

This article was downloaded by:[EPSCoR Science Information Group (ESIG) Dekker Titles only Consortium]
On: 12 September 2007
Access Details: [subscription number 777703943]
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of New Music Research

Publication details, including instructions for authors and subscription information:
<http://www.informaworld.com/smpp/title~content=t713817838>

Computational Models of Beat Induction: The Rule-Based Approach

Peter Desain; Henkjan Honing

Online Publication Date: 01 March 1999

To cite this Article: Desain, Peter and Honing, Henkjan (1999) 'Computational Models of Beat Induction: The Rule-Based Approach', Journal of New Music Research, 28:1, 29 - 42

To link to this article: DOI: 10.1076/jnmr.28.1.29.3123

URL: <http://dx.doi.org/10.1076/jnmr.28.1.29.3123>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

COMPUTATIONAL MODELS OF BEAT INDUCTION: THE RULE-BASED APPROACH*

Peter Desain¹ and Henkjan Honing^{1,2}

¹NICI, Nijmegen University, Nijmegen, The Netherlands

²Music Department, University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT

This paper is a report of ongoing research on the computational modeling of beat induction which aims at achieving a better understanding of the perceptual processes involved by ordering and reformulating existing models. One family of rule-based beat induction models is described (Longuet-Higgins and Lee, 1982; Lee, 1985; Longuet-Higgins, 1994), along with the presentation of analysis methods that allow an evaluation of the models in terms of their in- and output spaces, abstracting from internal detail. It builds on work described in (Desain and Honing, 1994b). The present paper elaborates these methods and presents the results obtained. It will be shown that they can be used to characterize the differences between these models, a point that was difficult to assess previously. Furthermore, the first results of using the method to improve the existing rule-based models are presented, by describing the most effective version of a specific rule, and the most effective parameter settings.

INTRODUCTION

Beat induction is the process in which a regular isochronous pattern (the beat) is activated while listening to music. This beat, often tapped along by

musicians, is a central issue in time keeping in music performance. But also for non-experts the process seems to be fundamental to the processing, coding and appreciation of temporal patterns. The induced beat carries the perception of tempo and is the basis of temporal coding of temporal patterns. Furthermore, it determines the relative importance of notes in, for example, the melodic and harmonic structure.

There are a number of aspects that make beat induction a process that is hard to model computationally. Beat induction is a fast process. Only after a few notes (5-10) a strong sense of beat can be induced (a “bottom-up” process). Once a beat is induced by the incoming material it sets up a persistent mental framework that guides the perception of new incoming material (a “top-down” process). This process, for example, facilitates the percept of syncopation, i.e., to “hear” a beat that is not carried by an event. However, this top-down processing is not rigidly adhering to a once established beat-percept, because, when in a change of meter the evidence for the old percept becomes too meager, a new beat interpretation is induced. This duality,

Correspondence: Peter Desain, NICI, Nijmegen University, P.O. Box 9104, NL-6500 HE Nijmegen, The Netherlands.

*This paper is a revised version of the paper which appeared in the worknotes of the AI & Music workshop, IJCAI, Montreal (Desain & Honing, 1995).

where a model needs to be able to infer a beat from scratch, but also to let an already induced beat percept guide the organization of more incoming material, is hard to model. This might be an explanation for the wide variety of computational formalisms that have been used to capture the process. Next to rule-based and symbolic search models, optimization, neural nets, and coupled oscillator systems have been used extensively (see Desain and Honing, 1994a for an overview of these models). This diversity makes it difficult to compare and evaluate them. Another problem is that the models implicitly address different aspects of the beat-induction process. For instance, some models explain the formation of a beat concept in the first moments of hearing a rhythmical pattern (initial beat induction), some model the tracking of the tempo once a beat is given, and others cover beat induction for cyclic patterns only.

This paper is part of a larger study that aims at achieving a better understanding of the beat induction process by ordering and reformulating the different models and the subprocesses involved. We restrict ourselves here to presenting the analysis of the family of rule-based models of initial beat induction.

RULE-BASED MODELS

Although symbolic rule-based models are not much *en vogue* anymore, rule-based models for initial beat induction pioneered the field of computational modeling of rhythm perception and perform amazingly well. Longuet-Higgins & Lee (1982) propose a rule-based model of beat induction that was unique at the time, because of its incremental nature and its focus on the initial stages of beat induction. In this paper we will compare the Longuet-Higgins & Lee (1982) model to two recent refinements of the original, i.e., Lee (1985) and Longuet-Higgins (1994). They will be referred to as LHL82, L85, and LH94. Related rule-based models are described in Lee (1991) and Scarborough, Miller & Jones (1992), but they will not be described in this paper since they are models of meter induction. However, the first an extension of L85, the second is an extension of LHL82.

All three models take note duration values as input (expressed as integral multiples of a 16th-note) rather than, for example, attempting to identify the note onsets in an expressive real-time performance. They initially assume the beat to be equal to the time interval between the first two onsets, and then work their way through the incoming material, shifting, doubling and stretching the beat. Each model postulates a state variable (the current beat hypothesis) and a small set of rules (test-action pairs) in which the test consists of a predicate on the rhythmic pattern and the current beat hypothesis, and the action modifies this beat hypothesis.

An example

As a concrete example, consider the musical fragment in Figure 1 showing a trace for a specific rhythmical pattern (3 1 6 2 3 1 6 2 3 1) for the LHL82 model. Time is read from left to right in discrete time steps, and from top to bottom in computation steps. The top line shows the input pattern in a time grid notation (with each “|” marking a note onset). The LHL82 model consists of only five rules: INITIALIZE, STRETCH, UPDATE, CONFLATE, and CONFIRM. For the pattern in Figure 1 the INITIALIZE rule makes the beat equal to the first note. Then the STRETCH rule recognizes a note (i.e., the third) that is longer than the note beginning on the end of the beat and extends the beat such that it coincides with the beginning of that note. The UPDATE rule is the next to fire, since that same note is even longer than the beat. This rule shifts the beat to the beginning of that long note. Because at the end of the

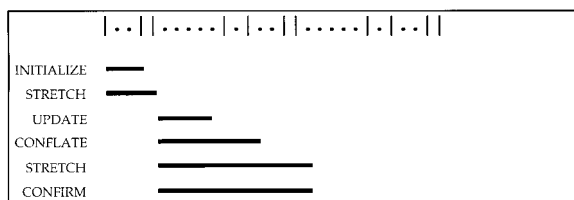


Fig. 1. A computation trace of the processing of the pattern (3 1 6 2 3 1 6 2 3 1) by the LHL82 model, showing subsequent modifications of the beat hypothesis (Musical time from left to right, computation steps from top to bottom).

next beat there is a note, the CONFLATE rule will fire, making the beat twice as long. Then, once more, the STRETCH rule fires and makes the beat so long that the CONFIRM rule stops further processing. The resulting beat for the pattern in Figure 1 is 12 time units long and 4 time units shifted with regard to input (an upbeat), the first beat being on the third, long note.

