

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/74722>

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

A connectionist and a traditional AI quantizer, symbolic versus sub-symbolic models of rhythm perception

Peter Desain

*Utrecht School of Arts, The Netherlands and Music Department, City University,
London, UK*

The Symbolic AI paradigm and the Connectionist paradigm have produced some incompatible models of the same domain of cognition. Two such models in the field of rhythm perception, namely the Longuet-Higgins Musical Parser and the Desain & Horning connectionist quantizer, were studied in order to find ways to compare and evaluate them. Different perspectives from which to describe their behavior were developed, providing a conceptual as well as a visual representation of the operation of the models. With these tools it proved possible to discuss their similarities and differences and to narrow the gap between sub-symbolic and symbolic models.

KEY WORDS: Rhythm perception, quantization, (sub-)symbolic processing, connectionism

Introduction

The so called Good Old Fashioned Artificial Intelligence has established itself firmly in the past decades as a research methodology. The methods and tools it uses are symbolic, highly structured representations of domain knowledge and transformations of these representations by means of formally stated rules. These rule based theories can function (and are vital) as abstract formal descriptions of aspects of cognition, constraining any cognitive theory. However, some authors go beyond that and claim that mental processes *are* symbolic operations performed on mental representations of rules (see Fodor, 1975). Until the connectionist paradigm emerged there was no real alternative to that view. But now, in this new paradigm, the departure from reliance on the explicit mental representation of rules is central, and thus the conception of cognition is fundamentally different (Bharucha & Olney, 1989). This holds regardless of the fact that the behavior of connectionist models could be formally described in rules. These distributed models consist of a large number of simple elements, or cells, each of which has its own activation level. These cells are interconnected in a network, the connections serving to excite or inhibit others. Connectionism opened the possibility of defining models which have characteristics that are hard to achieve in traditional AI, in particular robustness, flexibility and the possibility of learning (Rumelhart & McClelland, 1986). The connectionist boom brought forth much interesting work, also in the field of music (Todd & Loy, forthcoming). Although many researchers lost their critical attitude,

impressed by the good performance of some (prototypical) models, it became soon clear that more study was needed to the limitations of these models. A connectionist model that 'works' well, constitutes in itself no scientific progress, when questions like the sensitivity to parameter changes, the scalability to larger problems and the dependancy of the model on a specific input representation, cannot be answered. However, it is possible to describe the behavior of a connectionist model from different abstract perspectives that provide more insight in its limitations and its validity as a cognitive model than simulations or test runs alone. These perspectives are also fruitful for the analysis of traditional AI models. In this article we pursue this approach for a connectionist and a traditional AI model of rhythm perception as a case study for the wider issue how the paradigms themselves relate.

The quantization problem

In performed music there are large deviations from the time intervals as they appear in the score (Clarke, 1987). Quantization is the process by which the time intervals in the score are recovered from the durations in a performed temporal sequence; to put it in another way, it is the process by which performed time intervals are factorized into abstract integer durations representing the notes in the score and local tempo factors. These tempo factors are aggregates of intended timing deviations like rubato and unintended timing deviations like noise of the motor system. This process of separating different discrete and continuous aspects of musical timing, however simple at first sight, and indeed forming a rather basic musical skill, proved to be very hard to model (Desain & Honing, forthcoming). As an example one could try to recover the intended rhythmic interpretation of the following temporal sequence (in milliseconds):

476 : 237 : 115 : 135 : 174 : 135 : 155 : 240 : 254 : 118 : 112 : 118 : 138 : 476

This task, however hard by calculation, yields an obvious and simple answer when the data is converted to an auditory stimulus: a sequence of drumbeats (the solution is given in note 1).

Known methods

The simplest method of quantization, used by most commercially available music transcription programs, is the round-off of any point in time to the nearest point on a fixed time grid, with a resolution equal to, or an integral factor smaller than, the smallest duration to be expected. This method is totally inappropriate: even when enhanced with facilities like user control over the grid resolution, it yields results that makes no musical sense, even when the performer is forced to play along with a metronome.

However, this method can serve as the basis of more reasonable models in which the time grid is adapted if consistent deviations (notes being late or early) are detected. In this so called 'tempo-tracking' the design of the control behavior becomes crucial: the extraction of an error signal between time grid and note onsets, and the way in which this error influences the tempo of the grid. The most elaborate example is the 'real time foot tapper' (Dannenberg & Mont-Reynaud, 1987 and

Boulanger, 1990), but still a 30% error ratio is reported for this system.

A symbolic, rule based system for quantization was in place at the Stanford automatic transcription project (Chowning, Rush, Mont-Reynaud, Chafe, Schloss & Smith, 1984). It used knowledge about preferable ratios between time intervals as a basis for an optimal quantized description of the data. In such a knowledge based system it is relatively easy to use information from other domains (e.g. dynamic, harmonic) to help the quantization process, but one has to keep in mind that this increases the risk of style dependency and therefore brittleness. Because of its design as an unordered collection of rules it is, like all rule based systems, impossible to characterize its behavior in non-operational terms.

