A (De)Composable Theory of Rhythm Perception

PETER DESAIN
Nijmegen University, The Netherlands

A definition is given of expectancy of events projected into the future by a complex temporal sequence. The definition can be decomposed into basic expectancy components projected by each time interval implicit in the sequence. A preliminary formulation of these basic curves is proposed and the (de)composition method is stated in a formalized, mathematical way. The resulting expectancy of complex temporal patterns can be used to model such diverse topics as categorical rhythm perception, clock and meter inducement, rhythmicity, and the similarity of temporal sequences. Besides expectancy projected into the future, the proposed measure can be projected back into the past as well, generating reinforcement of past events by new data. The consistency of the predictions of the theory with some findings in categorical rhythm perception is shown.

Introduction

Many incompatible theories about temporal perception and memory exist, which explain a number of phenomena well, but fail to predict others. A common theoretical basis for such work would be desirable. Connectionism might be an attractive paradigm in the search for such a basis, but most of its models lack compositionality. This means that the model as a monolithic whole might perform well, but it is impossible to decompose its complex behavior into meaningful smaller parts. Chandrasekaran (1990) argues that the composability is a condition for successful cognitive modeling, even in the connectionist paradigm. In Desain (1990), the behavior of a subsymbolic (connectionist) model of temporal quantization was described such that it could be compared with an incompatible symbolic model from the traditional artificial intelligence paradigm. The paper concluded with an abstraction of the behavior of the quantizer in the form of an “expectancy of events” with a temporal pattern as prior context. Expectancy turned out to be (de)composable, which makes it possible to base a theory of perception of complex stimuli on a simple model for the perception of their constituting components. Be-
cause the expectancy concept seems to explain the dependence of perception of rhythmic structure on global tempo, the influence of context on categorical perception, and other complex phenomena, I propose to use it as a common basis for theories about temporal perception and memory. It is noteworthy that Povel (1984, 1985) has already remarked that high-level cognitive judgments like rhythmicity might be based on, or be a byproduct of, low-level rhythm perception processes that deal with quantization and tempo tracking.

The experience of rhythmicity is supposed to result from the process that updates the internal clock in the light of the incoming stream of events. It may be noted that this process makes part of the normal process of listening to music in which the listener constantly adjusts his internal clock (metrical frame) to local temporal irregularities and tempo variations. (Povel, 1984, pp. 19–20)

In this paper, I will focus on the explanation and formalization of the theory and the composability of the definition of expectancy. The interpretation of the resulting curves, their possible use, and their relations to other research aimed at a higher level of rhythm perception will be dealt with in Desain (in preparation). Although the theory looks attractive enough, I must warn the reader that this paper is an account of recent work that has yet to be empirically verified.

The Connectionist Quantizer

To illustrate how the theory developed from a connectionist approach to the quantization of temporal sequences (Desain and Honing, 1989), a brief overview of that system will be given here. One by one the interonset intervals of a performed sequence are passed to a network of cells. The network acts as a complex shift-register: new interonset intervals shift in, are processed on the way through, and then shift out of the network as quantized durations (rhythmic categories). Besides cells for performed time intervals, there are cells for intervals spanning several basic intervals. Two cells interact if they represent neighboring time intervals: one ending where the other starts. Their interaction is such that the ratio of the two intervals is adjusted towards an integer, if the ratio is already close to this goal. The change in length of each interval is proportionally propagated to all the basic intervals that form part of it. This interaction proceeds until a new interonset interval enters the network. All data are then shifted one position and the interaction process resumes.
Expectancy

To make the link from the updating of time intervals, as is done in the connectionist quantizer, to expectancy, we can study what would happen to an imagined new incoming event whose corresponding time interval has just been shifted into the network. Its change is completely determined by the context of a temporal pattern that was presented before. When all but the last new interval in the network are clamped to a fixed state, we can study the change of the new interval effectuated by the context while ignoring the influence of the new interval on the already perceived, but not yet fully processed data. The quantizer then can only effect a change in duration of this new basic interval. The imagined new onset ending this interval, now acting as a kind of measuring probe, will be moved to an earlier or a later time by the interactions. If the interval is increased, moving the onset to a later point in time, the model clearly has not yet “expected” an event. If we postulate a measure of expectancy for this “early” event, it will be larger later. Conversely, a decrease of that time interval, a movement of the new onset to an earlier point in time, indicates a falling expectancy: the expectancy at an earlier time was larger. Finally, if the context inflicts no change to the new onset, expectancy is constant at that time.