Shared framework

All three theories make use of the same notion of a current beat hypothesis and a set of rules that changes it. Many temporal patterns are treated differently by the three rule-based models and yield a different beat. Accordingly, different assertions about the state maintained during processing can be made for the different models. Note that in these programs some rules have the same name (e.g., UPDATE) but a different definition.

For LH94 the beat is always equal to or larger than the longest note in the pattern, while this is not necessarily the case for L85 and LHL82. For LHL82 and L85 the beat always grows (or stays the same duration) during processing, while for LH94 the beat sometimes can become smaller. For LH94 the end of the current beat hypothesis is always on a note, while this is not necessarily the case for L85 and LHL82.

It is hard to get firm conclusions about the behavior of the models by studying the detailed workings of the rules on a small set of musical examples, since the interaction between the rules is quite complex. A formal analysis of these models in the form of assertions and invariants, can be given once the models are sufficiently formalized. This was not as strait forward as expected, because of the present state of the models.

Status of the theories

LHL82 describes a beat induction theory with a collection of musical examples and computation traces, along with a clear description of the rule-set that made up the original program. Some rules were not described in a formal way, and interactions between the rules were not made explicit, and therefore had to be rationally reconstructed. The original program that was used to generate the output was not available anymore.

L85 describes a “paper and pencil” model of beat induction –it was never implemented. Several unformalized aspects, as well as interactions between the rules unforeseen by the author (Lee, personal communication) had to be filled-in, to be able to produce an implementation that could replicate the examples given in L85. Its informal presentation has also led to different implementations that give different results (see e.g., Essens, 1995, for an alternative interpretation).

LH94 is a refinement of LHL82, in the sense that some rules were combined and unformalized parts were made explicit. A small computer program in POP-11, describing the model, was made available by its author. The (modified) theory behind the program is not yet published.

Time scale of the input representation

Although at first impression, one would define the time scale used in these models as a discretized time grid in which each time interval is expressed as a multiple of a short time quantum, on closer inspection one discovers that the actual models do not rely on such a quantum. They never use the granularity of the time grid, but only require exact arithmetic for calculating if a note happens on a certain position in time and for deciding if a certain time duration is longer than another one. Without any change in the formalism the model could deal with, for instance, all times expressed as relative to the first note duration. In that sense their behavior is independent of global tempo. However, the parameters of the models, that are expressed in the arbitrary units of the time grid, control, for example, if a beat is long enough to be accepted. Furthermore, because in all the examples given in LHL82 the rhythms are represented on a time-grid with sixteenth note as time unit, the parameters can be assumed as well to be represented on a scale of score note durations (in quarter notes).

There are several difficulties that arise when interpreting these rule-based theories as models for beat induction. A first question that arises is, when is a beat a proper beat, and when just an intermediate state of computation? LH94 makes this explicit and makes a distinction between unconfirmed beats (i.e., an ongoing, yet incomplete state) and confirmed beats (see Fig. 1, last line).

After a beat is confirmed the processing stops. For LHL82 in certain states no other rule can ever fire anymore (i.e., the model becomes “deaf”): an implicit confirmation takes place. L85 is a special case, in the sense that it keeps processing its input, the SHIFT rule can keep moving the processing window through the material. Furthermore, this model has the somewhat awkward characteristic – from a perceptual standpoint –, that the UPDATE rule sometimes has to wait forever until it can execute its action (e.g., the pattern 2 3 2 2 2...).

Control structure and formalization

A complete and comparable formalization of the models has to take the control structure of the rule-based systems into account as well. The order and moment at which the rules fire is crucial, these issues are often left undiscussed in the presentation of the original models.

LHL82 is presented in a window-based way, each rule may look for occurrences of its trigger pattern somewhere in the range of the current beat hypothesis. The rules are executed in a specific order until one can fire, this rule then performs its associated action (i.e., change the phase and/or duration of the beat), after which the next rule in the series is allowed to fire on an updated window. When no rule can fire anymore on the window the model stops further processing.

In LH94 the processing was implemented in an event-based way, with each note onset in the input constituting an event. For each event all rules are given the chance to fire in a fixed order. When no rule fires a next event is processed. The CONFIRM rule explicitly stops the processing of input.

L85 is also described as an window based model, with all rules given a chance to fire before the window is shifted by the current beat duration by the SHIFT rule.

It is not a trivial task to make assertions and proof invariance over these different interpretations of the model’s control structure. However, it turned out that all three models can be formalized such that they can operate in either window-based or event-based mode, as well as in a grid-based mode. The latter mode makes clear how early certain decisions can be made (e.g., is the present note longer than the beat?), an issue that becomes

important in real-time applications and in predicting the times at which changes in beat response can be made. Furthermore, the control structure can be adapted such that for each processing step (be it window-based, event-based or grid-based) only one rule is allowed to fire and apply its action, which makes it much easier to define assertions about the state after each processing round. A full presentation and proofs will be given in a forthcoming paper.

STATISTICAL ANALYSIS

An alternative to a formal analysis is a statistical analysis, studying the global behavior of the models in a rough statistical way. Such an analysis characterizes the behavior of each model as the partitioning of the set of all possible inputs into classes of patterns that yield the same result and comparing these partitions. Different analyses can be made depending on the way the results are interpreted. In this paper we can only show a small selection of the full matrix (analysis method x input set x model).

