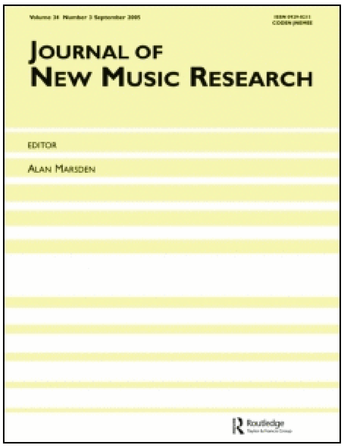

Journal of New Music Research
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713817838

A Positive-Evidence Model for Rhythmical Beat Induction

Douglas Eck

Online publication date: 09 August 2010

To cite this Article Eck, Douglas(2001) 'A Positive-Evidence Model for Rhythmical Beat Induction', Journal of New Music Research, 30: 2, 187 — 200
To link to this Article: DOI: 10.1076/jnmr.30.2.187.7110
URL: http://dx.doi.org/10.1076/jnmr.30.2.187.7110

Journal of New Music Research
2001, Vol. 30, No. 2, pp. 187–200

0929-8215/01/3002-187$16.00

A Positive-Evidence Model for Rhythmical Beat Induction

Douglas Eck

IDSIA, Galleria 2, Lugano, Switzerland

Abstract

The Normalized Positive (NPOS) model is a rule-based model that predicts downbeat location and pattern complexity in rhythmical patterns. Though derived from several existing models, the NPOS model is particularly effective at making correct predictions while at the same time having low complexity. In this paper, the details of the model are explored and a comparison is made to existing models. Several datasets are used to examine the complexity predictions of the model. Special attention is paid to the model's ability to account for the effects of musical experience on beat induction.

1 Introduction

A pattern – whether verbal, musical or otherwise – that evokes a sense of movement or pulse is called a *rhythm*. When onsets in a rhythm are related hierarchically, as they often are in music, the pattern is said to be *metrical*. Meter is the sense of strong and weak onsets that arises from interactions among hierarchical levels. This hierarchy is implied in Western music notation, where different levels are indicated by kinds of notes (whole notes, half notes, quarter notes, etc.) and where bars establish measures of equal length. In Figure 1 a metrical hierarchy is shown for a repeating length-16 rhythmical pattern. Here it can be seen that events at progressively higher levels in the hierarchy always line up with events in the lower levels. When an onset is marked at many levels it is metrically strong (such as note 9). By constraining the temporal organization of music, meter makes it easier for the listener to form expectations about when particularly important onsets will occur. Though patterns like this one generally have only one best metrical interpretation, more complex patterns can yield several plausible interpretations. For example, in repeating length-12 patterns it is often possible to have competing 3/4 and 4/4 meters. See Povel and Essens (1985), Experiment 2 for an exploration of this phenomenon. Another example of complex meter lies in polyrhythmic patterns where two candidate, incompatible meters are considered in parallel. See Handel (1984) for more on polyrhythms. Asymmetric meters (e.g., 5/4, 7/8) are complex in another way: though they fit into a single hierarchy and so pose no problem for the algorithms considered here, the ratio relationships between levels are irregular and often high.

One remarkable quality of music is the sense of movement that it evokes in a listener. This feeling of movement arises quickly and effortlessly; after just a few seconds of hearing a new song, a sense of coherence takes hold, and what was before a series of disconnected sounds now seems whole. By helping evoke this movement or pulse, meter can be seen as an important organizing force in music perception (Parncutt, 1994; Handel, 1989). There is evidence to suggest that cognition in non-musical domains is also guided by meter and rhythm. Speech, for example, has been shown to be rhythmical, at least in speech cycling tasks (Cummins & Port, 1998). Also, experiments with rhythmical finger wagging (Haken et al., 1985) reveal finger-to-finger relationships consistent with the low-integer ratios found in the metrical hierarchy.

One way to discover a listener's metrical interpretation of a musical segment is to ask them to find downbeats by tapping with their hands or feet. A downbeat is a particularly salient beat in a rhythmical pattern and, in general, is metrically important. Jones et al. (Jones, 1976; Jones & Boltz, 1989; Large & Jones, 1999) underscore the importance of downbeats by claiming that human attention itself is guided by the process of beat induction. They forward a theory

Thanks to Henkjan Honing and an anonymous reviewer for valuable comments. This work supported by SNF project 21-49144.96

Accepted: 10 July, 2001

Correspondence: Douglas Eck, IDSIA, Galleria 2, 6928 Manno-Lugano, Switzerland. E-mail: doug@idsia.ch.URL:http://www.idsia.ch/~doug

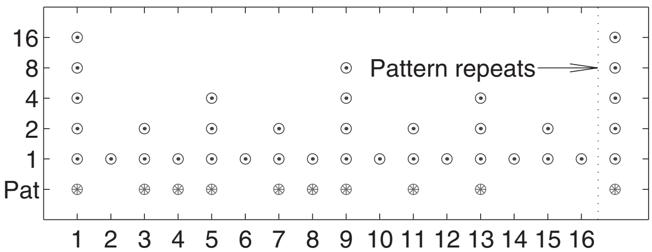

Fig. 1. One candidate metrical hierarchy for a rhythmical pattern. The pattern, at the bottom, repeats every 16 time units. Progressively higher levels in the hierarchy are numbered on the vertical axis. Relatively strong events occur when many levels are aligned, such as at note 9. Alternately, events in the input that coincide with few levels in the hierarchy are relatively weak, such as at note 6.

that uses entrainment to rhythmical pulse in order to account for attentional allocation not only to music but to behaviors as varied as mother-baby interaction and basketball playing. Whether or not a theory based on rhythmical beat induction can indeed account for such a wide range of attentional behaviors is still unanswered. However, by suggesting such a theory, Jones et al. highlight the importance of rhythm and meter not only for music but for a host of day-to-day behaviors.

2 Overview

Significant research has gone into algorithmic methods or finding downbeats in rhythms (Longuet-Higgins & Lee, 1982, 1984; Povel & Essens, 1985; McAuley & Semple, 1999; Parncutt, 1994; Miller et al., 1992). Though models differ in important ways, all of them work by applying rules to a symbolic representation of a rhythm. This strategy raises two questions. First, what rules are to be used? Second, how is a rhythm to be represented? We address these questions in this study by taking share the common strategy of applying rules to a symbolic representation of a rhythm.

Though this study offers no comprehensive review – see Desain and Honing (1999) for such an overview – the work of Longuet-Higgins et al. (Longuet-Higgins & Lee, 1982, 1984) deserves special mention due to its popularity. Their model predicts downbeats by using a set of rules to minimize syncopation. One rule (INITIALIZE) initializes processing by setting a hypothesis that the first note is the downbeat. A second rule (STRETCH) then enforces the assumption that long notes are more salient than short ones. A third rule (UPDATE) finds upbeats. That is, it performs phase adjustment, allowing the model to account for patterns where the first beat is not in fact the downbeat. A fourth rule (CONFLATE) attempts to move up the metrical hierarchy by doubling beat length when possible. A fifth rule (CONFIRM) stops processing. These rules combine to form a rules-based model that succeeds at finding downbeats in patterns. Though it has limitations (for example, it does not deal well with cycled patterns) it has many advantages including that it is a process model.

The model we introduce is a modified version of one taken from another popular rules-based model by Povel and Essens (1985). This model (described in detail below) operates on clocks that sample all phases at low and middle levels of a metrical hierarchy. Preference rules are used to choose the best clock from among the set of clocks. This winning clock predicts downbeat location. Our modification lies in the kind of pattern information used to drive the assignment of downbeats. In the original model, rests are the primary predictor of downbeat location. In our model, perceptually-accented onsets (Povel & Okkerman, 1981) are considered instead. To account for the effects of pattern presentation tempo we will modify the our model to be sensitive to a preferred tempo or *tactus* of 600 ms (Parncutt, 1994; Fraisse, 1982). Though rates as slow as 1000 ms to 1500 ms are sometimes suggested for beat induction, a value near 600 ms worked well for the model in the datasets examined. A linear function is used to implement tempo sensitivity.