The musical parser (Longuet-Higgins, 1987) comprises another symbolic AI approach to quantization, besides methods of tonal and articulation analysis that will be ignored here. It is highly hierarchical in its music representation and has a reasonable good performance. Furthermore it had the advantage of a published program being available. A Lisp version of this program is published in Desain (1990).

The connectionist quantizer (Desain & Honing, 1989, 1991; Desain, Honing & de Rijk, 1989) is a distributed model of fairly simple processing elements. This model displays desirable properties like robustness, graceful degradation and precedence of local context, but as a model it is hard to understand why it works so well, and what its limitations are.

These last two methods will now be described in more detail.

The Longuet-Higgins Musical Parser, a symbolic model

Using just a little knowledge about meter, and exploiting that to the extreme, the Longuet-Higgins Musical Parser builds a metrical tree from performance data, and thus implicitly manages to quantize it. This method is supplied with an initial notion of a time interval called the beat. This interval is subdivided recursively in 2 or 3 parts looking for onset times near the start of each part, until the interval contains no more onsets. The 'best' subdivision is then returned. At each recursive level the interval length is adjusted on the basis of the onsets found, just as in simple tempo-tracking models.

The output of the system consists of a list of trees, one for every analyzed beat. Each tree is of a combined binary-ternary nature, which means that each node has 0 (in case it is a leaf of the tree) or 2 or 3 sub-trees. During the construction of the tree there is a horizontal flow of information through the layers of the tree, seeking to maintain the same kind of subdivision at a certain level as long as possible. The description of the proposed subdivisions at each level of the tree is called meter. During the construction of the tree a strict left to right order is maintained, and new sub-trees are created on a generate-and-test basis. This means that a proposed (and constructed) binary sub-tree may be rejected in favour of a tertiary one. The generate and test procedure is non-standard in that it may, after checking and rejecting the first alternative, still reject the second in which case as yet the first alternative is chosen.

There is one parameter (called tolerance) identified in the program. It is used in different places as the allowed margin of deviation in deciding if notes start or stop at a certain time. In this way the model does depend elegantly on global tempo by

limiting the possibility of further subdivisions when an absolute time span (the tolerance) is reached: onsets that happen within the tolerance interval are considered synchronous.

The Desain & Honing connectionist quantizer: a sub-symbolic model

A class of connectionist models, known as interactive activation and constraint satisfaction networks generally behave so as to converge towards an equilibrium state given some initial state. The connectionist quantization model is designed to converge from non-metrical performance data to a metrical equilibrium. The network topology is fixed (hard-wired) and so is the kind of interaction between cells: no learning takes place.

The net comprises cells for each time interval in a temporal sequence, be it basic (one inter-onset: interval) or compound (spanning several notes). Two cells representing neighboring time intervals may interact and push each other to their 'perfect' values implied by an integer ratio, propagating the changes through the net. After a while the net stabilizes and a quantized temporal sequence can be read out. The interaction between cells, the change of their ratio, depends on the ratio of their durations, via a so called interaction function. Since the ratio of two time intervals is the only determinant of local behavior, the quantization result does not depend on absolute global tempo, nor can it handle polyphony. The interaction function is a section-wise polynomial with 2 parameters called peak and decay; the first reflects the size of the 'capture' range around an integer ratio, the second represents the decreasing influence of higher ratios. It has to be stressed that all aspects of the global behavior of the system are determined completely by these parameters.

A model that takes a whole temporal sequence into consideration at once is not feasible when the aim is to develop a cognitive model. Luckily, it proved quite simple to design a version of the quantizer which operates upon a window of events. In such a model tempo tracking can handle slow global tempo changes. For reasons of space this part of the connectionist quantizer will not be described here.

Differences

The models described can be characterized as complete antipoles in a number of aspects. They are summarized roughly in the table in figure 1. The huge differences made comparing them quite hard, but in the end the work was gratifying. Because the systems are prototypical for the two main AI paradigms the results may well generalize to other cases.

Different perspectives

Different perspectives for describing these models will now be given, each at its own level of abstraction. Some perspectives will generalize over sets of inputs or parameters, some will reduce the amount of variability by keeping certain concepts fixed. I hope to show that this search for different representations of the behavior

Longuet-Higgins musical parser	Desain & Honing connectionist quantizer
symbolic central search hierarchical knowledge based	numerical distributed optimization heterarchical knowledge free

Figure 1 A summary of the differences of the two models under study

of a computational cognitive model, conceptual as well as visual, is fruitful, even for analyzing traditional symbolic AI programs.

The most direct and raw representation of a computational model is a trace of the computation itself, an overview of how the internal state of the system changes in the course of a complete calculation as a function of the computation-time or the number of computation steps taken. A visualization of such a trace for the connectionist model is shown in Desain & Honing, (1989) and for the Longuet-Higgins parser similar graphic representations can be devised.

A deficit of these representations is that they can only be given for one example input at a time, and thus are extremely dependent on the choice of input – in a sense it is easy to ‘lie’ with these examples by picking one that behaves well. But, on the other hand, these representations show in full detail the ongoing processes and thus enable interpretations and hypothesis forming.