Sets

The universes of temporal patterns that we used are a collection of nested, abstract sets that are combinatorially complete (we only use two of them in this paper), and one large corpus of composed rhythms. The first test-set is the universe of all grid-based temporal patterns of a certain duration (referred to as *All*). This set is almost completely free of assumptions about musical knowledge and structure and will encompass, next to rhythms that can be easily remembered and performed musically, many examples that will be hard to interpret rhythmically at all. Removing all patterns from the previous set that cannot be generated from a simple metrical grammar (using only binary and tertiary subdivisions) gives us the subset of strictly metrical sequences (referred to as *Metric*). These patterns have a simple metrical interpretation in which each durational interval fits one level of a metrical hierarchy directly. The patterns are strictly metrical in the sense that there are no synchopations or tied notes. Note that they can still be

ambiguous –some patterns can be generated from different meters. Finally, to stay in line with the beat induction literature that shows a preference for musical ditty's, and especially anthems, we use the set of all national anthems (Shaw & Coleman, 1960; referred to as *Anthems*). This set consists of ca. 90% duple (70% is in 4/4) and 10% triple meters.

Monte Carlo method

Theoretically, the combinatorially complete test sets could be fed into the models and the exact size of each class of same-beat patterns calculated. However, the enormous size of the sets prohibits this (e.g., the size of the set of all grid-based temporal patterns of duration n is in the order 2^n). We used a practical way to yield reasonable estimates by Monte Carlo simulation: sampling the sets in a fair way and counting the response categories that arose. This method form the basis for a global statistical characterization of the behavior of the models. The sample size used for the set of all patterns and the set of strict metric sequences is 1000. The size of the Anthems set is 105, this set is always used as a whole.

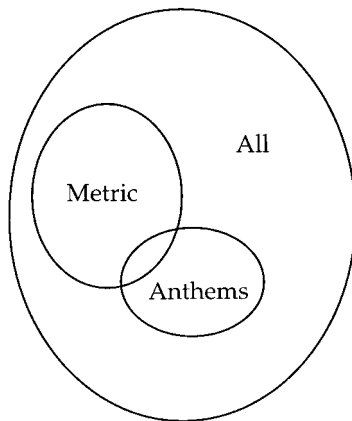


Fig. 2. Venn diagram of the corpora of temporal patterns as used for the analyses. The set of all grid-based temporal patterns (*All*), the set of strictly metric sequences (*Metric*), and the set of national anthems (*Anthems*).

BEAT-SPACE

First we will try to characterize the models in terms of their output for the different test sets, to get an insight in the range of the beat durations and phases, and to identify possible preferences. We will use Beat-space diagrams to show the distribution of beat duration for specific sets of patterns (see Fig. 3). The diagram shows the output of the

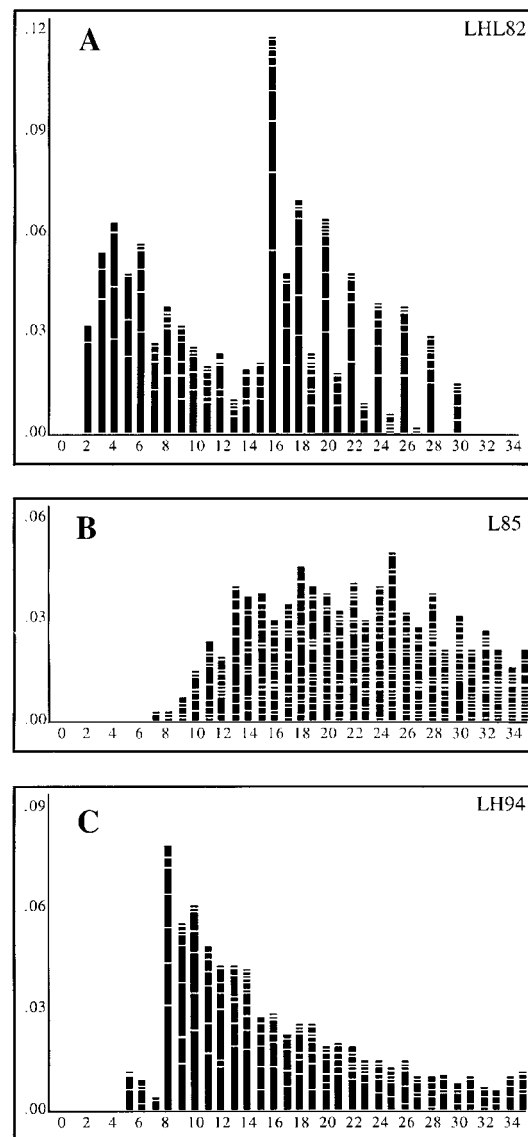


Fig. 3. Beat-space diagrams for the three models for the set of all patterns. (Proportion of patterns yielding a beat with a specific duration vs. beat duration counted in sixteenth-notes).

three models for the set of All rhythms. The x-axis indicates beat duration (in grid units that can be interpreted as a sixteenth note), the height of a bars indicates the proportion of patterns that yielded that beat duration. The subdivisions in each bar indicate the distribution of phases of the beat, with zero phases represented at the bottom of the diagram (later we will look at these phases in some more detail). For example, it can be seen in Figure 3a that the LHL82 model assigns to about 12% of all patterns a beat of duration 16.

If we look more globally at this measurement, we can see that the LHL82 model prefers beat durations of 16 grid-units or longer, L85 has some preference for beat durations in the range of 18 to 24 grid-units, and LH94 has a decaying preference for longer beat durations. Note that the distribution of durations is quite smooth (most notably for L85), which is surprising, considering the symbolic, discrete nature of the models. Furthermore, it can be observed that L85 has an relatively even distribution of phases, while LHL82 and LH94 have a preference for low phases.

For the Metric and the Anthems set (not shown) these spaces are sparser, with clearer differentiated beat durations, although the contour and overall distribution of phases stays the same. Next, we will take a closer look at these phases.

PHASE-SPACE

Phase-space diagrams depict the distribution of phases for a specific set (see Fig. 4 for the Phase-space for the set of all patterns). The x-axis indicates the phase duration (e.g., 0 is no upbeat, 2 is an upbeat of 2 grid-units), while the height of a bar indicates the proportion of a specific phase duration with respect to size of the whole pattern set.

In Figure 4 it can be seen that both LHL82 and LH94 have a clear preference for beats with zero phase, i.e., an interpretation without upbeats. This in contrast with L85 that has no particular preference for a particular phase. (Note that, because the beat-space was only analyzed for patterns with a duration of upto 35 gridunits, for L85 and LH94 the proportion of phases do not add to 1).