Our model is similar in many ways and owes much to the one presented in Parncutt (1994). Both models weigh positive evidence (i.e., onset rather than rests), and both models are sensitive to pattern presentation rate. However, the Parncutt model is much more ambitious than the simple model presented here and contains a large set of initial assumptions yielding a rather complicated algorithm. This is not meant as a criticism of the model but rather a recognition that it is attempting to account for a larger set of behaviors than concern us here. For example, it accounts for phenomenal accents using an exponentially-thresholded function of timing, loudness, timbre and pitch change while the Povel and Essens framework only processes changes in timing with linear functions. The Parncutt model deals with human memory capacity (echoic storage), perceptual centering effects and normalization issues while the Povel and Essens framework ignores these. Even in the case of preferred tempo, the Parncutt model uses a Gaussian function of the logarithm of tempo to predict salience, while we use a linear functions.

Clearly, more complicated models can account for a greater portion of variance in certain datasets. Another example of such a model is found in Van Noorden and Moelants (1999) where a resonance-based model is shown to account for tempo preference. The framework we are using here has weaknesses including an inability to deal with continuous-valued IOIs and an inability to act as a process model. However, this framework suits our goal of exploring the predictive power of a relatively low-complexity model. We use only two initial assumptions. First, we assume that solely accented onsets (Povel & Okkerman, 1981) are used for beat induction. Second, we assume that an ability by musicians to tap near tactus (600 ms) regardless of pattern presentation rate explains the differences between musicians and non-musicians (McAuley & Semple, 1999). From these assumptions we derive a beat induction model NPOS in the framework of the Povel and Essens (1985) rule-based model. The implementation and behavior of this model is explored in the following sections.

3 Details of four rule-based models

All four of the rule-based models considered in this paper use the same underlying algorithm. What differs among them is the nature of evidence they most strongly weigh: positive evidence in the form accented onsets, counter or negative evidence in the form of rests or both. To make clear the evidence most strongly considered, we shall name the models NEG, POS, HYBRID, and NPOS for negative evidence, positive evidence, hybrid evidence and normalized positive evidence. The models are as follows.

1. NEG, the negative-evidence model from Povel and Essens (1985);
2. POS, a simple positive-evidence model from McAuley and Semple (1999);
3. HYBRID, the full hybrid model from McAuley and Semple (1999);
4. NPOS, a normalized positive evidence model.

3.1 Application of preference rules

The first step in simulating a rule-based model involves applying perceptual accents to certain onsets in the pattern. These onsets are located by using the following preference rules (Povel & Essens, 1985; Povel & Okkerman, 1981):

1. An isolated[1] onset is accented;
2. The second of two adjacent onsets is accented;
3. The first and last of three or more adjacent onsets are accented.

The effects of all three of these rules can be seen in Figure 2, bottom, where unaccented onsets are shown as unfilled circles and accented onsets as filled circles. For example, note 1 is due to rule one, note 13 is due to rule two, and note 5 is due to rule three.

Preference rules like these are what distinguish the rule-based model as a generative grammar model in the tradition of Chomsky (1965) and, for music, Lerdahl and Jackendoff (1983). One criticism of rules like these is that they are ad-hoc and make the model untestable. That is, so long as more rules can be added, a generative model can always be modified to account for new data and thus can never be disproved. We are generally in agreement with this criticism. However it should be mentioned that in this case only three rules are used, and that these rules are derived from experiments in human rhythmical behavior. (See Povel and Okkerman (1981)) for a defense of these particular rules, and the chapter on rhythm in Handel (1989) for a general overview of perceptual accenting of rhythmical patterns.

3.2 Clock generation and ranking

Once pattern accents have been assigned using the preference rules, a set of clocks is generated. In general, only clocks of less than half the period of the pattern are generated. (Longer clocks are likely not considered by the listener. See Povel & Essens (1985)). Also, clocks with periods that do not divide evenly into the pattern period are discarded. For any period considered, clocks for all possible phases are generated. This strategy insures that the clocks are hierarchical, that they occupy a low or middle level in the metrical hierarchy, and that they exhaustively search the metrical levels considered. For example, a length-16 pattern yields 7 possible clocks: 4 period-four clocks (at each possible phase), 2 period-two clocks, and 1 clock at period one. Clocks 3,5,6,7 are discarded because they do not divide evenly into 16. Clocks 8 and higher are discarded because they are half the period of the pattern or greater. See Figure 2.

These clocks are then processed by counting how many times each clock coincides with accented onsets, unaccented onsets and rests in the pattern. For example, the clock at period 4, phase 4 (the uppermost clock in Fig. 2 aligns with

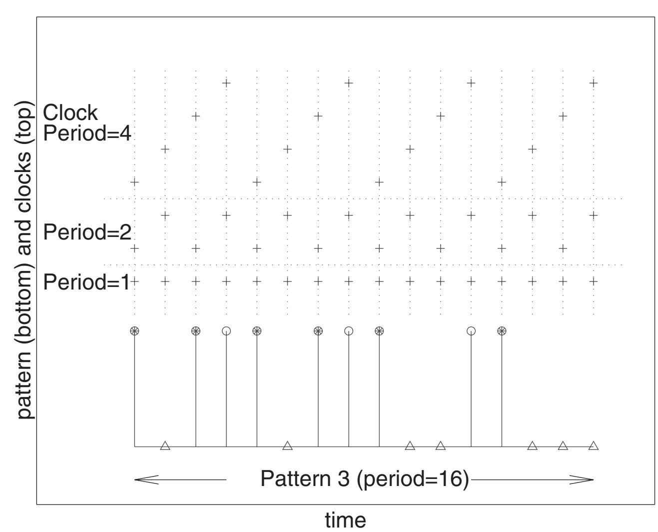

Fig. 2. On the top, all seven clocks for Pattern 3 from Experiment 1 are shown as crosses. On the bottom, the pattern with accents is shown. Rests are shown as triangles, unaccented onsets as raised unfilled circles, and accented onsets as raised filled circles. Accents are assigned using the algorithm from Povel and Okkerman (1981). The horizontal axis denotes discrete time with base IOI undefined.

[1] Note that isolation is arbitrary. For example, the same metronome could consist entirely of isolated onsets or entirely of unisolated onsets depending only upon the size of the grid used. Because this can lead to confusing results, we consider it a (minor) weakness in the formulation.

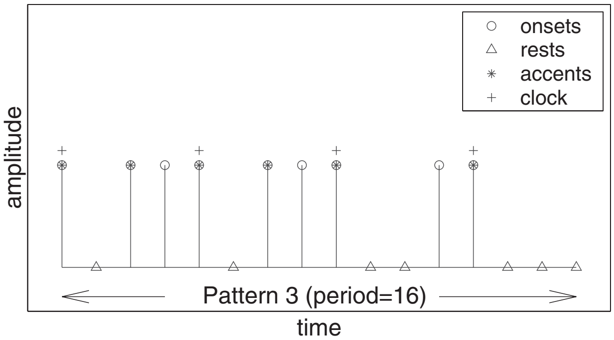

Fig. 3. The best clock as predicted by all four competing models for Pattern 3 from Experiment 1 of Povel and Essens (1985). The best clock is shown above the pattern as a series of crosses. Accents are from Povel and Okkerman (1981).

3 unaccented onsets and 1 rest. This information is used to drive the core matching algorithms of the different models. Though these algorithms differ from model to model, they all fit into the general form

$$S = f(O^+, O^-, R) \tag{1}$$

where S is the score of a particular clock and f is a function of the number of accented onsets O^+, unaccented onsets O^- and rests R to coincide with the clock. As an example of a strongly-induced clock (for all of the models considered here) see Figure 3. Here, the best period-four beat induction for Pattern 3 from Experiment 1 of Povel and Essens (1985) is shown as crosses above the onsets.

3.3 The Povel & Essens negative-evidence model (NEG)

The Povel and Essens (1985) model (NEG) is a negative-evidence model. That is, it ranks potential clocks based on the proportion of relatively unsalient onsets (rests R and unaccented onsets O^-) to coincide with the clock. The model ignores the most salient onsets (accented onsets O^+) found in the pattern. The function f for the NEG model is

$$NEG = wR + O^- \tag{2}$$

where w weighs the relative importance of a rest. In general, $w > 0$. Because the model measures negative-evidence, the higher the NEG score, the lower the induction strength of the clock.