At the other end of the spectrum of possible perspectives is the statistical method, reducing all the information to a number of correct responses. We can assume that, when a skilled performer plays a rhythm, the performed temporal sequence should be quantized as the presented score. Collecting a set of performances and counting the numbers quantized correctly by the model gives us then an indication of its validity. The number of correct quantizations will in general be a function of the possible parameter values given to the model. Visualization of this dependency is useful in the study of the parameter-sensitivity of the models. Often connectionist models behave badly in this respect. They might need specific parameter settings for different problems. Or they might not ‘scale-up’: for larger problems the model only works for an increasingly smaller range of parameter settings specific to the problem at hand. Parameters might also have no cognitive relevance, and as such could not be used to control the global emerging behavior of the model, or they might be highly dependent. A visualization of the parameter space can detect such problems. Both models behave well in this respect, but because of space limitations we cannot present the parameter spaces here.

Both perspectives have their drawbacks, one being too specific, the other one too general. If we give up some detail on the speed and order of processing that was available in the computation trace, and give up the free choice of musical material that was available in the parameter space we can characterize the precise behavior of the system for a family of sequences: all possible sequences of a fixed small length. This set can be considered to constitute a rhythm space: the problem space of quantization.

Rhythm space perspective

Let us consider the 3 dimensional space of all possible temporal sequences of 3

inter-onset intervals (four bangs on a drum). Every point in this space represents a unique temporal sequence. One could envisage this space projected in a room, with one corner as the origin. The distance along one wall represents the length of the first time interval, the distance along the second wall represents the length of the second interval, and the height above the floor represents the third. In this space certain points will be perfectly metronomical sequences, other points will represent performed deviations from them. Let us call this space 'rhythm space' although 'temporal sequence space' would be more appropriate. A quantization process maps each point in this space to another; it assigns to each sequence a solution of the quantization. Thus the problem space of a quantization method is the whole rhythm space, the solution space is a set of points within this rhythm space. A further characterization of the solution space (e.g., what constraints limit the set of permissible quantizations – is, for instance, a complex temporal pattern such as 7 : 11 : 2 to be considered allowable?) cannot be given at the moment, which is part of the reason why quantization is a difficult problem to define.

Trajectories in rhythm space

If the model has intermediate processing states that are temporal sequences themselves, as is the case with the connectionist model, the computation trace becomes a trajectory through this rhythm space. Otherwise a simple straight line can indicate the mapping from problem to solution. Easy visualization requires mapping of this space to 2 dimensions which can be done by assuming that the whole time-length of the temporal sequence is kept constant, the third interval then follows from the first two and the first two durations can be taken as the only independent variables: the x and y axes of a diagram. This normalization, which factors out global tempo, reduces the general applicability of the method if the theory is itself dependent on global tempo. The connectionist quantizer does not (but we admit that to model human rhythm perception accurately, it should). The Longuet-Higgins model does depend on global tempo and for this model the rhythm space can only be shown for one global tempo at a time.

If all three intervals are restricted between a minimum and a maximum time span, the allowed portion of the 2 dimensional projection forms a parallelogram. In figure 2a and 2b the rhythm space for the two models is shown, given an input sequence of three notes between a sixteenth and a double dotted quarter note. The whole sequence has a total duration of three quarter notes. The different solutions are indicated by small circles. Note that in the Longuet-Higgins model some solutions contain inter-onset intervals of length zero. That is because this model interprets two onsets that happen within the tolerance as synchronous.

Regions in rhythm space

Because the behavior of the connectionist model is completely determined by a temporal sequence, any point on a trajectory will be mapped to the end point of that trajectory. This means that the connectionist model 'carves up' rhythm space into little compartments around each solution. Each compartment or region contains all the sequences that will be quantized equivalently. Now we can abstract from the trajectory from initial state to solution and only characterize the compartments. These areas, the so called basins of attraction, can be shown as a partitioning of the