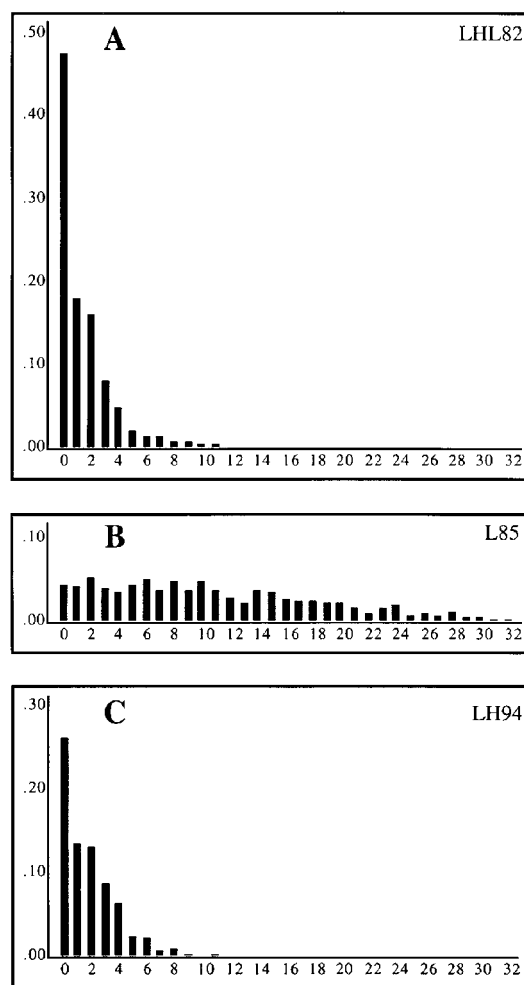


Fig. 4. Phase-space diagrams for the three models for the set of all patterns. (Proportion of patterns yielding a beat with a specific phase vs. beat phase counted in sixteenth-notes).

AGREEMENT

Having shown how the overall distributions of the model's results differ, the question arises what the relation between is between the results of the models for a specific input pattern? For that, patterns are taken from the sets and are categorized into four classes: the class of patterns for which the three models agreed on the same beat, the three classes of patterns for which only two models agreed, and the class of patterns that resulted in three different answers. We allowed an integer multiple of the beat duration to count as an agreed beat, provided that the phases matched as well.

The result of this measurement can be depicted in a histogram. The x-axis indicates the duration of the pattern that is used as input to the model. The height of the bar shows the proportion of patterns of that duration for which at least two models agreed. The black part of each bar indicates the proportion of patterns for which all three models agreed on the beat. As an example, consider Figure 5a. It can be seen that for all patterns of duration 25 there is 30% agreement between all three (the black part of each bar), and about 85% agreement between at least two models (the total bar height).

The diagrams show in general that agreement between the three models (black part of the bars) increases with the amount of musical structure in the sets, upto 50% for the Anthems set. This may indicate that part of the differences of the models are exhibited mainly when they are applied outside the domain of input patterns for which they were conceived. Furthermore, it can be observed that the agreement between LHL82 and LH94, with L85 differing, increases with longer patterns (light gray part of the bar).

SPEED OF BEAT INDUCTION

Now we have shown how the models can arrive at the same or different answers, we come to the ques-

tion how fast these answers are arrived at. Both LHL82 and LH94 have an explicit point at which processing stops and a result is returned. The distribution of the proportion of patterns yielding such a confirmed beat can be depicted as a function of the pattern length. However, since it is quite natural to need more time to establish a long beat, a different, and possibly fairer representation of the same data is made by expressing the time needed for confirmation relative to the duration of the beat found. What this analysis (see Fig. 6) shows is that, roughly spoken, both models can establish a beat of a certain duration relatively fast, on the basis of between two and three beats worth of rhythmical material. However, in some cases LHL82 needs a much longer fragment to confirm a beat. Here it turns out that the models clearly predict a very fast beat induction process that contrasts with, for example, the much larger amount of material that coupled oscillator models (e.g., Large & Kolen, 1994) need to establish locking.

CORRECTNESS

Now we have looked at the speed at which the models arrive at an answer, we should look at its correctness. We plan to compare the models to empirical data of human subjects, but for some of

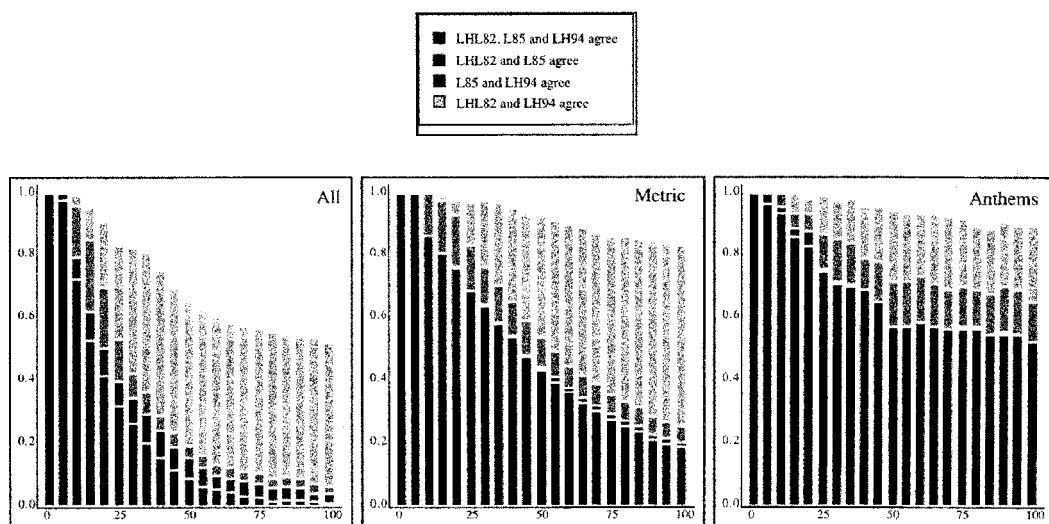


Fig. 5. Agreement diagrams for the different sets. (Proportion of patterns for which the models yield compatible beats vs. pattern duration counted in sixteenth-notes).