For certain datasets, this model correlates well with pattern reproduction error. However, McAuley and Semple (1999) show that the model does not work as well when tempo variation is considered. Tempo is implicitly encoded in the model via the base inter-onset interval used to encode a pattern (200 ms). That is, by choosing a very fine-grained grid, most or all beats would be isolated and so therefore treated as perceptually salient. However there is no means to control tempo explicitly in the model.

3.4 A simple positive evidence model (POS)

McAuley and Semple (1999) propose a simple positive evidence model (POS). The function f for this positive evidence model is

$$POS = wO^+ + O^- \tag{3}$$

where w weighs the relative importance of an accented onset.

This model is fundamentally flawed because it will invariably give the highest induction score to the *fastest* clock being considered. This is the case because the fastest (generally period 1) clock coincides with more onsets than any of the slower clocks. Because of this failing, the POS model cannot be considered a serious model of pattern complexity. The failure of the POS model illustrates, perhaps, why Povel & Essens chose to use negative evidence. Their NEG model achieves remarkably better performance by simply looking at rests instead of accented onsets. Otherwise the two models are identical.

3.5 The McAuley & Semple hybrid model (HYBRID)

McAuley and Semple (1999) take note of the general ineffectiveness of the POS model. Their main reason for introducing POS is as a building block for a hybrid model (HYBRID). The function f for the HYBRID model is

$$HYBRID = w_{pos}POS - w_{neg}NEG \tag{4}$$

where w_{pos} and w_{neg} are new weights that consider influence of positive evidence and negative-evidence respectively. The two weights w_{pos} and w_{neg} can be collapsed into one free parameter by setting $w_{pos} = 1 - w_{neg}$ where $0 \leq w_{neg} \leq 1$. This model reduces to the NEG model by setting $w_{pos} = 0$ and reduces to the POS model by setting $w_{neg} = 0$.

This model was used successfully by McAuley and Semple (1999) to account for mediating effects of tempo in beat induction for musicians. McAuley and Semple claim that, while non-musicians attend only to negative evidence, musicians attend to positive and negative evidence differently based upon pattern tempo. Specifically they posit a gradual shift for musicians from negative evidence to positive evidence as pattern presentation rate increases. Thus, while they use the NEG model alone for non-musicians, they use all three models for musicians: POS for slow rates, HYBRID for medium rates and NEG for fast rates. This musician strategy can be restated in terms of the HYBRID model alone: as pattern presentation rate goes from slow to fast, w_{neg} goes from 0 to 1 while w_{pos} goes from 1 to 0. We will return to this issue in Section 5.

3.6 A normalized positive-evidence model (NPOS)

Here we introduce the normalized positive evidence model (NPOS) an alternative to the NEG and HYBRID models. This model is similar to a model by Parncutt (1994) that also uses positive evidence. In fact our model could be thought of as a simplification of the Parncutt model, implemented in the framework of a Povel and Essens-style model. However, our model differs in that it is less complex than the Parncutt model and so can be fit into the Povel and Essens framework considered here. (Predictions from the Parncutt model are considered in Section 5). The function f for the NPOS model is

$$NPOS = \Omega O^{+} \tag{5}$$

The model is similar to the POS model except that here the predicted score is normalized by Ω, the period of the clock being matched. This solves the problem with the POS model that the fastest clock will always win. At first glance it may seem that, by adjusting for clock period, NPOS has access to information that the other three models do not. In fact, clock period Ω can by found by simply counting all onsets and rests that the clock encounters. That is, $\Omega = 1/(O^{+} + O^{-} + R)$ meaning that $NPOS = O^{+}/(O^{+} + O^{-} + R)$.

With modification, the NPOS model can be used to account for the interaction between musical training and pattern presentation tempo observed by McAuley and Semple (1999). We believe that all listeners attend primarily to positive evidence; that is, we reject the notion that listeners change the kind of information they attend to based upon pattern presentation rate. We account for differential behavior between musicians and non-musicians by suggesting that musicians have an enhanced ability to tap at or near a comfortable tapping rate regardless of pattern presentation rate. Accordingly, we modify our NPOS model with a term Δ which indicates the degree to which the period of a candidate beat assignment differs from a preferred rate of 600 ms (Fraisse, 1982). For these experiments, $\Delta = abs(\Omega - 600)/4800$. 4800 ms was chosen arbitrarily as an upper bound. Because this choice modifies the slope of the function determining Δ, in theory it could affect the results. However in practice we found that values between 4000 ms and 10000 ms seconds worked well. This yields the equation for TNPOS:

$$TNPOS = NPOS - v\Delta \tag{6}$$

The variable v scales the relative importance of tactus. When v is set to zero, tactus plays no role in clock induction strength and TNPOS reduces to NPOS. When v is large, the induction score for clocks of period 600 ms is the same as that of *NPOS*, while scores for clock periods far from tactus are reduced. TNPOS is only used to model experimental data where presentation rate was manipulated (e.g., Section 5 and Section 4). For data where presentation rate was kept constant, the simpler NPOS is used (e.g., Section 6).

The differences between NPOS and TNPOS are perhaps best understood by considering the Pulse pattern from Parncutt (1994). For this pattern – a simple pulse train having no rests – the NPOS model (weight $v = 0$ in Table 1) predicts equal induction strengths for all clocks regardless of clock period or phase. The TNPOS model, on the other hand, predicts that clocks with periods near a comfortable tapping rate will be preferred. For example, with weight $v = 3$ a clock at 600 ms is almost twice as likely (1.00) as a clock at 1600 ms (0.53).

Table 1. The Pulse pattern (a metronome) with an inter-onset interval of 200 ms is presented to the TNPOS model. As the weight v increases, a clock far from a preferred tapping rate of 600 ms become less likely. Weight $v = 0$ reduces to the NPOS model.

TNPOS strengths for the Pulse pattern					
Clock Ratio	$v = 0$	$v = 1$	$v = 2$	$v = 3$	$v = 4$
1 : 1 (200 ms)	1.00	0.94	0.88	0.81	0.69
1 : 2 (400 ms)	1.00	0.97	0.94	0.91	0.84
1 : 3 (600 ms)	1.00	1.00	1.00	1.00	1.00
1 : 4 (800 ms)	1.00	0.97	0.94	0.91	0.84
1 : 6 (1200 ms)	1.00	0.91	0.81	0.72	0.53
1 : 8 (1600 ms)	1.00	0.84	0.69	0.53	0.22

4 Dataset 1 – McAuley and Semple (1999), Experiment 1

McAuley and Semple (1999) investigate an interaction between musical training and pattern presentation rate in the task of beat induction. 28 listeners were presented with six simple patterns and instructed to tap along regularly. Results for 5 participants were discarded due to equipment failure, leaving 8 male and 15 female subjects. 14 had more than 5 years of formal musical training while the other 9 had no musical training. The authors successfully show that musically-trained subjects have a strong tendency to tap at or near a preferred tapping rate of 600 ms regardless of presentation rate. Untrained subjects do not show this bias and tend instead to match the rate of their beat assignments to the rate of pattern presentation. One likely explanation for this effect is that musically-trained subjects are able to find beat assignments at many levels of the metrical hierarchy while non-musicians find only the most strongly suggested ones, usually at a shallow to middle level of the hierarchy. Therefore, as pattern presentation rate increases musicians (but not non-musicians) are able to continue tapping at a comfortable rate by choosing downbeat assignments deeper in the metrical hierarchy Drake and Botte (1993); Drake et al. (2000). In Figure 4 the differences between musicians and non-musicians in McAuley and Semple (1999), Experiment 1, can be clearly seen. Note that the analysis is of mean tapping rates and so can only be interpreted as showing a general trend towards slower rates for non-musicians.