We can thus view the change of an imagined new onset time as the slope of an expectancy measure of an onset at that time. The pattern of expectancy forms a curve with extremes at places where an event, were it to happen there, would stay in place, because the derivative of the curve at this point (which is the change) is zero. The local maxima form the expected, “perfect” places of onsets, and the local minima form points of maximal confusion, places of unexpected events. Take, for example, the pattern [3,1,2,x]. If x is just below 1, the quantizer network would effectuate a positive change. If x is equal to 1, the change would be zero. If x is just above 1, the quantizer would adjust it downward, a negative change. Around 2 a similar situation occurs, but the changes are more pronounced because of the strong 1:1 interaction between the last two intervals. Around 3, 6, and even 12, the pattern is similar as well: positive below, zero on, and negative above that value. If we now integrate the change over x, an expectancy curve results with local maxima at the values 1, 2, 3, etc. Figure 1 shows some expectancy curves that were measured with longer temporal patterns represented as prior context in the network. They all have the same global characteristic, 2/4 meter, and produce very similar curves. Note the prominent peaks at the half-bar and bar boundary (at time 2 and 4, counting in eighth notes) and lesser peaks dividing these intervals further.
Expectancy is defined as the integral of the change generated by the sum of all interactions in the network. We can exchange integration and summation in this formulation and redefine the expectancy, given a prior complex pattern, as the sum of all the expectancies generated by each interval in that pattern. This effectively decomposes the theory for processing complex rhythms into a set of simple components, one for each time interval implicit in the pattern. Figure 2 is an attempt to depict this kind of decomposition. At the bottom left, the presented temporal pattern is shown. Above it, all the intervals implicit in this pattern are indicated (by bold lines). To the right of each interval, the pattern of expectancy projected by that interval is shown (light lines). The basic expectancy is a function of two parameters: the length of a time interval and the time elapsed after the end of that interval. It peaks when the second parameter is an integer divisor or an integer multiple of the first. All the projected basic expectancies are summed and yield the curve at the lower right: the global expectancy curve projected by the complex temporal pattern. It has
peaks at time points that can be considered as “good continuations” of the pattern presented.

This concept of expectancy seems closely related to ideas of Jones, and could function as a formalization of these.

It is such psychological trajectories that rhythmically guide our attentional energies along ideal paths. Attention is cast from some reference event at one point in time toward a target event scheduled for a later time. This approach demonstrates that attention itself is a dynamic, many levelled affair based upon nested internal rhythms. We continually cast ourselves forward by rhythmically anticipating future events that may occur within small and larger time intervals. These paths form the patterns of mental space and time and so can establish for us that sense of continuity and connection that accompanies comprehension. (Jones, 1981, p. 571)

A more precise formulation of this principle of decomposition, which constitutes the core of the theory presented in this paper, will be given later. But first it is interesting to consider what happens if the grain of analysis is made a bit more coarse by lumping together the expectancy contributions of intervals that end at the same point in time. This gives a decomposition based on events instead of intervals. In Figure 3, the contributions of each new event are shown, incrementally building the total expectancy. The expectancy evolves during the presentation of the
pattern and extends from the end of the pattern into the future, as was shown in Figure 2. This kind of curve can show how a temporal pattern fails to realize a high expectancy (a syncope), or comes up with an unexpected event. It also enables one to see how a new event reinforces the already existing pattern of future expectancy or introduces new elements in it.