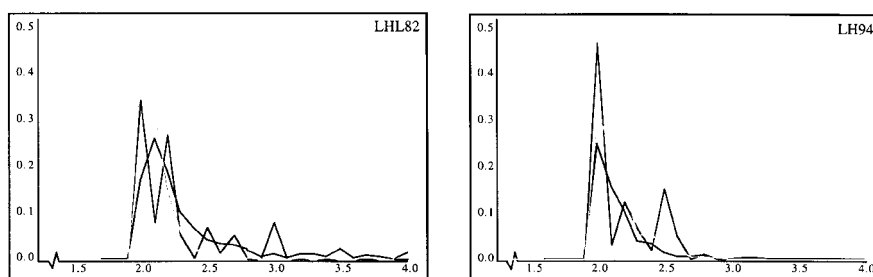


Fig. 6. Moment of confirmation for the sets All (black line), Metric (gray line), and Anthems (light gray line). (Proportion vs. moment of confirmation relative to the beat-length).

the subsets a rough approximation of the correctness of the results can already be derived. For the set of strictly metrical patterns, correct beats can be defined to be those that fit one of the metrical levels of one of the generating meters. For the set of national anthems, a correct beat can be defined as a beat that is compatible with the meter notated in the score. For short patterns that form the beginning of more than one anthem we counted a beat as correct whenever it fitted the meter of one of those anthems. Because this measure is not very stringent, it may not be valid as an absolute level of performance, but it can function well in comparisons between models.

The set of anthems may very well contain examples where the meter is conveyed by the melody and not by the metrical structure. In that sense the measure of correctness could underestimate the performance of the models, which have access to the rhythm only.

It is easy to yield small beats that conform the meter of a piece at a very low level (these beats are much more likely to fit than large ones, because large beats have more degrees of freedom in choosing a phase). And it is difficult to judge the merit of a beat that, though having a proper phase, spans several bars. Therefore the correctness was differentiated according to the metrical level of the resulting beat.

In Figure 7 these measures are depicted as histograms. The x-axis indicates the duration of the pattern that is used as input to the model (in eighth notes). The height of the bar shows the proportion of patterns that yielded a beat compatible with the notated meter of any of the anthems starting with that pattern. The black part of each bar indicates

the proportion of hyper-meter results, the beat spanning more than one measure. Below that, the dark gray area indicates the proportion of patterns that yielded a proper bar as output of the model. The gray area below that indicates the proportion of correct beat-level answers, i.e., the bar divided by 2 or 3 according to the time signature. Finally, the light-gray areas indicate answers that can still be considered correct but that aligned with lower levels of the metrical hierarchy. As an example, consider Figure 7a in which it can be seen that L85 rated 36% of the patterns with a duration of 24 eighth-notes correctly. About 15% were rated correctly at the bar level, about 15% at the beat level, 3% above the bar level and 3% below the beat level.

In these figures it can be seen how in subsequent stages of the processing the beat hypothesis shifts upwards through the metrical levels, becoming larger and larger. The overall performance that the models arrive at finally is quite remarkable, considering that in some anthems the meter might be communicated through the melodic structure, information that is ignored by these models. The LHL82 model reaches about 60% of correct answers at the bar or beat level and seems to have the best performance (but see the section on optimal parameter settings).

However, these absolute figures have to be read with caution. They have to be compared, e.g., to the likelihood of arriving at a correct answer by guessing. This baseline can easily be established by doing the same measurement for a statistical model that only knows the distribution of bars and beats in the total set and randomly selects a duration and a phase according to that distribution. The cor-

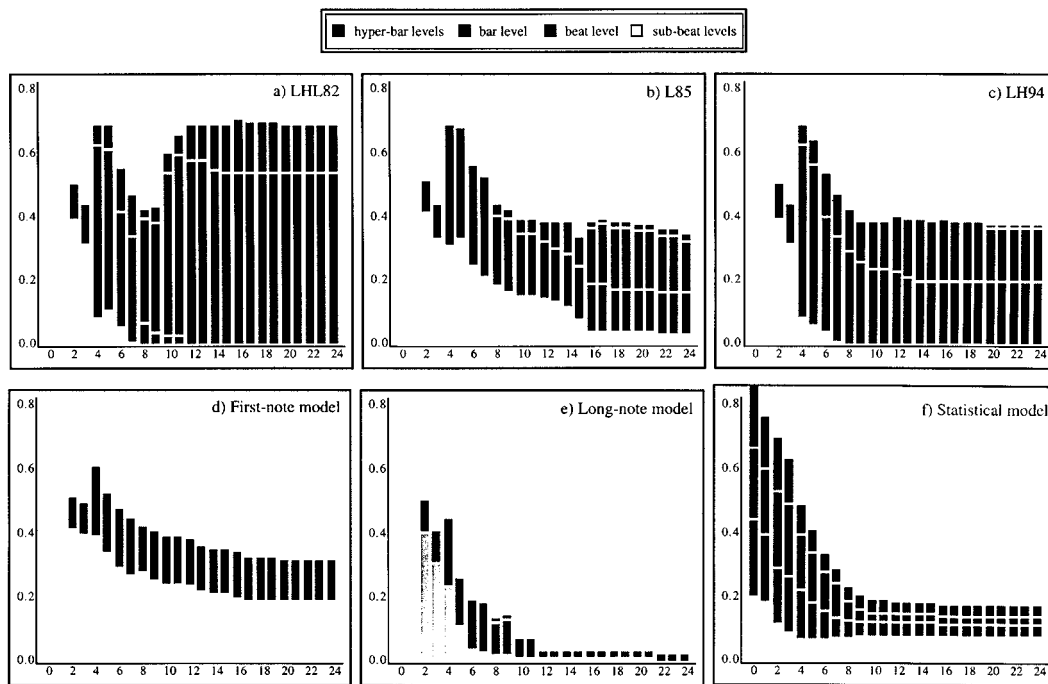


Fig. 7. Correct-level diagrams for the three rule-based models and three baseline models using the Anthems set. (Proportion of patterns yielding a correct beat vs. pattern duration counted in eighth-notes).

rectness of this model is shown in Figure 7f, it forms a reference for judging the correctness of the models.

Another baseline that can be used is a model that just assumes that the first note is the proper beat. This hypothesis, which all models use initially, turns out to be not such a bad one, as is shown in Figure 7d, it does yield a correct result in about 35% of the cases. However, these results only reach the beat level in 10% of the cases.

A third baseline assumes that the longest note encountered is the proper beat. And in one way or another the detection of long notes play a crucial role in each model. However, after some initial success this strategy turns out to be quite unsuccessful, it even performs below chance level, as can be seen in Figure 7e.