4.1 Patterns

The six patterns used were taken from Parncutt (1994). They are shown in Table 2. The first four patterns, Pulse, March,

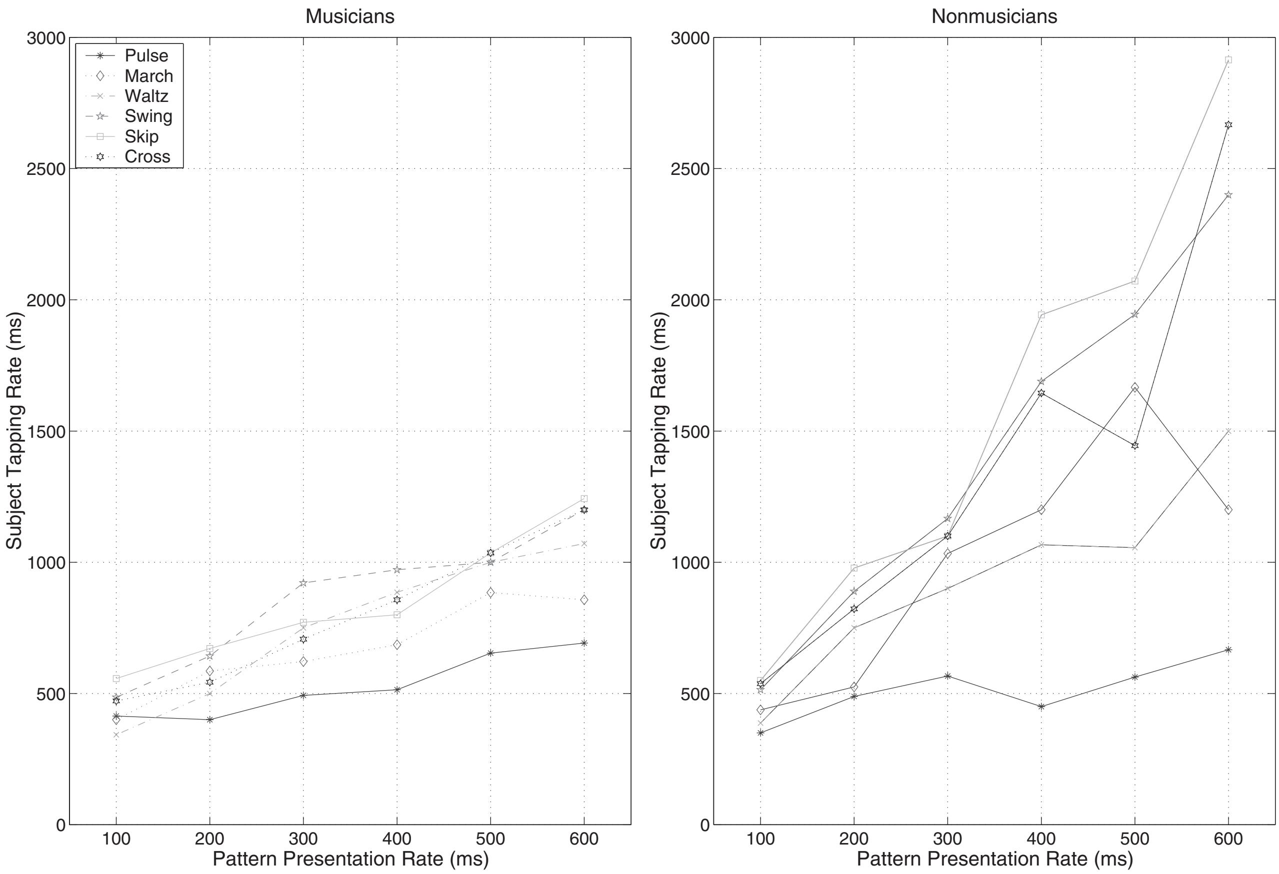

Fig. 4. Data from McAuley and Semple (1999). Musicians (left) versus non-musicians (right) tapping rates. The y-axis shows the mean rate of tapping while the x-axis shows the pattern presentation rate. Notice that musicians show a stronger tendency to tap at near 600 ms regardless of presentation rate. This requires the ability to find downbeat assignments at multiple levels of the metrical hierarchy.

Table 2. The six patterns from Parncutt (1994) also used in McAuley and Semple (1999) Experiment 1. Rests are shown with a "." and onsets are show with an "x".

Six patterns from Parncutt (1994)			
Pulse	xxxxxxxxxxxx	March	x.xxx.xxx.xx
Waltz	x.xx.xx.xx.x	Swing	x..x.xx..x.x
Skip	x..xx.x..xx.	Cross	x.xxx.x.xxx.

Waltz and Swing are relatively simple metrically while the last two are more difficult to categorize. Though the patterns can be economically encoded in patterns of length six or less, they are shown here as length-24 patterns. By coding them at this length, the rule-based models were forced to consider clocks of period 8, 6, 4, 3, 2 and 1. Note that in generating length-24 patterns we did not modify the grid size – that is, we did not stretch the pattern using rests – as doing so would change which onsets are considered isolated by the algorithm. Instead we simply concatenated multiple copies of the patterns to generate length-24 versions.

4.2 Method

The method used here is almost identical to the analysis method used in McAuley and Semple (1999). The six patterns were encoded as inter-onset intervals and simulated on the four rule-based models using a general clock model simulator created for this purpose. The resulting clock induction scores were translated into subject response probabilities. Then root mean squared error (RMSE) values were computed using the response probabilities and the experimental data.

To generate response probabilities from the induction scores, the Luce choice rule (Luce, 1963) was used:

$$P(C) = exp\left(\frac{\zeta M_c}{\sum_{i=1}^{n} exp(\zeta M_i)}\right) \quad (7)$$

where $P(C)$ is the probability that a particular clock interval is chosen, M_c is the induction strength for that clock and ζ is a sharpening factor. In general a high sharpening value ($\zeta >> 0$) results in just a few clocks receiving all of the response probability while a low value spreads the probability among many clocks. Model predictions are generated by scaling the response probabilities $P(C)$ by the number of subjects n. In

McAuley and Semple (1999) the sharpening constant ζ was set to 0.2. For the TNPOS model a sharpening value of 4.0 was used. This was done to adjust for the relatively low-magnitude induction strengths of the TNPOS model. Similar results are had by keeping the sharpening constant $\zeta = 0.2$ and multiplying the TNPOS values by a constant to increase their overall magnitudes.

Different TNPOS v weights were used for musicians and non-musicians. For musicians we chose a fairly high v weight of 5.0. This caused the TNPOS model to strongly prefer a tactus of 600 ms. For non-musicians we set v to 0.0. This reduced the TNPOS to the NPOS model, resulting in no preference for tactus. We also added the musicians and non-musicians together in a merged dataset. For this merged dataset we chose a medium v weight of 3.3. For all simulations.

NEG was simulated with a weight $w = 4.0$, POS with a weight $w = 4.0$ and HYBRID model with weights of $w_{pos} = 0.5$ and $w_{neg} = 0.5$. This is in keeping with the weights used in Povel and Essens (1985) and McAuley and Semple (1999).

Note that we did not search for the best w value for the simulations. (We do vary w_{pos} and w_{neg} values for comparison with TNPOS. See the results in Table 4). Given that we did search for the best TNPOS v setting, this may seem unfair. We believe it is not unfair for several reasons. First, simulation revealed that only very extreme w values had any effect on performance, and those effects were minor. For example, when $w >= 10$ NEG performed slightly better in some cases and when $w = 0$ NEG performed slightly better in other cases. Overall performance is best for $w = 4$ (and for many values near 4.0). Furthermore, for instances where NEG performance increased due to an inflated w value, TNPOS performance also increased with the same value. Simulations aside, we also note that there is no principled reason to vary w based upon the musical experience of subjects. For the TNPOS v, on the other hand, we predict that a high v will best fit musicians and a low v non-musicians. Though we did perform a search to find the optimal v, the search values support our hypothesis. In short, we admit that TNPOS has a free variable that NEG does not, and it should not be surprising that it performs better in instances where that variable is used (Musicians and All Subjects). Though we do not believe optimization of v is unfair, we admit the presence of the variable should be taken into account when comparing the results of the two models. All simulations were done using Matlab 5.3.

4.3 Simulation results

The full set of TNPOS predictions are found on the web at http://www.swets.nl/jnmr/jnmr.html. To save space, the full set of NEG, POS and HYBRID predictions are not reproduced; they can be found in McAuley and Semple (1999). In Table 3 the best model fit is shown for all patterns and all presentation rates. Three tables are shown, detailing model performance for musicians, non-musicians and all subjects. In general, the TNPOS model performed very well. For musicians, TNPOS with weight $v = 5.00$ was the best fit in 26 of 36 cases (72% of the time). For non-musicians, TNPOS with weight $v = 0.00$ (reducing TNPOS to NPOS) was the best fit in 22 of 36 cases (61% of the time). For all subjects, TNPOS with weight $v = 3.30$ was the best fit in 24 of 36 cases (67% of the time).

