. *Hearing in Time: Psychological Aspects of Musical Meter*. New York: Oxford University Press.
Google Scholar Google Preview WorldCat COPAC

London, Justin. 2012. *Hearing in Time*. 2nd ed. New York: Oxford University Press.
Google Scholar Google Preview WorldCat COPAC

London, Justin, Rainer Polak, and Nori Jacoby. 2017. “Rhythm Histograms and Musical Meter: A Corpus Study of Malian Percussion Music.” *Psychonomic Bulletin & Review* 24 (2): 474–480.
Google Scholar WorldCat

Longuet-Higgins, Hugh Christopher. 1976. “Perception of Melodies.” *Nature* 263:646–653.
Google Scholar WorldCat

Longuet-Higgins, Hugh Christopher. 1978. “The Perception of Music.” *Interdisciplinary Science Reviews* 3 (2): 148–156.
Google Scholar WorldCat

Longuet-Higgins, Hugh Christopher. 1979. “The Perception of Music.” *Proceedings of the Royal Society of London* B 205 (1160): 307–322.
WorldCat

Longuet-Higgins, Hugh Christopher, and Christopher S. Lee. 1982. “The *Perception* of Musical Rhythms.” *Perception* 11 (2): 115–128.

Longuet-Higgins, Hugh Christopher, and Christopher S. Lee. 1984. “The Rhythmic Interpretation of Monophonic Music.” *Music Perception* 1 (4): 424–441.
Google Scholar WorldCat

Longuet-Higgins, Hugh Christopher, and Mark J. Steedman. 1971. “On Interpreting Bach.” In *Machine Intelligence*, edited by Bernard Meltzer and Donald Michie, 6:221–241. Edinburgh: Edinburgh University Press.
Google Scholar Google Preview WorldCat COPAC

MacKay, David J. C. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.
Google Scholar Google Preview WorldCat COPAC

Marcus, Scott. 2001. “Rhythmic Modes in Middle-Eastern Music.” In *Garland Encyclopedia of World Music, Volume 6: Middle East*, edited by Virginia Danielson, Dwight Reynolds, and Scott Marcus, 89–92. London: Routledge.
Google Scholar Google Preview WorldCat COPAC

Marr, David. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: W. H. Freeman.
Google Scholar Google Preview WorldCat COPAC

McAuley, J. Devin. 1993. *Learning to Perceive and Produce Rhythmic Patterns in an Artificial Neural Network*. Technical report. Computer Science Department, Indiana University.
Google Scholar Google Preview WorldCat COPAC

McAuley, J. Devin. 1994. “Finding Metrical Structure in Time.” In *Proceedings of the 1993 Connectionist Models Summer School*, edited by Michael C. Mozer, David S. Touretzky, and Paul Smolensky, 219–227. New York: Psychology Press.
Google Scholar Google Preview WorldCat COPAC

McAuley, J. Devin. 1995. “Perception of Time as Phase: Toward an Adaptive-Oscillator Model of Rhythmic Pattern Processing.” PhD diss., Indiana University.

Mehr, Samuel A., Manvir Singh, Dean Knox, Daniel M. Ketter, Daniel Pickens-Jones, S. Atwood, Christopher Lucas, et al. 2019. “Universality and Diversity in Human Song.” *Science* 366 (6468). DOI: 10.1126/science.aax0868
Google Scholar WorldCat

Morrison, Steven J., Steven M. Demorest, and Marcus T. Pearce. 2019. “Cultural Distance: A Computational Approach to Exploring Cultural Influences on Music Cognition.” In *The Oxford Handbook of Music and the Brain*, edited by Michael H. Thaut and Donald A. Hodges, 42–56. New York: Oxford University Press.
Google Scholar Google Preview WorldCat COPAC

Newell, Allen, and Herbert A. Simon. 1976. “Computer Science as Empirical Inquiry: Symbols and Search.” *Communications of the ACM* 19 (3): 113–126.
Google Scholar WorldCat

Palmer, Caroline, and Carol L. Krumhansl. 1990. “Mental Representations for Musical Meter.” *Journal of Experimental Psychology: Human Perception and Performance* 16 (4): 728–741.
Google Scholar WorldCat

Patel, Aniruddh D., and Steven M. Demorest. 2013. “Comparative Music Cognition: Cross-Species and Cross-Cultural Studies.” In *The Psychology of Music*, 3rd ed., edited by Diana Deutsch, 647–681. London: Academic Press.
Google Scholar Google Preview WorldCat COPAC

Pearce, Marcus T. 2018. “Statistical Learning and Probabilistic Prediction in Music Cognition: Mechanisms of Stylistic Enculturation.” *Annals of the New York Academy of Sciences* 1423 (1): 378–395.
Google Scholar WorldCat

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.
Google Scholar Google Preview WorldCat COPAC

Polak, Rainer, Nori Jacoby, Timo Fischinger, Daniel Goldberg, André Holzapfel, and Justin London. 2018. “Rhythmic Prototypes Across Cultures: A Comparative Study of Tapping Synchronization.” *Music Perception* 36 (1): 1–23.
Google Scholar WorldCat