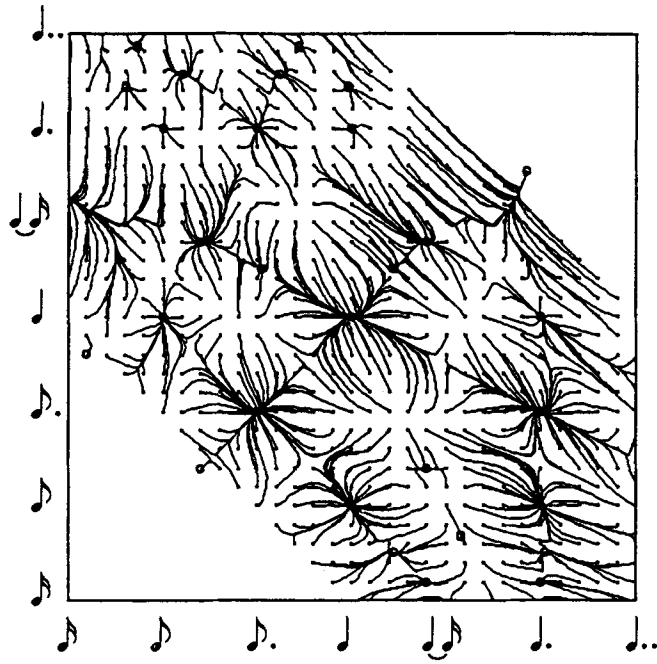


Figure 2a Trajectories through rhythm space in the connectionist model

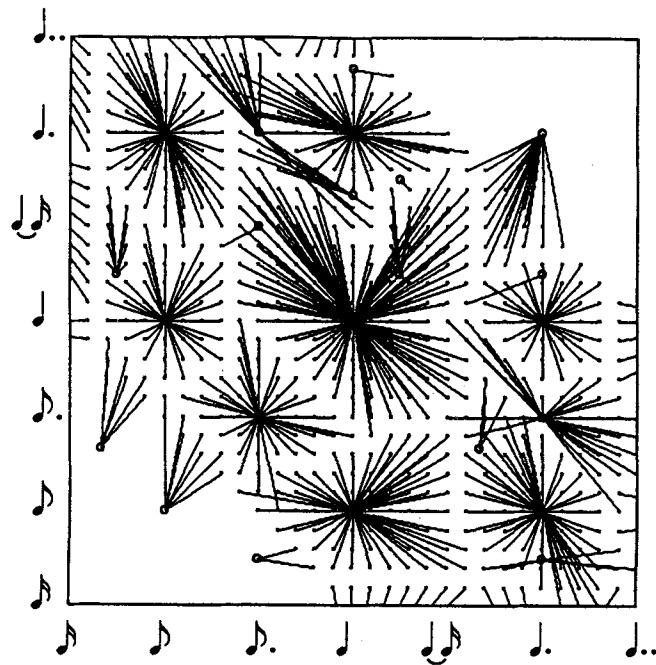


Figure 2b Mapping in rhythm space in the traditional AI model

rhythm space, as is depicted in figure 3a. The Longuet-Higgins model does not behave such that the solution itself will always lie within the region that will map to that solution. But still the region of all points that map to the same solution (the equivalence classes of the mapping) can be shown as in figure 3b. We can now check the differences between the models. E.g. the region that maps to the rhythm of three quarter notes (marked A in the figures) is much larger in the Longuet-Higgins model than it is in the connectionist model. Another difference is the behavior around the region marked C in figure 3a, which corresponds to a 2 :1 :2 rhythm. This solution is not present in the Longuet-Higgins model, because it is based on a five-fold division.

Influence of context

A good way to understand the influence of context (previously presented musical material) is to consider how these maps change under the influence of it. In figure 4 a context of two dotted quarter notes was presented before the notes shown in the rhythm space. This context heavily biases the behavior of the connectionist model to quantize the following inter-onset intervals in subdivisions thereof as is shown by the enlargement of the area marked B. The area marked C completely disappears in the light of the contextual evidence for these subdivisions. Also in the Longuet-Higgins model the quantization is influenced (or even guided) by context. It does so by propagating the established meter to the processing of the remaining data. Using a duple meter as context, a very similar distortion of the regions in rhythm space can be seen (figure 4b).

Influence of parameters

These maps can also be used to understand the influence of the parameters of the model by interpreting their changes under the influence of different values. In figure 5a the so called peak parameter of the connectionist quantizer is slightly increased. This yields a denser map of smaller regions and adds new regions around 'difficult' rhythms. In the Longuet-Higgins model we can achieve a similar change by decreasing the tolerance. Now the model will behave more 'precisely' and new solutions and small regions around them emerge.

Cognitive interpretation

We really need to compare these maps now to the corresponding maps for the human listener to be able to judge the cognitive validity of the models. In principle it is possible to obtain this empirical data in categorical perception experiments, presenting subjects with temporal sequences from the space in a transcription task. But mapping out the whole space will be a paramount task, even for such short sequences. Data about the borderlines between some regions can be found in Schulze (1989) and Clarke (1987).