One problem with displaying only the best fitting model for each case is that it is not possible to see the degree by which a certain model prevailed over the others. To address this we computed normalized root mean squared error (NRMSE) for all patterns and all presentation rates. One property of NRMSE is that a model which only predicts the mean of the data will yield an NRMSE of 1.0. For this reason, NRMSE values near to or greater than 1.0 are considered poor. By taking an average of the 36 NRMSEs from each dataset (6 presentation rates and 6 patterns) we obtain a global metric of model performance. These averages are displayed in Table 4 for all of the rule-based models. To test the hypothesis of McAuley and Semple (1999) the column MIN was introduced. The authors claim that musicians shift from positive evidence to hybrid evidence to negative evidence as tempo increased. To test this, we took the minimum of the error from HYBRID, POS at slow rates (500 ms and 600 ms) and NEG at fast rates (100 ms and 200 ms) and use them to form the column MIN. As can be seen, MIN does indeed outperform NEG, POS or HYBRID alone. However, TNPOS outperformed all other rule-based models and MIN for all three datasets. MIN scores were not included for the non-musicians case because McAuley and Semple (1999) do not hypothesize a similar shift in non-musicians.

4.4 Discussion

The TNPOS model performed better than any of the other rule-based models for all three datasets (musicians, non-musicians and all subjects). These results support our hypothesis that a positive-evidence tactus-sensitive model is the simplest and most effective way to account for these data. However, since the NEG model accounted for some of the non-musician variance and the MIN combination performed adequately well on musicians, we cannot conclusively rule out the explanation of McAuley and Semple (1999). The TNPOS model failed to predict the Pulse pattern and the Waltz pattern for non-musicians. Though we are not concerned with the Pulse pattern because it has no meter, the Waltz pattern should yield reliable downbeat predictions. We have no easy explanation for this result.

As one reviewer noted, perhaps the difference in behavior between musicians and non-musicians is primarily due to a misunderstanding of the task at hand by non-musicians. Non-musicians may, for example, tap slower because it is easier to wait for repetition in the patterns. In this is the case then musicians and non-musicians were in fact performing different tasks, even if the instructions were the same. We agree that this is a reasonable hypothesis; more research in this area is warranted.

Table 3. Model fits for musicians (above), non-musicians (middle) and combined (bottom) in McAuley and Semple (1999) Experiment 1. The TNPOS weight v was set to 5.0 for musicians, reflecting a sensitivity to preferred tempo; it was set to 0.0 for non-musicians, reflecting no sensitivity. For the merged data (all subjects) a weight v of 3.3 was chosen.

Best model fits (Musicians, TNPOS v = 5.00)

	100 ms	200 ms	300 ms	400 ms	500 ms	600 ms
Pulse	NEG	NEG	NPOS	NPOS	NPOS	NPOS
March	NEG	NPOS	HYBRID	NPOS	NPOS	POS
Waltz	HYBRID	HYBRID	NPOS	NPOS	NPOS	NPOS
Swing	NPOS	NPOS	HYBRID	NPOS	NPOS	NPOS
Skip	NPOS	NPOS	NPOS	NPOS	NPOS	NPOS
Cross	NPOS	HYBRID	HYBRID	NPOS	NPOS	NPOS

Summary: NEG = 3 POS = 1 HYBRID = 6 NPOS = 26

Best model fits (Nonmusicians, TNPOS v = 0.00)

	100 ms	200 ms	300 ms	400 ms	500 ms	600 ms
Pulse	NEG	HYBRID	HYBRID	HYBRID	HYBRID	HYBRID
March	NPOS	NPOS	NPOS	NPOS	NPOS	POS
Waltz	HYBRID	HYBRID	HYBRID	HYBRID	HYBRID	HYBRID
Swing	NPOS	NPOS	NPOS	NPOS	NPOS	NPOS
Skip	NPOS	NPOS	NPOS	NPOS	NPOS	NPOS
Cross	NEG	NPOS	NPOS	NPOS	POS	NPOS

Summary: NEG = 2 POS = 2 HYBRID = 11 NPOS = 22

Best model fits (All subjects, TNPOS v = 3.30)

	100 ms	200 ms	300 ms	400 ms	500 ms	600 ms
Pulse	NEG	NEG	NPOS	HYBRID	NPOS	NPOS
March	NEG	NPOS	NPOS	NPOS	NEG	POS
Waltz	HYBRID	HYBRID	NPOS	NPOS	NPOS	NPOS
Swing	NPOS	NPOS	NPOS	NPOS	POS	NPOS
Skip	NPOS	NPOS	NPOS	NPOS	NPOS	HYBRID
Cross	NPOS	HYBRID	NPOS	NPOS	NPOS	POS

Summary: NEG = 4 POS = 3 HYBRID = 5 NPOS = 24

Table 4. Average normalized root mean squared errors (NRMSE) for all patterns and all presentation rates for the four models. Higher values indicate poorer performance. A value of 1.0 indicates that the model is simply predicting the mean. "MIN" combines NEG, POS and HYBRID in keeping with the hypothesis of McAuley and Semple (1999).

Average NRMSE for all models

Data	NEG	POS	HYBRID	MIN	TNPOS
Musicians	1.13	1.18	0.93	0.68	0.52
Non-Musicians	0.92	1.42	1.15	N/A	0.85
All Subjects	1.08	1.39	1.05	0.81	0.72

5 Dataset 2 – Parncutt (1994), Experiment 1

Parncutt (1994) reports on two experiments in beat induction. (In fact, the work was about more than simply finding downbeats; it dealt with the more general issue of pulse salience. However, the experimental results are applicable here.)

In Experiment 1 (similar to the McAuley and Semple (1999) experiment above) listeners are asked to tap along with simple cycled rhythmical patterns. The patterns are presented at six different rates ranging from very fast to very slow (400, 264, 174, 155, 76 and 50 onsets per minute). In Experiment 2, listeners are presented with these same patterns and, in addition, a second periodic sound. They are asked to judge whether this second sound is on or off of the beat. In addition to these two experiments, Parncutt offers a detailed quantitative model. Because the model is structured differently than Povel and Essens-type rule-based models, it is not described here, and readers are referred to Parncutt (1994) for details.

5.1 Patterns

The patterns from Parncutt (1994) were also used by McAuley and Semple (1999) in Experiment 1. These patterns are discussed in Section 5.1 and show in in Table 2. They are Pulse, March, Waltz, Swing, Skip and Cross.

5.2 Method

The method here was identical to the method used to analyze the McAuley and Semple (1999) data in Section 4. The six patterns were encoded as inter-onset intervals and simulated on the four rule-based models using a general clock model simulator created for this purpose. The resulting clock induction scores were translated into subject response probabilities. Then RMSE values were computed using the response probabilities and the experimental data.

The best fit to the data was achieved using an TNPOS tactus weight v of 5.3, indicating a fairly strong preference for a 600 ms tactus. As the subjects in the experiment had a mean of 12 years of musical training, this v weight is consistent with our hypothesis that musical experience allows musicians to remain close to a preferred tapping rate regardless of pattern presentation rate. Since the data in Parncutt (1994) was not categorized by musical experience, musicians and non-musicians could not be treated separately, making it impossible to test this hypothesis further.

The Parncutt (1994) Experiment 1 data included period and phase information. That is, the tapping rate of the subject was recorded, along with the pattern-relative position of the tapping. For these analyses, the phase information was discarded by summing responses across phases for each period. This transformed the data into the same format as that of McAuley and Semple (1999) Experiment 1.

The other rule-based model weights were as follows: the NEG model was simulated with a weight $w = 4.0$, the POS model with weight $w = 4.0$ and the HYBRID model with weight $w_{pos} = 0.5$ and weight $w_{neg} = 0.5$. As a comparison, the predictions of the Parncutt (1994) model ("PNCT") were also used. This model was not simulated computationally; instead prediction values were taken directly from the paper. All simulations were done using Matlab 5.3

As in Dataset 1, we did not search for optimal w values. See our comment in Section 4.2.