RULE CALLS

Before studying the contribution of the individual rules to the model's performance one needs to check how often the different rules fire? In Figure

8 the proportion of cases in which each rule fires is given for each model and each set.

It can be seen how the LHL82 LONGNOTE rule applies relatively infrequent compared to the other rules (LONGNOTE fires when a note is longer than twice the current beat). The STRETCH rule, in which a long note is encountered that does not align with the beat, applies more often while processing random rhythms than in the structured sets. Conversely, the CONFLATE rule, that fires whenever a note is encountered on a next beat, is more often called in the Metric and the Anthems set. Both results confirm the expectation about the amount of musical structure in the different sets.

ROBUSTNESS

After we know how often the rules are applied, it can be questioned how crucial the application of a rule's action is when its matching pattern is encountered in the input. It can be argued that meaningful musical material does contain many redundant cues to the meter (as can be experienced when tuning the

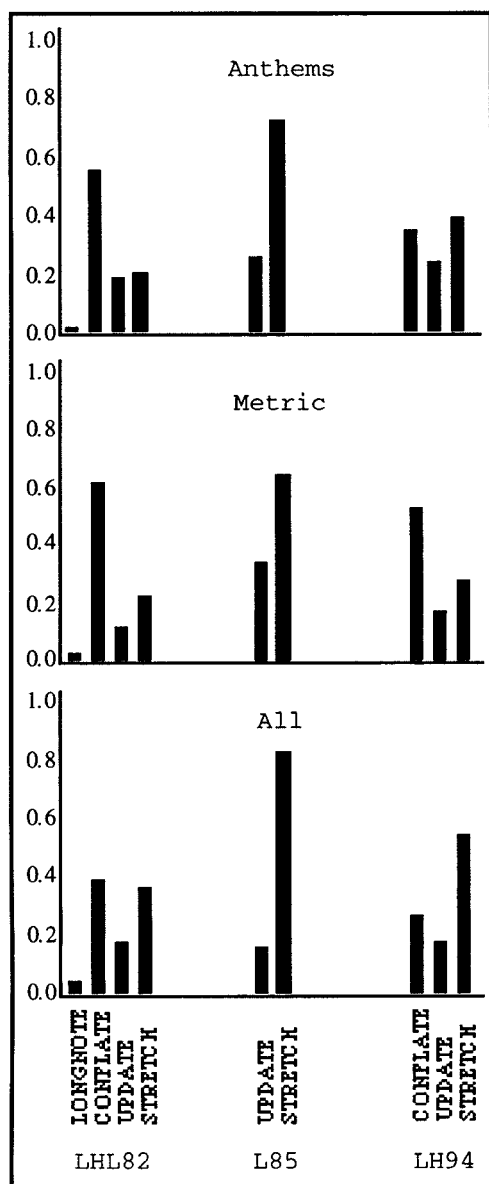


Fig. 8. Proportion of rule calls for the three models for the set of all patterns (top), the Metric set (middle), and the Anthems set (bottom).

radio and finding oneself suddenly listening to the middle section of an anthem), and the system might get a second chance at getting it right. This issue was resolved in the form of a measurement that gave one rule a change of not firing in a situation where it otherwise would. In Figure 9 the resulting performance (the proportion of correct answers at the bar or beat level) is given as a function of the chance that a rule fires when it should. The correct-

ness measurement was taken at the point where processing stopped, this is why the level of correctness when all rules fire as they should (at 1.0) cannot be directly compared to the final correct proportion at the bar and beat level in Figure 7a, 7b and 7c.

It turns out that all three models behave quite robustly under this condition, even in case of the complete removal of a rule. Overall, there seem to be multiple cues in the music that allow a later repair of the situation caused by a broken rule. This holds especially for the CONFLATE rule which can double the beat. Furthermore, it can be observed that STRETCH has quite an important role in both LHL82 and LH94. Remarkably, the performance of L85 improves when the STRETCH and the UPDATE rule are not always used. And indeed L85 seems to take long notes too seriously and often changes the beat at syncopations.

EFFECTIVENESS OF THE RULES

We will continue focussing in more and more on the rules themselves and next address the question how beneficial the actions of the different rules are when they fire. This was measured by counting the cases in which a rule succeeds in repairing a wrong beat, i.e., the beat being wrong just before and correct (at any level) just after the application of a rule. This number can then be compared to the number of cases in which a rule's action breaks the beat, i.e., it was correct before and wrong after application. In Figure 10 the proportion of these cases is shown for each rule (given that it fires). Both the INITIALIZE and CONFIRM rule are not depicted in Figure 10. The former because it invariably repairs a missing beat hypothesis into a correct one in 70% of the cases, the latter because it doesn't alter the beat hypothesis, and therefore it never breaks nor repairs.

It can be seen that the STRETCH rule does a good job in repairing wrong beat hypotheses. The UPDATE rule in LH82 never breaks or repairs the beat since it simply shifts the beat. The absence of a rating for LONGNOTE in LH82 is due to an artifact in our measurement procedure. The large proportion of cases in which the UPDATE rule breaks a correct beat hypothesis in model LH94 is puzzling and topic of further study.

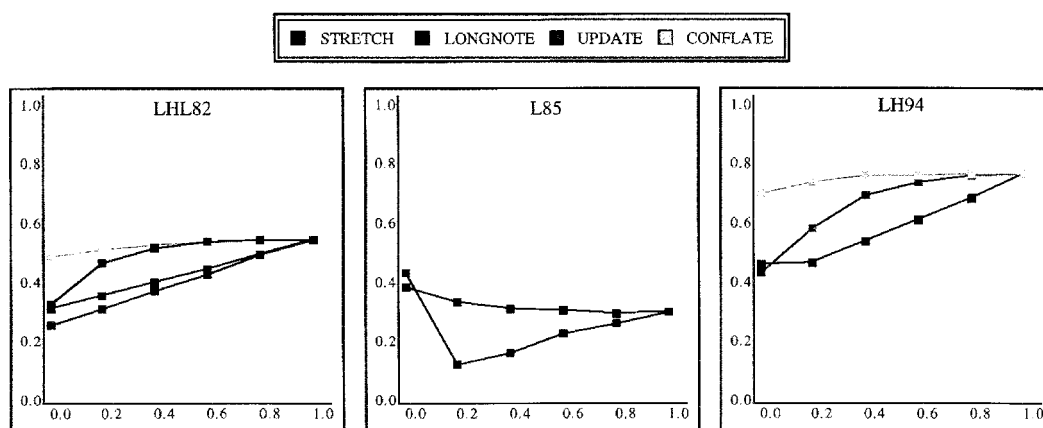


Fig. 9. Robustness for the three models for the Anthems set. (Proportion of patterns yielding a correct beat vs. application probability of each rule).