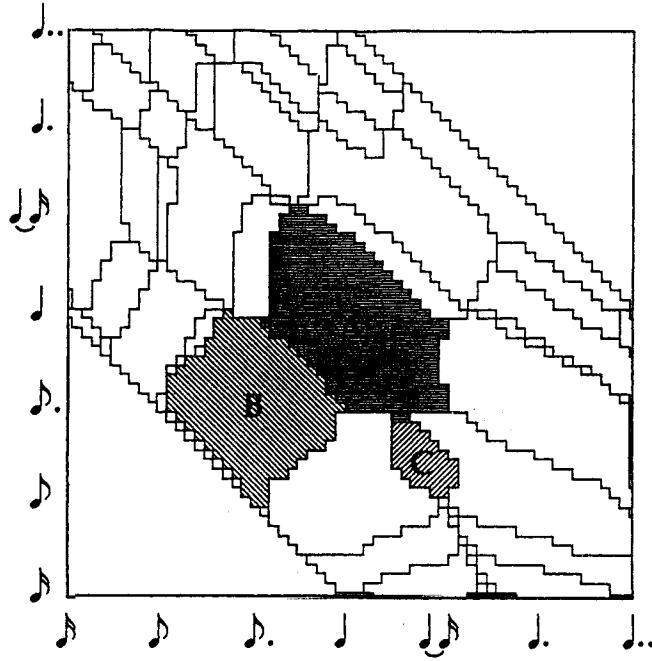


Figure 3a Regions in rhythm space of the connectionist model

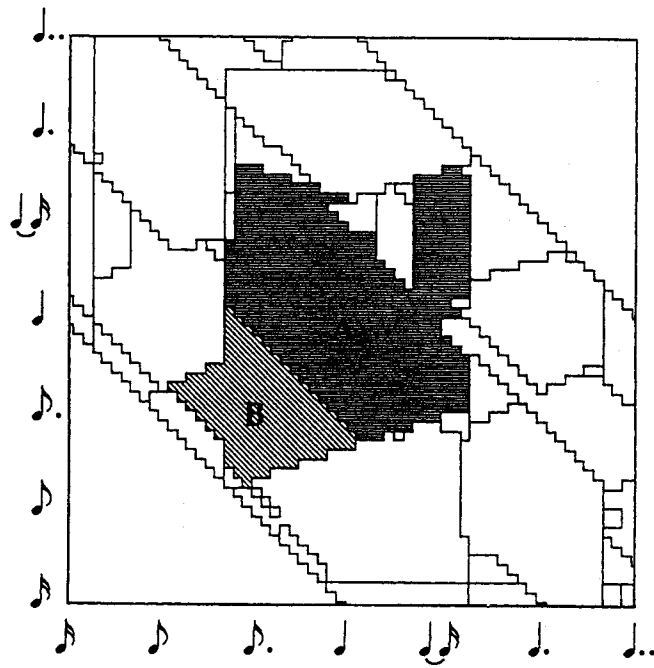


Figure 3b Regions in rhythm space in the traditional AI model

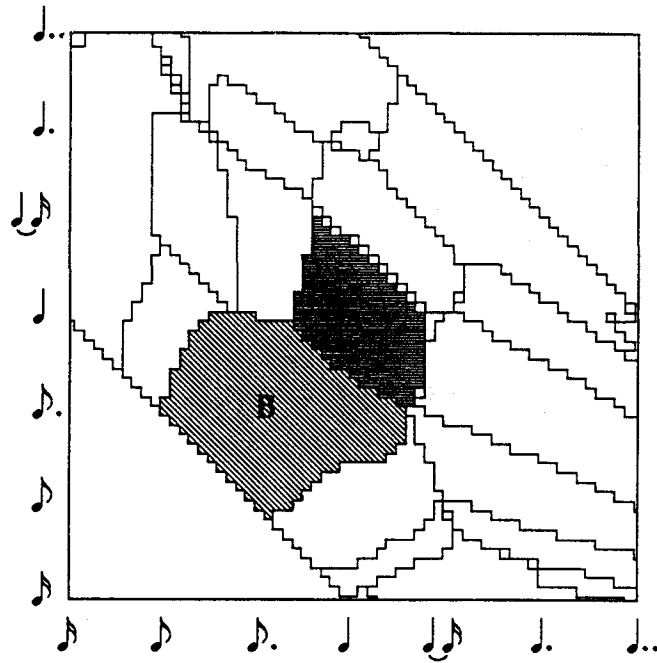


Figure 4a Influence of context (two dotted quarter notes) in the connectionist model