Polak, Rainer, Justin London, and Nori Jacoby. 2016. “Both Isochronous and Non-Isochronous Metrical Subdivision Afford

Precise and Stable Ensemble Entrainment: A Corpus Study of Malian Jembe Drumming." *Frontiers in Neuroscience* 10:1–11.
Google Scholar WorldCat

Povel, Dirk-Jan, and Peter Essens. 1985. "Perception of Temporal Patterns." *Music Perception* 2 (4): 411–440.
Google Scholar WorldCat

Rao, Rajesh P. N., and Dana H. Ballard. 1999. "Predictive Coding in the Visual Cortex: A Functional Interpretation of some Extra-Classical Receptive-Field Effects." *Nature Neuroscience* 2 (1): 79–87.
Google Scholar WorldCat

Repp, Bruno H. 1995. "Expressive timing in Schumann's "Träumerei:" An Analysis of Performances by Graduate Student Pianists." *The Journal of the Acoustical Society of America* 98 (5): 2413–2427.
Google Scholar WorldCat

Repp, Bruno H. 2005. "Sensorimotor Synchronization: A Review of the Tapping Literature." *Psychonomic Bulletin & Review* 12 (6): 969–992.
Google Scholar WorldCat

Repp, Bruno H., and Yi-Huang Su. 2013. "Sensorimotor Synchronization: A Review of Recent Research (2006–2012)." *Psychonomic Bulletin & Review* 20 (3): 403–452.
Google Scholar WorldCat

Sadakata, Makiko, Peter Desain, and Henkjan Honing. 2006. "The Bayesian Way to Relate Rhythm Perception and Production." *Music Perception* 23 (3): 269–288.
Google Scholar WorldCat

Savage, Patrick E., Steven Brown, Emi Sakai, and Thomas E. Currie. 2015. "Statistical Universals Reveal the Structures and Functions of Human Music." *Proceedings of the National Academy of Sciences* 112 (29): 8987–8992.
Google Scholar WorldCat

Schaffrath, Helmut, and David B. Huron. 1995. The *Essen Folksong Collection in the Humdrum Kern Format*. Center for Computer Assisted Research in the Humanities. Accessed 2018. https://kern.humdrum.org/cgi-bin/browse?l=/essen.
WorldCat

Shannon, Claude Elwood. 1948. "A Mathematical Theory of Communication." The *Bell System Technical Journal* 27 (3): 379–423.
WorldCat

Shmulevich, Ilya, and Dirk-Jan Povel. 2000. "Measures of Temporal Pattern Complexity." *Journal of New Music Research* 29 (1): 61–69.
Google Scholar WorldCat

Shove, Patrick, and Bruno H. Repp. 1995. "Musical Motion and Performance: Theoretical and Empirical Perspectives." In *The Practice of Performance: Studies in Musical Interpretation*, edited by John Rink, 55–83. Cambridge: Cambridge University Press.
Google Scholar Google Preview WorldCat COPAC

Soley, Gaye, and Erin E. Hannon. 2010. "Infants Prefer the Musical Meter of Their Own Culture: A Cross-Cultural Comparison." *Developmental Psychology* 46 (1): 286–292.
Google Scholar WorldCat

Song, Chunyang, Andrew J. R. Simpson, Christopher A. Harte, Marcus T. Pearce, and Mark B. Sandler. 2013. "Syncopation and the Score." *PLOS One* 8 (9): 1–7.
Google Scholar WorldCat

Steedman, Mark J. 1977. "The *Perception* of Musical Rhythm and Metre." *Perception* 6 (5): 555–569.

Temperley, David. 2000. “Meter and Grouping in African Music: A View from Music Theory.” *Ethnomusicology* 44 (1): 65–96.
Google Scholar WorldCat

Temperley, David. 2001. *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press.
Google Scholar Google Preview WorldCat COPAC

Temperley, David. 2007. *Music and probability*. Cambridge, MA: MIT Press.
Google Scholar Google Preview WorldCat COPAC

Temperley, David. 2009. “A Unified Probabilistic Model for Polyphonic Music Analysis.” *Journal of New Music Research* 38 (1): 3–18.
Google Scholar WorldCat

Temperley, David. 2010. “Modeling Common-Practice Rhythm.” *Music Perception* 27 (5): 355–376.
Google Scholar WorldCat

Temperley, David. 2013. “Computational Models of Music Cognition.” In *The Psychology of Music*, 3rd ed., edited by Diana Deutsch, 327–368. London: Academic Press.
Google Scholar Google Preview WorldCat COPAC

Temperley, David, and Daniel Sleator. 1999. “Modeling Meter and Harmony: A Preference-Rule Approach.” *Computer Music Journal* 23 (1): 10–27.
Google Scholar WorldCat

Thompson, Bill, Simon Kirby, and Kenny Smith. 2016. “Culture Shapes the Evolution of Cognition.” *Proceedings of the National Academy of Sciences* 113 (16): 4530–4535.
Google Scholar WorldCat