IS THERE A BEST UPDATE RULE?

In the models three different UPDATE rules are used. The UPDATE rule is intended to skip over upbeat. All UPDATE rules fire whenever they

encounter a relatively long note (under specific conditions) in the input. But all models apply a different action in that case. LHL82 shifts the beat, maintaining its duration. The LH94 model shifts and elongates the beat to the onset of the next note. The L85 model shifts and elongates the beat to the onset of the next note that happens to fall on a beat. These actions are illustrated in Figure 11. It shows an example pattern (a line is a note onset, the curves indicate the beats) before the UPDATE rule fires, and the situation after it has done its action.

To test the different UPDATE rules, we transplanted the different variants into the three models. Because in LHL82 the function of the UPDATE rule is closely intertwined with the LONGNOTE rule, it was transplanted both with and without the LONGNOTE rule to the other two models. In two cases the transplanted LONGNOTE rule was

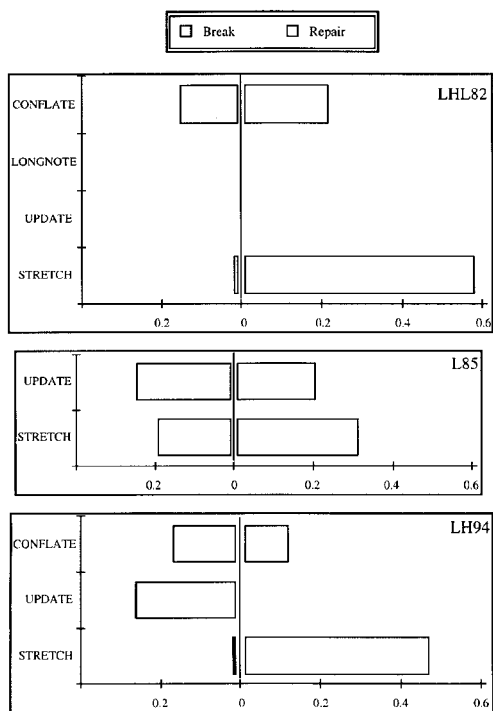


Fig. 10. Effectiveness of the rules for the three models for the Anthems set. (Rule vs. proportion of rule calls that are breaks and repairs of the current beat hypothesis).

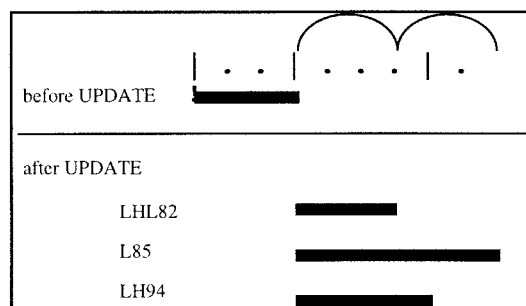


Fig. 11. Definitions of the UPDATE rule.

made inoperative by the rest of the rule set (the UPDATE rule from L85 makes the beat so long that LONGNOTE never applies anymore). These combinations were eliminated. The resulting rule cocktails were all tested for correctness, measuring at the point of confirmation for LHL82 en LH94 and at the end of the anthem for L85. The results are given in Figure 12, with the subdivisions of the bars as used in the figures for the Correct-level analysis: black is above, dark gray on and gray just below the bar level (i.e., beat level), light gray being the correct answers below the beat level.

The results are aligned such that the proportions can be compared per rule-cocktail for the beat and bar levels (gray and dark gray). The correctness at the sub-beat levels (light gray) and hyper-bar levels (black) constitute less useful answers.

In Figure 12 we can see that LHL82 obtains the best score (i.e., 55% correct at the beat and bar level), followed by LHL82 with the UPDATE rule form LH94. No difference between the performance of LH94 and of the LH94 model with a transplanted UPDATE en LONGNOTE rule from LHL82. However, these results were calculated with the parameter settings supplied by the author, and different settings yield a different result, as will be shown next.

WHAT ARE THE OPTIMAL PARAMETER SETTINGS?

Finally, we will use the correctness analysis (considering only bar and beat level answers as correct) to search for the optimal parameter setting of the models, i.e., the setting that produces the highest proportion of correct results.

We will show here the results for the LHL82 and LH94 model (the L85 model has no parameters). In Figure 13 the performance of the models is shown as function of their parameters. The LHL82 model, which was described in the literature with an unformalized “near-beginning” predicate that controlled whether the update rule is still allowed to fire, was augmented with an update-interval parameter that specified this point in time.

The LHL82 model achieves an optimal performance with the *maximum-beat* parameter (which controls whether beats may still be conflated) set around 20 sixteenth notes (original parameter setting is 30 sixteenth notes) and the *update-interval* parameter (which controls until when the UPDATE rule still allowed to fire) set around 16 sixteenth notes (no value given in the original model). The optimal performance level then rises to about 60% correct at the bar or beat level. The LH94 model

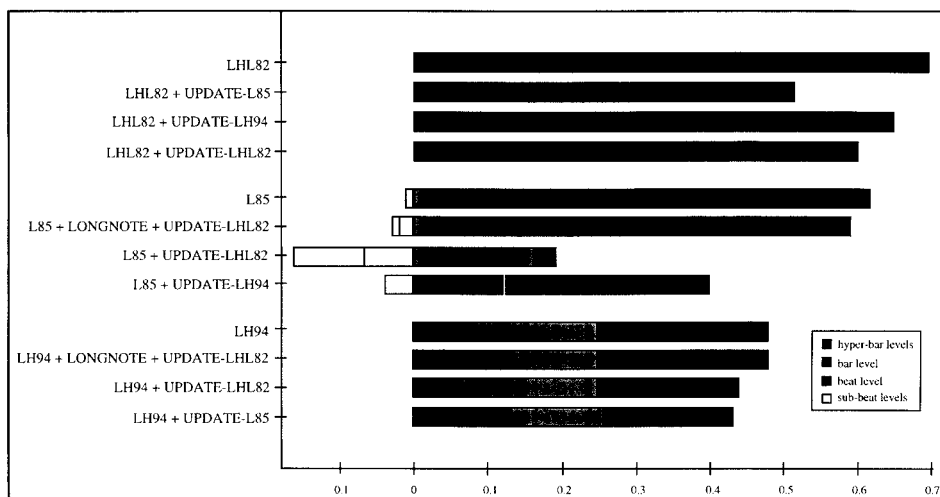


Fig. 12. Correctness of the rule-cocktails and the original models for the Anthems set. (Proportion of patterns yielding a correct beat vs. rule cocktails).