5.3 Simulation results

The full set of TNPOS predictions are on the web at http://www.swets.nl/jnmr/jnmr.html. The full set of NEG, POS and HYBRID predictions are found in McAuley and Semple (1999). In Table 5 the best model fit is shown for all patterns and all presentation rates. In general, the TNPOS model performed very well. It outperformed POS, HYBRID and TNPOS in 28 of 36 cases (78% of the time). Even when the Parncutt model predictions are considered, TNPOS works adequately well with 16 of 36 cases (44%) compared to PNCT's 14 of 36 cases (39%). The dataset is too small to support the claim that TNPOS is *better* than PNCT. Given that the Parncutt model is much more sophisticated, it is likely to generate better predictions for large datasets.

To see global performance, we computed the average normalized root mean square error (NRMSE) for all patterns and all presentation rates (Table 6). Higher values indicate poorer performance. By normalizing the RMSE scores, the NRMSE values are calibrated such that a value of 1.0 indicates a prediction of the mean. Thus, values greater than 1.0 reveal particularly poor performance. To test the hypothesis of McAuley and Semple (1999) the column MIN was introduced, which combined NEG, POS and HYBRID evidence in keeping with the claim of McAuley and Semple (1999) that, as tempo increases, musically-trained listeners systematically shift from positive to hybrid to negative evidence. With an average NRMSE value of 0.48, the TNPOS performed better than all other models, including PNCT and MIN.

5.4 Discussion

Given that the PNCT model relies on four free parameters (Parncutt, 1994) while TNPOS contains one free parameter, this is a promising result for TNPOS. Note, however, that we are not predicting the phase of downbeat location, only the period. It remains untested whether TNPOS can match the performance of PNCT at the more difficult task of predicting both the periods and phases of downbeats.

With performance better than any of the other rule-based models, these simulations support the claim that a positive-evidence tactus-sensitive account of the data is the correct one. However, as was the case in the McAuley and Semple (1999) simulations, the MIN model performed adequately well (in fact, better than the PNCT model). Thus, we conclude that these results are not wholly conclusive and that further research is warranted.

6 Dataset 3 – Povel and Essens (1985), Experiment 1

In Experiment 1 of Povel and Essens (1985), 35 patterns were presented to listeners in a pattern reproduction task. The method was as follows: a pattern was presented via headphones until a subject indicated familiarity with the pattern by pressing a button. Then the subject attempted to reproduce four periods of the sequence they had just heard. Mean reproduction errors were collected (Fig. 5).

Table 5. Model predictions for subjects in Parncutt (1994) Experiment 1. The top table takes into account NEG, POS, HYBRID and TNPOS predictions. The bottom table adds the predictions from the Parncutt model (PNCT). Tempo values are in events per minute (epm) rather than ms. See (Parncutt, 1994) for details. The TNPOS weight $v = 5.3$ indicating a fairly strong preference for tactus.

Best model fits (All subjects, TNPOS $v = 5.30$)

	400 epm	264 epm	174 epm	115 epm	76 epm	50 epm
Pulse	NPOS	NPOS	NPOS	NPOS	NPOS	POS
March	NEG	NPOS	NPOS	NPOS	NPOS	POS
Waltz	HYBRID	HYBRID	NPOS	NPOS	NPOS	NPOS
Swing	NPOS	NPOS	NPOS	HYBRID	NPOS	NPOS
Skip	NPOS	NPOS	NEG	NPOS	NPOS	NPOS
Cross	NPOS	NPOS	HYBRID	NPOS	NPOS	NPOS

Summary: NEG = 2 POS = 2 HYBRID = 4 NPOS = 28

Best model fits (All subjects, TNPOS $v = 5.30$)

	400 epm	264 epm	174 epm	115 epm	76 epm	50 epm
Pulse	NPOS	PNCT	PNCT	NPOS	NPOS	PNCT
March	PNCT	PNCT	PNCT	PNCT	PNCT	POS
Waltz	HYBRID	HYBRID	NPOS	PNCT	NPOS	NPOS
Swing	NPOS	NPOS	NPOS	HYBRID	NPOS	PNCT
Skip	NPOS	NPOS	NEG	NPOS	PNCT	PNCT
Cross	PNCT	PNCT	HYBRID	NPOS	NPOS	NPOS

Summary: NEG = 1 POS = 1 HYBRID = 4 NPOS = 16 PNCT = 14

Table 6. Average normalized root mean squared errors (NRMSE) across all patterns and all presentation rates for the four models. Higher values indicate poorer performance. A value of 1.0 indicates that the model simply predicted the mean. "MIN" combines NEG, POS and HYBRID in keeping with the hypothesis of McAuley and Semple (1999).

Average NRMSE for all models

Data	NEG	POS	HYBRID	PNCT	MIN	TNPOS
All Subjects	0.95	1.25	0.92	0.62	0.60	0.48

The authors argue that their NEG rule-based model accounts for these errors and attribute listener behavior to an internal clock mechanism. In this section we will compare the NEG model to other candidate models using the same data. We will argue that our NPOS model also accounts for the pattern reproduction error but with a smaller set of categories. Furthermore we argue that our model is simpler because it only relies on particularly salient positive evidence (perceptually-accented onsets) and so requires no free parameter to modulate the relative effect of accented and unaccented onsets.

6.1 Patterns

Each of the 35 patterns consists of 16-elements: nine onsets and seven rests. All patterns are permutations of the interval sequence **1 1 1 1 1 2 2 3 4**. The intervals 1, 2, 3 and 4 were presented to listeners as respectively, of 200 ms, 400 ms, 600 ms and 800 ms. The patterns are shown in Table 7.

6.2 Method

The 35 patterns were encoded as inter-onset intervals and simulated on the four rule-based models using a general clock model simulator created for this purpose. The weights were set as follows: NEG weight $w = 4.0$, POS weight $w = 4.0$, HYBRID weights $w_{neg} = 0.5$ and $w_{pos} = 0.5$. As tempo was not a factor in this experiment, the NPOS model was used in place of the TNPOS model. NPOS has no adjustable weights.

Two statistical analyses were performed. First, four linear regressions were performed (one for each rule-based model) that regressed mean deviation errors onto induction strengths. In all cases, the strongest induction strength (for a given model) was used to predict the complexity of the pattern. For these regressions, neither the period nor the phase of the strongest induction strength mattered; only the magnitude of the score was taken into account.

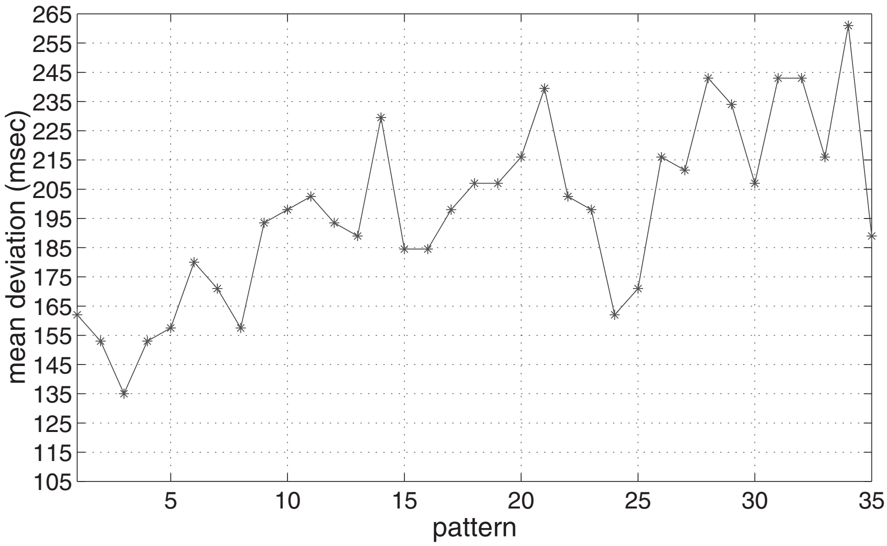

Fig. 5. Mean deviation errors are shown for all 35 patterns from Experiment 1 of Povel and Essens (1985). The patterns are ordered from 1 to 35 using the NEG model. Since this model does a good job of predicting pattern complexity, this results in a general upward linear trend. The actual data points are mean deviation in milliseconds from the target pattern as measured during the reproduction phase of the experiment.

Table 7. The 35 patterns from Povel and Essens (1985) Experiment 1. Rests are shown with a "." and onsets are show with an "x". The table is ordered left to right, top to bottom.