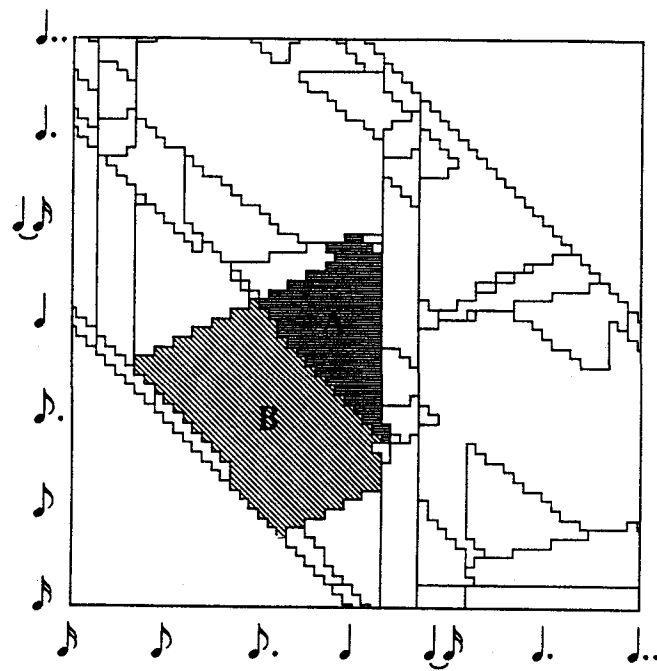


Figure 4b Influence of context (duple meter) in the traditional AI model

Expectancy space perspective

The previous representations were based on the abstraction of a whole temporal sequence that served as input of the system. Since the full models work incrementally, a representation that makes explicit how a previously established context influences future decisions would be useful. We have to ignore here any influence of new incoming data back to the previously processed results, which is a reduction for both models. In the full process model of the connectionist quantizer we can 'clamp' the whole of the network state to the partial solution obtained and study what would happen to a new incoming onset. This virtual new onset, acting as a measuring probe, will be moved by the model to an earlier or a later time. If it is given a positive time shift to a later time, the model clearly had not yet 'expected' an event. If we postulate a measure of expectation of an event, it has to be larger at a later time for this 'early' event. Vice versa: a negative movement, a shift to an earlier time, indicates a dropping expectancy: the event is late. So we can integrate the movement to yield an expectancy measure. It forms a curve with peaks at places where an event, were it to happen there, would stay in place. We could also rephrase this explanation in terms of potential and energy. The potential curve projected into the future by the network is then the inverse of the expectancy. But in the context of cognitive models expectancy seems a more appropriate concept. This process of calculating an expectancy can even be done in an incremental way: the expectancy is calculated until a real new event happens, that event is added to the context, and the process starts all over again. In figure 6 this curve is shown for a rhythm in 2/4 and the peaks in between and at the note onsets clearly are positioned at important metrical boundaries. Note that for the sake of clarity the input sequence is already idealized here to a metronomical performance.

To show that these curves capture indeed an abstract property of the input data we can look at the last part of the curve in figure 6 (the last measure between time 16 and 20), and study that for different 2/4 contexts as is done in figure 7. It shows how the different rhythms project a very similar expectancy into the future. This even prompts the challenging thought that these curves constitute a kind of rhythmic 'signature' that can be compared to produce a kind of distance measure, a metric, of rhythms.

One further corroboration of the usefulness of these curves is shown in figure 8. Here the expectancies of two rhythms are compared: one in 6/8 (a division in 2 and then in 3) and the other with time signature 3/4 (a division in 3 and then in 2). The prominent peak in the curve of the first one is located at half the measure length, lesser peaks appear at 1/6 and 2/6, and at 4/6 and 5/6. The curve of the second one has prominent peaks at 1/3 and 2/3 of the measure, and somewhat less pronounced peaks at 1/6, 1/2 and 5/6. These findings clearly correspond with the musical notion of the importance of the different points in time given these meters.

For the Longuet-Higgins model it is a bit difficult to 'clamp' the internal state to a partial solution because of possible backtracking. However, no backtracking can take place across beat boundaries, and after each beat the model only propagates the established meter and the length of a beat (the tempo) to the processing of the next beat, expecting them to apply there too. So given a beat length and a meter, they will determine the points in time where notes will be found and assigned to a metrical level. Together with the resolution of this decision (the tolerance), a comparable kind of expectancy of future onset times can be postulated. Of course