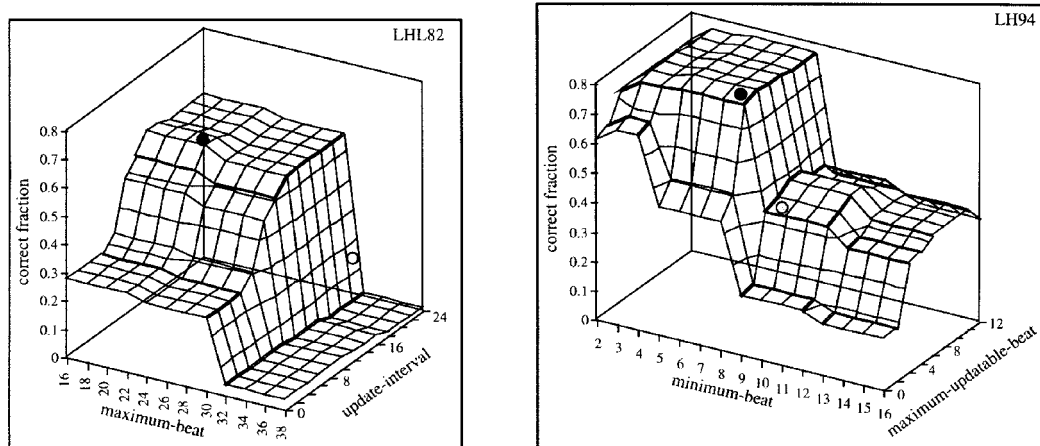


Fig. 13. Parameter-spaces for the Anthems set (gray = original setting, black = optimal setting). (Proportion of patterns yielding a correct beat vs. parameter values).

yields its best performance with the *minimum-beat* parameter (which prohibits the confirmation of small beats) set at 6 sixteenth notes (original parameter setting is 8 sixteenth notes) and the *maximum-updatable-beat* parameter (until when the UPDATE rule still is allowed to fire) set at 6 sixteenth notes (original parameter setting is 4 sixteenth notes). The level then rises to 80% correct, which makes this model an improvement indeed. (A larger section of the parameter-space will be given in a future paper).

CONCLUSION AND FUTURE RESEARCH

We also hope to have shown that these methods are a promising way of characterizing the behavior of computational models of beat induction. Using these methods we were able to answer a set of questions that could not have been addressed otherwise. Our plans are to extend this method to families of models that are based on alternative formalisms. Furthermore, we plan to elaborate a perceptual measure of correctness that is based on empirical data. After a further round of formalization we will attempt to generalize the idea of rule cocktails to systems that have their rules specified in the form of a formal pattern-matching language, such that the members of this family may be enumerated and tested, possibly using genetic algorithms. The

rule-based models, even though they are simple and ignore effects of like tempo, timing, melody and harmony, turn out behave surprisingly well.

ACKNOWLEDGMENTS

Special thanks to Christopher Longuet-Higgins for his contribution in several discussions, and providing access to his programs. Part of this work was done while visiting CCRMA, Stanford University on kind invitation of Chris Chafe and John Chowning, supported by a travel grant of the Netherlands Organization for Scientific Research (NWO). The research has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW).

REFERENCES

- Desain, P. & H. Honing (1994a). A brief introduction to beat induction In *Proceedings of the 1994 International Computer Music Conference*. 78–79. San Francisco: International Computer Music Association.
- Desain, P. & H. Honing (1994b). Rule-based models of initial beat induction and an analysis of their behavior. In *Proceedings of the 1994 International Computer Music Conference*. 80–82. San Francisco: International Computer Music Association.
- Desain, P. & H. Honing (1995). Computational Models of Beat Induction: The Rule-based Approach. In G. Wid-

- mer (ed.), *Working Notes: Artificial Intelligence and Music*. 1–10. Montreal: IJCAI.
- Essens, P. (1995). Structuring Temporal Sequences: Comparison of Models and Factors of Complexity. *Perception & Psychophysics*. 57(4), 519–532.
- Large, E.W. & J.F. Kolen (1994). Resonance and the perception of musical meter. *Connection Science*, 6(1), 177–208.
- Lee, C.S. (1985). The rhythmic interpretation of simple musical sequences: towards a perceptual model. In R. West, P. Howell, & I. Cross (eds.) *Musical Structure and Cognition*. 53–69. London: Academic Press.
- Lee, C.S. (1991). The perception of metrical structure: Experimental evidence and a model. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure*. 59–127. London: Academic.
- Longuet-Higgins, H.C. & C.S. Lee (1982). Perception of musical rhythms. *Perception*. 11, 115–128.
- Longuet-Higgins, H.C. (1994). Unpublished computer program in POP-11, describing an algorithm named “shoe”.
- Miller, B.O., D.L. Scarborough, & J.A. Jones (1992) On the perception of meter. In M. Balaban, K. Ebcioglu, & O.Laske (eds.), *Understanding Music with AI: Perspectives on Music Cognition*. 428– 447. Cambridge: MIT Press.
- Shaw, M. and H. Coleman (1960) *National Anthems of the World*. London: Pitman.



Peter Desain
NICI, Nijmegen University
P.O. Box 9104
NL-6500 HE Nijmegen
The Netherlands
E-mail: desain@nici.kun.nl
www.nici.kun.nl/mmm



Henkjan Honing
NICI, Nijmegen University
P.O. Box 9104
NL-6500 HE Nijmegen
honing@nici.kun.nl
www.nici.kun.nl/mmm
and
Music Department
University of Amsterdam
Spuistraat 134
NL-1012 VB Amsterdam

Peter Desain and Henkjan Honing direct the Music, Mind, Machine group at the Nijmegen Institute of Cognition and Information (NICI), University of Nijmegen. This research project is concerned with the computational modeling of musical knowledge and music cognition, concentrating on the temporal aspects of music perception and music performance such as rhythm, timing and tempo.