Now we will make a slight detour to the concept of memory. According to Mari Riess Jones, expectancy and memory are closely related:

Paradoxically, a third implication of including time as a part of subjective structure results in an alternative view of memory. This new view casts remembering as a dynamic attentional process unfolding in negative time. That is, we can conceive of both expectancy and remembering as activities tied to the time dimension. . . . Expectancy and remembering then are opposite sides of the same coin. (Jones, 1981, pp. 574–575)

These rather puzzling remarks become clearer when one studies the influence that a new incoming event might have on the prior context. The

![Fig. 3. Expectancy contributions of events in a temporal pattern.](image-url)
new event can support one or the other of the previous interpretations and in retrospect contribute to a limited extent to the salience of already perceived stimuli. To visualize this, we can simply construct all the new intervals created by a new incoming event and project their expectancy of events into the past (see Figure 4). This gives the amount of reinforcement given to each event in the pattern by the new incoming event. Temporal patterns that behave in a well-formed way, with high support of events by later ones, might be remembered better. This model thus predicts how a later event can facilitate or inhibit the memory of past ones, a rather spectacular feature. An example of this phenomenon is the often encountered "closure" of a temporal pattern, which concludes with an event in a highly expected place: an important metrical position.

Concluding, we can state that the concept of expectancy as presented here has no time direction in itself. It is determined completely by two time intervals, but whether each of the time points marking the intervals was presented as stimulus in the past, or has yet to happen in the future, is irrelevant to the theory. One can thus speak about expectancy of an event at a future time generated by two time points in the past, the reinforcement of an event in the past by two time points happening later, or even the expectancy of an event at a certain time in between two time points. All these notions are equivalent at this level of the theory and yield the same numerical values if the distance between the first and second time point is the same as the distance between the second and third time point in the three cases. This does not imply that the theory is symmetric with respect to time: swapping the distances between the first and second and between the second and third time point might yield different values, because the perception of a time interval followed by a multiple of its length might

Fig. 4. Relation of expectancy and memory.
be different from the perception of that interval followed by a division of it by the same factor.

Besides the ratio of the two time intervals, the basic expectancy function is supposed to depend on the absolute time duration of both of its parameters as well. This makes the theory sensitive to the time scale used (the absolute tempo). In Figure 5 the same pattern as in Figure 2 is used, but at half the tempo (note that for the sake of easy comparison the horizontal axis is "renormalized" such that the figures have the same size). It is clear that what is often called "the shift of level of attention through the levels of metrical hierarchy, prompted by different tempi" can be found here in the shift in relative importance of expectancy peaks.

We now get back to the details of the model. The shape of the basic expectancy curves is the part of the model that still has to be "plugged in" to yield the full theory. The theory is "generic": given any method for calculating the basic expectancy of a time interval pair, it defines the method to calculate expectancies for any complex temporal pattern. I will discuss a first approximation of these basic expectancy curves and the possibilities of deriving them empirically.

![Diagram](image-url)

Fig. 5. Expectancy at slow tempo.
Preliminary Theory of Basic Expectancy

Before its mathematical formulation, I will first give a graphic representation of the proposed basic expectancy. We can depict the basic expectancy of a time interval pair \((A, B)\) for fixes \(A\) as a curve (see Figure 6). One can see clear peaks in expectancy when \(B\) equals \(A, 2A, 3A \ldots\) and when \(B\) equals \(A/2, A/3 \ldots\). The shape of the expectancy curve is determined by our capacity to perceive serial duration ratios, with higher ratios being more difficult and less expected (Jones & Boltz, 1989). This may also be true for more complex ratios in terms of their prime divisors. These curves, projecting expectancy into the future, were used for Figure 2. Another visualization is given in Figure 7. It shows the expectancy of a subdivision of a unit time interval into the interval pair \(A, 1 - A\). One can see here that the time-reversed interval pairs are still assigned the same expectancy in this preliminary theory (e.g., the pair \(1/3, 2/3\) and the pair \(2/3, 1/3\)).

Sternberg, Knoll, and Zukofsky (1982) showed that perceptual judgment is dependent both on the ratio of the intervals involved and on their absolute duration. Very long and very short time intervals are difficult to perceive accurately. The maximum in sensitivity occurs at about 600 msec, which is in the preferred tempo range (Fraisse, 1982). The total length of the interval pair will be used as a second determinant of the expectancy to model this dependency on the absolute time scale.

Because the nationwide polynomials used in Desain and Honing (1989) are a bit difficult to treat mathematically, basic expectancy is defined as a sum of several Gaussian curves, one around each relevant ratio.
The parameters of the Gaussians are determined by the ratio and the absolute tempo (the size of the sum interval in proportion to the preferred tempo). These last two values ($R$ and $S$) determine the height (via function $C$) and the width (via function $D$) of the expectancy curve peak at each integer ratio.