35 patterns from Povel and Essens (1985), Experiment 1			
1 xxxxx..xx.x.x...	2 xxx.x.xxx..xx...	3 x.xxx.xxx..xx...	4 x.x.xxxxx..xx...
5 x..xx.x.xxxxx...	6 xxx.xxx.xx..x...	7 x.xxxx.xx..xx...	8 xx..xxxxx.x.x...
9 xx..x.xxx.xxx...	10 x.xxx.xxxx..x...	11 xxx.xx..xx.xx...	12 xx.xxxx.x..xx...
13 xx.xx.xxxx..x...	14 xx..xx.xx.xxx...	15 x..xxx.xxx.xx...	16 xx.xxxx.xx..x...
17 xx.xxx.xxx..x...	18 xx.xxx..xx.xx...	19 xx..xx.xxxx.x...	20 xx..xx.xxx.xx...
21 xxxxx.xx.x..x...	22 xxxx.x..xxx.x...	23 xxx..xx.xxx.x...	24 x.xxx..x.xxxx...
25 x.x..xxxx.xxx...	26 xxxx.x.x..xxx...	27 xx.xxx.x..xxx...	28 xx.x..xxx.xxx...
29 x.xxxx.x..xxx...	30 x..xxxxx.xx.x...	31 xxxx.xxx..x.x...	32 xxxx..xx.xx.x...
33 xx.xxxx..xx.x...	34 xx.x..xxxxx.x...	35 x.x..xxx.xxxx...	

Each of the models generated a fixed number of complexity categories for the 35 patterns. For example, the NEG model categorized all patterns as having complexity 1, 2, 3, 4, 5, 6 or 7. The linear regression is not sensitive to the number of categories generated by the rule-based models. However, we were interested in knowing whether any of the models succeeded at predicting the data at the expense of a large number of statistically-unreliable categories. Accordingly, we performed an analysis of variance (ANOVA) to test validity of the categories. All simulations were done using Matlab 5.3.

6.3 Simulation results

6.3.1 Linear regression analysis

All four models were fit to the data using linear regression. The analyses revealed that all models except for the POS model are able to account reliably for roughly half of the reproduction error ($0.50 \leq R^2 \leq 0.58$, $p << 0.05$). For the successful models the R^2 values were very close, with a NEG $R^2 = 0.56$, a HYBRID $R^2 = 0.58$ and an NPOS $R^2 = 0.50$. The POS model failed completely, with $R^2 = 0.01$ and $p = 0.67$. In Figure 6 the regression lines are plotted with the experimental data for all four models. On the x-axis of all four plots are the categories generated by the models. The y-axis indicates mean deviation in ms. The straight lines in the plots indicate the prediction made by the linear regression.

6.3.2 Analysis of variance (ANOVA)

Though NEG, HYBRID and NPOS all return good linear regression results, they achieve success with different numbers of categories (NEG = 7, HYBRID = 5, NPOS = 3). This

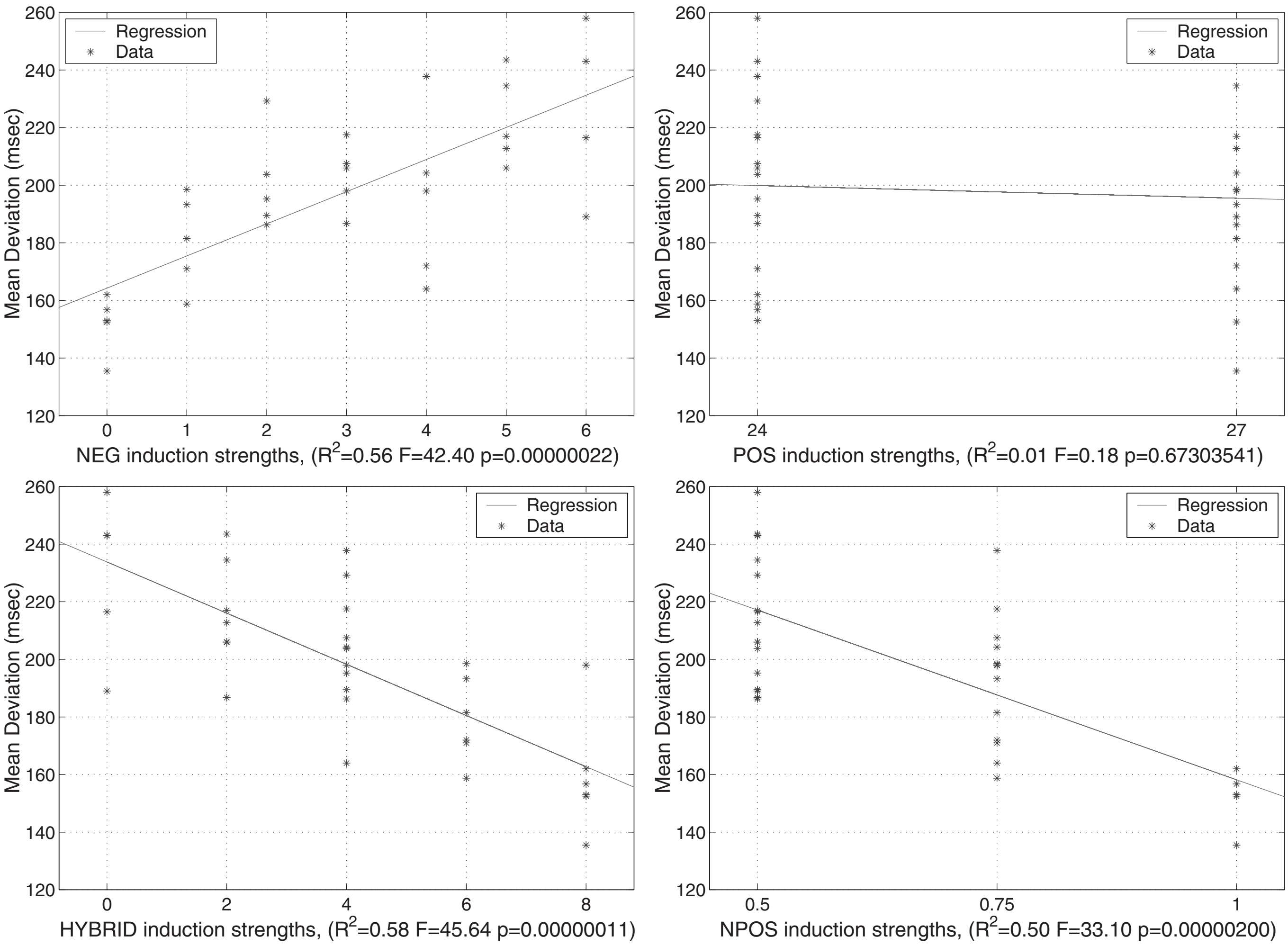

Fig. 6. The regression results are shown for all four clock models. To form the regression line, the data from Experiment 1 of Povel and Essens (1985) were regressed onto the clock strengths. The reversed slope for the NEG model is due to the fact that a high NEG value indicates a low clock induction strength.

raises the question of whether any or all of the models succeed at the expense of having a large number of ineffective categories. To test the categorization abilities of the models, an ANOVA was performed.

All models returned statistically significant results for the ANOVA with $p << 0.05$. However, the sensitivity of the ANOVA brought about a reordering of the models in terms of effectiveness. While the linear regression R^2 values suggest that the best model is HYBRID, the ANOVA suggests that the best model is NEG. More specifically, if we order the models by how much variance they account for (R^2) we get: HYBRID (0.58), NEG (0.56), NPOS (0.50) and POS (0.01). But if we order them by how well the categorize (ANOVA F) we get: NPOS (17.23), HYBRID (10.93), NEG (9.06), POS (0.18). The ANOVA tables are shown in Table 8. In short, NPOS accounts for slightly less variance than NEG or HYBRID but it uses only three categories to achieve its success.

7 Discussion

Because NPOS has no free weights, it is not surprising that it accounts for less variance and generates fewer categories than does NEG, which has a single adjustable weight w, or HYBRID which uses two adjustable weights w_{neg} and $wpos$ (collapsible into a single weight) to modulate the effects of NEG and POS, each in turn having adjustable weights. What may be surprising is that NPOS did as well as it did, accounting for 50% of the variance with three robust categories.