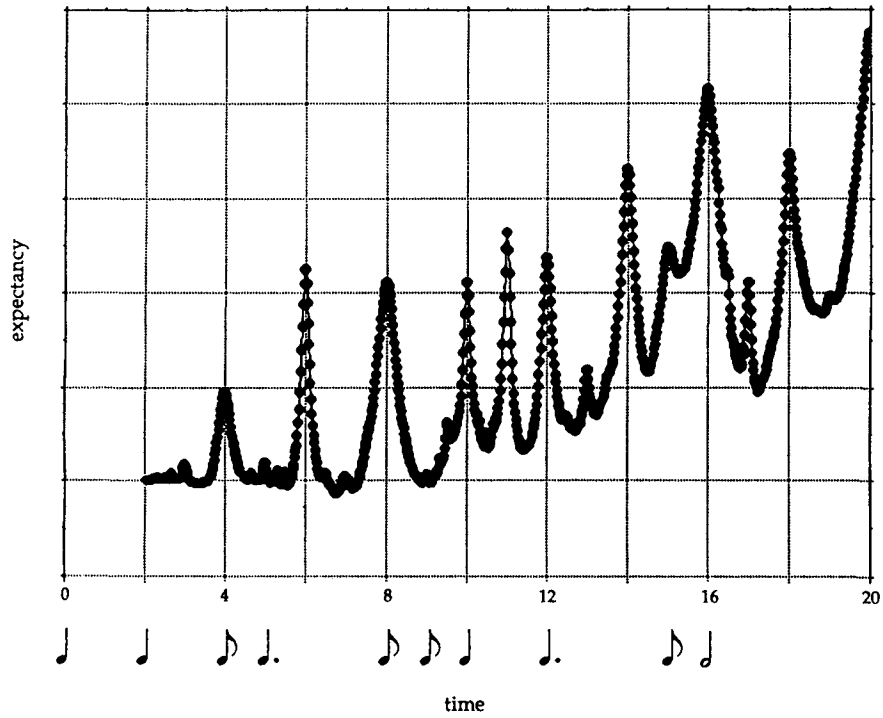


Figure 6 Expectancy of onsets in the connectionist model

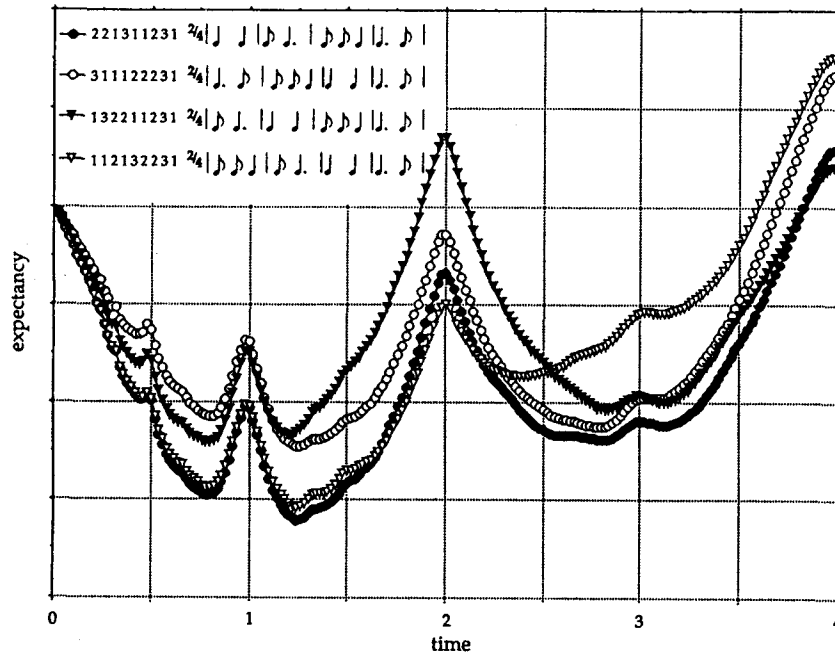


Figure 7 Expectancy of onsets in different 2/4 contexts in the connectionist model

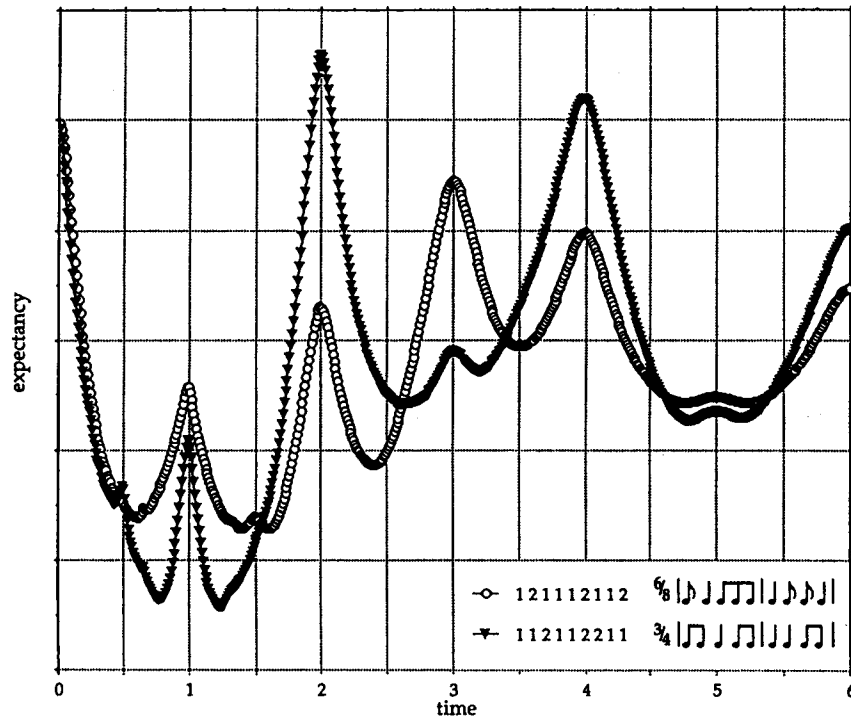


Figure 8 Expectancy of onsets in 6/8 versus 3/4 rhythm in the connectionist model