**Measuring Basic Expectancy**

To be able to proceed from a theory to a tested cognitive model, we need a way to operationalize and measure basic expectancy. Although the material under consideration is very simple (just two successive temporal intervals), collecting empirical data on perceived expectancy might still be quite difficult. Carolyn Drake (personal communication, 1990) has proposed a measure of accuracy in an adjustment task as used in her work on accents (Drake, Botte, & Gérard, 1989). Goodness-of-fit judgments and memory confusion in discrimination judgments as used by Palmer and Krumhansl (1990) should also be considered. A more indirect measure might be derived from calculating the probabilities of time interval pairs $A, B$ in a body of musical pieces. This is not the same as the frequency...
A (De)Composable Theory of Rhythm Perception

449

counts approach used by Palmer and Krumhansl (1990), as the latter makes use of a priori knowledge of meter.

Complex Expectancy

When the basic expectancy function $E_b(A,B)$ is given, either by measurement or construction, the expectancy generated by an interval in a complex temporal pattern can be derived. Let $X$ be a vector of basic time intervals $X_i$ ($1 \leq i \leq N$) and $S(X,p,q)$ the time period spanned by intervals $p$ through $q$.

$$S(X,p,q) = \sum_{i=p}^{q} X_i \quad \text{with } p \leq q \leq N \quad (3)$$

When $X$ contains the interonset intervals of a temporal fragment presented from time 0, we can define the interval expectancy $E_i$ of a new onset at time $T$, generated by the interval spanning interonset intervals $p$ to $q$ in $X$, by applying the basic expectancy function $E_b$ to the relevant time intervals.

$$E_i(X,p,q,T) = E_b[S(X,p,q), T - S(X,1,q)] \quad \text{with } T \geq S(X,1,q) \quad (4)$$

The complex expectancy $E(X,T)$ of an event at time $T$, generated by the temporal pattern $X$ presented from time 0, is then a sum of the interval expectancies $E_i$ over all intervals implicit in that pattern:

$$E(X,T) = \sum_{p=1}^{N} \sum_{q=p}^{N} E_i(X,p,q,T) \quad \text{with } T \geq S(X,1,N) \quad (5)$$

Figure 8 has the same structure as Figure 2, but it labels the relevant time intervals according to the formalism given above. The event-based expectancy of Figure 3 and the concept of reinforcement of Figure 4 may be formalized in an analogous way.

Because the basic expectancy $E_b(A,B)$ will be small for large $A$ or $B$, there is a natural limit to the size of the context that will contribute to the total expectancy $E(X,T)$, and the vector $X$ can function automatically as a short-term memory construct.

A nice consequence of the bidirectionality of the expectancy concept is that the sum of the corroborations of each event in a pattern by a virtual new onset is the same as the expectancy of that onset generated by the pattern.
Fig. 8. Time intervals used in calculating expectancy.

Related Work

The measure of expectancy presented here can be related to a number of different theories and concepts. The Gestalt principle of Good Continuation, which is one of the underlying assumptions in much research in grouping mechanisms (Deutsch, 1982), can be linked closely to the expectancy construct. It is tempting to interpret the relative height of the peaks in the expectancy curve directly as a measure of metrical boundary strength. However, instead of deriving a symbolic notion of meter from these curves, it might be more productive to rethink the concept of meter as a continuous concept, an idealized expectancy curve, as discussed in Desain (in preparation). This allows the construct to be applied directly to the difficult areas of ambiguous rhythms, change of meter and amount of metricality. The expectancy curves are also promising for the study of the perceived rhythmicity of temporal patterns and their degree of syncopation (Povel, 1985). This is because it is easy to formalize the violation of a maximum in expectancy by the absence of an event at that time.