Tichko, Parker, and Edward W. Large. 2019. “Modeling Infants’ Perceptual Narrowing to Musical Rhythms: Neural Oscillation and Hebbian Plasticity.” *Annals of the New York Academy of Sciences* 1453 (1): 125–139.
Google Scholar WorldCat

Todd, Neil P. M., and Christopher S. Lee. 2015. “The Sensory-Motor Theory of Rhythm and Beat Induction 20 Years On: A New Synthesis and Future Perspectives.” *Frontiers in Human Neuroscience* 9:1–25.
Google Scholar WorldCat

Trehub, Sandra E., Judith Becker, and Iain Morley. 2015. “Cross-Cultural Perspectives on Music and Musicality.” *Philosophical Transactions of the Royal Society B* 370:1–9.
Google Scholar WorldCat

Velasco, Marc J., and Edward W. Large. 2011. “*Pulse Detection in Syncopated Rhythms using Neural Oscillators.*” In *Proceedings of the 12th International Society for Music Information Retrieval*, 185–190. Curitiba, Brazil
Google Scholar Google Preview WorldCat COPAC

Vuust, Peter, and Maria A. G. Witek. 2014. “Rhythmic Complexity and Predictive Coding: A Novel Approach to Modeling Rhythm and Meter Perception in Music.” *Frontiers in Psychology* 5:1–14.
Google Scholar WorldCat

van der Weij, Bastiaan, Marcus T. Pearce, and Henkjan Honing. 2017. “A Probabilistic Model of Meter Perception: Simulating Enculturation.” *Frontiers in Psychology* 8:1–18.
Google Scholar WorldCat

Wiese, Wanja, and Thomas Metzinger. 2017. “Vanilla PP for Philosophers: A Primer on Predictive Processing.” In PPP - *Philosophy and Predictive Processing*, edited by Wanja Wiese and Thomas Metzinger, 1–18. Frankfurt am Main.

Wilson, Andrew D., and Sabrina Golonka. 2013. "Embodied Cognition Is Not What you Think It Is." *Frontiers in Psychology* 4:1–13.
Google Scholar WorldCat

Witek, Maria A. G., Eric F. Clarke, Mikkel Wallentin, Morten L. Kringelbach, and Peter Vuust. 2014. "Syncopation, Body-Movement and Pleasure in Groove Music." *PLOS One* 9 (4): 1–12.
Google Scholar WorldCat

Notes

1 Music Cognition Group, Amsterdam Brain and Cognition, Institute for Logic, Language, and Computation, University of Amsterdam, Amsterdam, Netherlands

2 Music Cognition Lab, School of Electrical Engineering and Computer Science, Queen Mary University of London, London, United Kingdom

3 Huron (2008, p. 457) illustrates this point anecdotally, describing an episode in which he, joining an expedition of biologists, encountered subsistence hunters in the western Amazon who, thanks to their transistor radios, were familiar with Western popular music.

4 A deadpan (or mechanical) performance is one that exactly reproduces the note duration ratios dictated by a notated musical score. For musicians performing music from a score, the goal is rarely to produce such (mechanical-sounding) performances. Instead, expressive tempo changes and deviations from deadpan timing are the norm (Clarke 1989; Repp 1995).

5 This procedure is used in a computer program that can transcribe a melody played on an organ console connected to a high-speed paper tape punch (a MIDI keyboard would nowadays suffice) into musical notation. The program has to be supplied with an initial beat interval, similar to a drummer's count off before a performance. It then tries to track and, if necessary, subdivide this beat throughout a performance. No detailed description of the program is provided, but its source code (written in the POP2 programming language), was made available. The program, as well as a translation into the LISP programming language, is described by Desain and Honing (1992, p. 294).

6 Coupled oscillation models do arguably incorporate some domain-specific constraints: Adaptive and neural oscillators need to be tuned to frequencies that are relevant to musical rhythms. Period coupling and temporal receptive fields in adaptive oscillator models are explicitly introduced to account for music perception and do not occur in physical coupled oscillation systems such as clocks suspended from the same beam or metronomes on the same moving platform.

7 Mode-locking is a generalization of phase-locking that describes states in which one oscillator aligns its phase with another oscillator exactly every n cycles (where n is an integer).

8 By a multilayered generative model, we mean a generative model that conditions observations on underlying latent variables.

9 We use the term "phase" for what Temperley calls "metrical position."

10 It should be noted that there are subtle issues, not mentioned by Temperley, involved in comparing these results between different (grid, interval, or phase) representations. For example, a grid representation of a rhythm is (10101), which is a sequence of five binary outcomes. This rhythm is one of $2^5 = 32$ possible rhythms. In an interval representation, each outcome is one of x possible intervals. For a model that considers $x = 8$ intervals per outcome with non-zero probability, the same rhythm, represented as (22), is one of $8^2 = 64$ possible outcomes. In a phase representation, assuming atomic temporal units of quarter notes and a meter with a period of four quarter notes, the same rhythm is represented as (020), which is one of $4^3 = 64$ possibilities. Predicting one out of sixty-four possibilities is more difficult than predicting one out of thirty-two possibilities. Grid models are thus likely to have an advantage over interval and phase models.