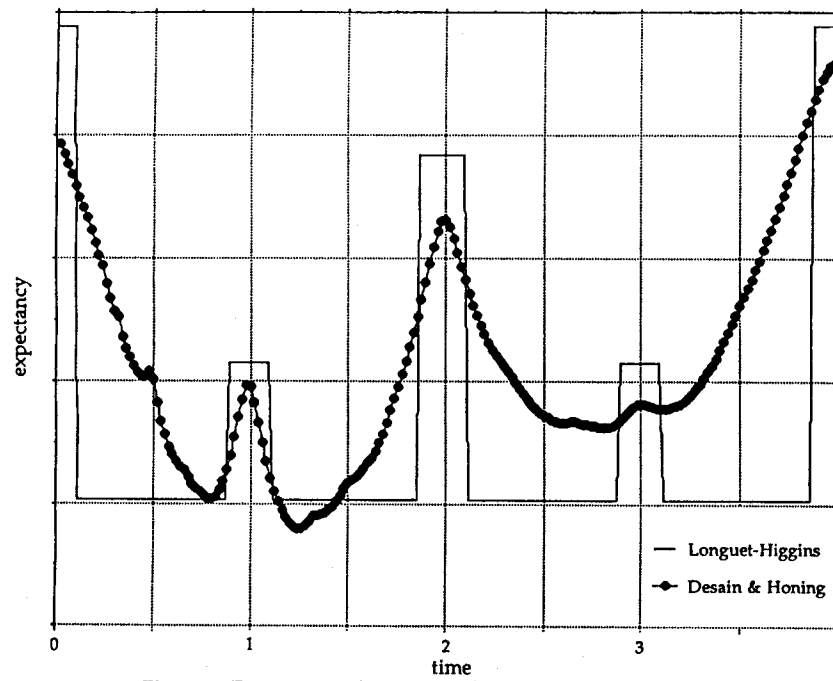


Figure 9 Expectancy of onsets in 2/4 context of the two models

the expectancy can only be given on an ordinal scale: onsets at higher metrical levels are expected 'more'. In figure 9 such a measure is shown for a twofold two-division ($2/4$ meter) and at a beat length of one bar, together with the expectancy curve of the connectionist model from figure 6. It is striking to see how the peaks in both curves now coincide. I feel that this is the point where the two models meet. Meter, a symbolic, structural concept at the very heart of the Longuet-Higgins parser, emerges out of the global and abstracted behavior of the connectionist quantizer. Here we are on the verge of the possibility of 'reading-out' symbolic representations from a sub-symbolic model.

Conclusion

It is possible to represent the behavior of two incompatible models of rhythm perception, the symbolic Longuet-Higgins musical parser and the Desain & Honing connectionist quantizer in different perspectives that make them comparable. These perspectives – the process state trace, the parameter and rhythm space, and the expectancy perspective – highlight different aspects of the models. Visualizations of these representations turned out to be crucial – even if the dimensionality or the flexibility had to be reduced.

These methods also showed the richness of the topic of quantization, a process that lies at the heart of rhythm perception. It is central because it separates two fairly different kinds of timing data: the discrete and the continuous, each of which forms the postulated input of different theories of temporal perception. It is well known that the concept of meter is of great importance in encoding, interpretation and memory in musical tasks (Palmer & Krumhansl, 1990) and it is not surprising that this symbolic concept, even though not represented explicitly in the connectionist models is still present implicitly and can emerge from the net with the help of an appropriate measuring method.

Acknowledgements

I would like to thank the colleagues who helped in this research: Eric Clarke for providing a very stimulating research environment at City University. Christopher Longuet-Higgins for encouragement and fruitful discussions about his parser. Steve McAdams for his support. Michel Koenders who automated the categorical perception experiment and Jeroen Schuijt who programmed a test suite for the quantizer, for their work and enthusiasm. Siebe de Vos and Peter van Oosten for commenting on drafts of this text. And especially Henkjan Honing for answering my midnight telephone calls.

Notes

1. The solution is the rhythm:



References

- Bharucha J.J. & Olney, K.L. (1989) Tonal Cognition, Artificial Intelligence and Neural Nets. *Contemporary Music Review*, 4.

- Boulanger, R. (1990) Conducting the MIDI Orchestra, part 1: Interviews with Max Mathews, Barry Vercoe, and Roger Dannenberg. *Computer Music Journal*, 14(2).
- Clarke, E.F. (1987) Levels of Structure in Musical Time. *Contemporary Music Review*, 2 (1).
- Clarke, E.F. (1987) Categorical Rhythm Perception: An Ecological Perspective. In A. Gabrielsson (Ed.) *Action and Perception in Rhythm and Music*. Stockholm: Royal Swedish Academy of Music, vol. 55.
- Chowning, J., Rush, L., Mont-Reynaud, B., Chafe, C., Andrew Schloss & Smith, J. (1984) Intelligent systems for the Analysis of Digitized Acoustical Signals. CCRMA Report No. STAN-M-15.
- Dannenberg, R.B. & Mont-Reynaud, B. (1987) An on-line Algorithm for Real Time Accompaniment. *Proceedings of the 1987 International Computer Music Conference*. San Francisco: Computer Music Association
- Desain, P. & Honing, H. (1989) Quantization of Musical Time: A Connectionist Approach. *Computer Music Journal*, 13 (3), also to appear in (Todd & Loy, forthcoming).
- Desain, P., Honing, H. & de Rijk, K. (1989) A Connectionist Quantizer. *Proceedings of the 1989 International Computer Music Conference*. San Francisco: Computer Music Association.
- Desain, P. & Honing, H. (forthcoming) The Quantization Problem: Traditional and Connectionist Approaches. In *Musical Intelligence* edited by M. Balaban, K. Ebcioglu and O. Laske, Menlo Park: AAAI book.
- Desain, P. (1990) Parsing the Parser, a Case Study in Programming Style. *Computers in Music Research*, 2.
- Fodor, J.A. (1975) *The Language of Thought*. New York: Crowell.
- Palmer, C. & Krumhansl, C.L. (1990) Mental Representations for Musical Meter. *Journal of Experimental Psychology: Human Perception and Performance*.
- Longuet-Higgins, H.C. (1987) *Mental Processes*. Cambridge, Mass.: MIT Press.
- Rumelhart, D.E. & McClelland, J.E. (Eds.) (1986) *Parallel Distributed Processing*. Cambridge: MIT Press.
- Schulze, H. (1989) Categorical Perception of Rhythmic Patterns. *Psychological Research*, 51.
- Todd, P. M. & Loy, D.G. (Eds.) (forthcoming) *Music and Connectionism*. Cambridge, Mass.: MIT Press.