7.1 The role of rule-based models

Can a simple set of rules be used to find downbeats in patterns? More specifically, can a grammar-based algorithm succeed at beat induction? We believe the short answer is, Yes, but only to a limited extent. Many factors influence the induction of downbeats in musical patterns, some of which

Table 8. The ANOVA results are shown for all four clock models. Except for POS, all four models contained statistically-relevant categories.

NEG Catagories (p = 0.000015)					POS Catagories (p = 0.673035)				
Source	SS	df	MS	F	Source	SS	df	MS	F
Columns	20476.7	6	3412.8	9.06	Columns	169.5	1	169.5	0.18
Error	10543.0	28	376.5		Error	30850.3	33	934.9	
Total	31019.7	34			Total	31019.7	34		
HYBRID Catagories (p = 0.000014)					**NPOS Catagories (p = 0.000008)**				
Source	SS	df	MS	F	Source	SS	df	MS	F
Columns	18399.4	4	4599.9	10.93	Columns	16084.3	2	8042.2	17.23
Error	12620.3	30	420.7		Error	14935.4	32	466.7	
Total	31019.7	34			Total	31019.7	34		

are not considered by existing rule-based models (Povel & Essens, 1985; McAuley & Semple, 1999). For example, melodic contour and differences in note amplitudes (neither of which are shown to the models mentioned above) play a role in the metrical interpretation of a piece. See Handel (1989); Fraisse (1982); Cooper and Meyer (1960) for reviews. Even if all musical information like pitch contour and amplitude is made available, listeners with differing amounts of musical experience make different downbeat assignments for the same music (McAuley & Semple, 1999). Thus, as many have observed (Longuet-Higgins & Lee, 1982, 1984; Povel & Essens, 1985), beat induction is not solely a property of the rhythmical pattern but rather of pattern and listener together.

More importantly, the rule-based models explored here (though not some others, such as the Longuet-Higgins models discussed above) are not process models and so are unable to make ongoing, developing predictions during pattern presentation. This poses a problem for a system which must respond to an unfolding pattern, such as a robot or online speech recognizer. It also makes it difficult for rule-based models to account for phenomena where global pattern structure is insufficient to make valid predictions. For example, listeners are highly likely to treat the first onset they hear as the downbeat, regardless of global pattern structure (Garner & Gottwald, 1968; Longuet-Higgins & Lee, 1982). Though a rule-based model could perhaps account for this behavior by placing special markers on the first onset or onsets, the requirement for global access to an entire pattern renders the model insensitive to non-stationary temporal pattern structure in general. For these reasons it seems unlikely that a simple grammatical approach will capture all of the important details of beat induction, and we agree with Large and Jones (1999) that a process model such as a beat-inducing nonlinear oscillator (Large & Kolen, 1994; McAuley, 1994; Gasser et al., 1999) is a more appropriate mechanism. See Eck (2001a,b) for our recent attempts.

8 Future research

8.1 The relative importance of unaccented and accented onsets

One important way in with NPOS/TNPOS differs from NEG, POS and HYBRID is that it pays no attention to unaccented onsets. This seems to us to be counter-intuitive. Unaccented onsets have a physical reality, and so it is reasonable to assume that they would be processed differently than rests. Thus, we were surprised that TNPOS did as well as it did. For the McAuley and Semple (1999) and the Parncutt (1994) datasets, adding unaccented onsets to a modified TNPOS model did not help at all. In fact, for low TNPOS w settings – adding more salience to unaccented onsets – performance was greatly diminished. This suggests that, for beat induction tasks, unaccented onsets are irrelevant or at least very unimportant. We believe that more research in this area is called for.

8.2 The relative importance of early versus late accented onsets

In Eck (2001b) we observed that particular oscillator models exhibit differential sensitivity to early and late accented onsets. By “early” and “late” we simply mean first and last in a series of unisolated events. This distinction is wholly phrasal: we do not mean early or late in the pattern as a whole, or early versus late with respect to the bar. Integrate-and-fire relaxation oscillators are particularly sensitive to single onsets and to early onsets in a series of unisolated onsets, but they are not as sensitive to late onsets in series. This is in conflict with the preference rules used in the models here (Povel & Okkerman, 1981), where all accented onsets (including the last of an uninterrupted series) are weighted equally. We recommend a study which focuses on "early versus late" by using a set of patterns designed to focus on this issue. Insights might also come from looking at appropriate rules for modeling this behavior and from expectancy-based theories such as (Desain, 1992), which deal explicitly with how past experience with some temporal signal determines prediction making.

9 Conclusions

The NEG, HYBRID and NPOS/TNPOS models can all account for the variance in the datasets they were designed to handle. We believe, however that NPOS/TNPOS is preferable among the three due to its simplicity and effectiveness. For the beat assignment datasets (McAuley & Semple, 1999; Parncutt, 1994) the TNPOS accounted for more variance than any of the other models. It also dealt well with the interaction between musical experience and pattern presentation rate. For the pattern memorization and reproduction dataset (Povel & Essens, 1985) the NPOS model performed nearly as well as the other models despite having no free parameters. In our view, the positive evidence hypothesis underlying the NPOS/TNPOS model is the simplest one considered here. Though these datasets are too incomplete to yield conclusive results, we believe that the simulations included here provide convincing evidence in support of the NPOS/TNPOS model.

References

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press, Cambridge, Mass.

Cooper, G. & Meyer, L.B. (1960). *The Rhythmic Structure of Music*. The University of Chicago Press.

Cummins, F. & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*, 145–171.

Desain, P. (1992). A (de)composable theory of rhythm perception. *Music Perception*, *9*, 439–454.

Desain, P. & Honing, H. (1999). Computational models of best induction: The rule-based approach. *Journal of New Music Research*, *28*, 29–42.

Drake, C. & Botte, M. (1993). Tempo sensitivity in auditory sequences: Evidence for the multiplelook model. *Perception and Psychophysics*, *54*, 277–286.

Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception*, *18*, 1–23.

Eck, D. (2001a). A network of relaxation oscillators that finds downbeats in rhythms. In *Proc. ICANN 2001, Int. Conf. on Artificial Neural Networks*, Vienna, Austria. IEE, London. Accepted.

Eck, D. (2001b). Finding downbeats with a relaxation oscillator. *Psychological Review* (in press).

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 149–180). Academic Press, Inc., New York.

Garner, W. & Gottwald, R. (1968). The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology*, *20*, 97–109.

Gasser, M., Eck, D., & Port, R. (1999). Meter as mechanism: A neural network model that learns metrical patterns. *Connection Science*, 11, 187–216.

Haken, H., Kelso, J., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347–356.

Handel, S. (1984). Using polyrhythms to study rhythm. *Music Perception*, *1*, 465–484.

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

Jones, M. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*, 323–335.

Jones, M.R. & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*, 459–491.

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

Large, E.W. & Kolen, J.F. (1994). Resonance and the perception of musical meter. *Connection Science*, *6*, 177–208.

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press, Cambridge, Mass.

Longuet-Higgins, H. & Lee, C. (1982). The perception of musical rhythms. *Perception*, *11*, 115–128.

Longuet-Higgins, H. & Lee, C. (1984). The rhythmic interpretation of monophonic music. *Music Perception*, *1*, 424–441.

Luce, R. (1963). Detection and recognition. In R. Luce & R. Bush (Eds.), *Handbook of Mathematical Psychology*. Wiley.

McAuley, J.D. (1994). Finding metrical structure in time. In M. Mozer, P. Smolensky, D. Touretsky, J. Elman, & A.S. Weigend (Eds.), *Proceedings of the 1993 Connectionist Models Summer School* (pp. 219–227). Hillsdale, NJ. Erlbaum.

McAuley, J.D. & Semple, P. (1999). The effect of tempo and musical experience on perceived beat. *Australian Journal of Psychology*, *51*, 176–187.

Miller, B.O., Scarborough, D.L., & Jones, J.A. (1992). On the perception of meter. In M. Balaban, K. Ebcioğlu, & O. Laske (Eds.), *Understanding Music with AI: Perspectives on Music Cognition* (pp. 429–447). MIT Press, Cambridge, Mass.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, *11*, 409–464.

Povel, D. & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*, 411–440.

Povel, D. & Okkerman, H. (1981). Accents in equitone sequences. *Perception and Psychophysics*, *30*, 565–572.

Van Noorden, L. & Moelants, D. (1999). Resonance in the perception of musical tempo. *Journal of New Music Research*, *28*, 43–66.