Given the numerous links with different aspects of the literature on rhythm perception, I will restrict myself here to some remarks on the predictions of the model for categorical rhythm perception.
Categorical Rhythm Perception

In general, categorization occurs when objects, on the basis of some continuously variable attribute, are placed in a small number of groups (see Repp, 1984). Sloboda argues for the existence of categorical perception in rhythm by noting how different the perception of rhythm and the perception of expressive timing are:

Identification of intended rhythm is a commonplace accomplishment for listeners, who are continually faced with the potentially confusing phenomenon of rubato and gradual changes in speed. In contrast, accurate perception of deviations from metricality is difficult, and requires much specific training. It is almost impossible for one performer to imitate another exactly. All this strongly suggests that listeners achieve a categorization of the duration of the notes they hear into crotchets, quavers etc. . . . one would not wish to claim that categorical perception makes finer temporal discriminations impossible. We can hear rhythmic imprecision and rubato with appropriate training, but fine differences in timing are more often experienced not as such, but as differences in the quality (the "life" or "swing") of a performance. (Sloboda, 1985, p. 30)

Thus, in quantization the deviations from a strict metrical performance are not thrown away, but timing is separated into structural and expressive components and then handled by different processes.

A discussion of the use of expectancy curves for a categorical perception model (a quantizer) will be left aside here, since it involves many technical points about the architecture of static versus process models, the use of global tempo tracking, etc. The models proposed in Desain and Horing (1991, addendum) use hill-climbing in an expectancy landscape, but they vary in the extent to which the different dimensions ("older" and "newer" time intervals) are allowed to vary. The following material will discuss the expectancy curves themselves, assuming that if local maxima and minima emerge they can be used somehow to segment the continuous time axis into discrete regions, one for each rhythmic category.

Although categorical perception is a well-established phenomenon in speech research, its existence is much harder to demonstrate in the rhythm domain. In a well-known set of experiments, skilled musicians were asked to judge the length of temporal intervals in rather simple integer ratios (Sternberg, Knoll, & Zukofsky, 1982). Surprisingly, they were not able to do this accurately, and boundaries between different categories turned out to be rather vague. Schulze (1989) was somewhat more successful in showing the existence of a categorical boundary in interpolations between patterns of three interonset intervals. In Clarke (1987), subjects were given
a context of five or six interonset intervals in different metrical contexts, before the two experimentally manipulated durations were presented. He was able to show clear identification and discrimination curves that also showed the predicted shift for the different context conditions. The general line in these findings thus seems to be that categorical perception is facilitated by context. This is in agreement with the expectancy theory presented above: the expectancy curves become more pronounced if more context is available. In Figure 6 one can see that in an impoverished context the perception of simple patterns like \([3, 2]\) is not possible anyhow: there is no peak in expectancy at \(B = 2/3A\). A bit more context will allow perception of a 2/3 ratio as is shown in Figure 9, where the expectancy of an onset after the context \([2, 3, 3]\) is given. Note how the local maximum at \(B = 2\) emerges here. This general idea is consistent with the findings of Povel (1981) concerning an imitation task.

**Conclusion and Discussion**

The presented theory of expectancy seems a promising candidate for a common basis for many incompatible theories of rhythm perception and memory. Its decomposability into simple components that model perception of time-interval pairs is attractive, not in the least because empirical results for simple stimuli can be “plugged” into the theory to yield predictions for more complex temporal patterns. The theory elegantly links expectancy projected into the future and reinforcement of past events by new data. Predictions following from the theory are consistent with
some findings in categorical rhythm perception. Empirical verification of the theory will be the next step that is needed to further the research in this direction.

Another field in which work needs to be done is the formalization of the use of expectancy in the different cognitive processes mentioned above, such as quantization, meter and beat inducement, rhythmicity and similarity of rhythms. It is clear that a full theory of rhythm perception cannot be based on time alone but has to take other musical parameters into account as well. A possible approach could be the use of a notion of salience to weigh expectancy contributions of events.

It still is an open question whether an expectancy pattern is available as a whole for input to other processes, or whether expectancy is merely a changing sense of present anticipation and no access to future expectations is possible.1

References


---

1 I would like to thank Piet Vos for organizing the Rhythm Perception and Production Workshop in Horssen and Ian Cross for his Music and Cognition Conference in Cambridge, where many of these ideas were born. Research done with Eric Clarke at City University started my interest in issues of expressive timing, it was very stimulating to work with him. Johan den Biggelaar of the Utrecht School of the Arts did his best to provide me with research facilities there. I also would like to thank Carolyn Drake, Mari Riess Jones, Jeroen Schuit, Marie-Jose Tienhooven, Siebe de Vos, and Caroline van der Wal, who gave many helpful comments on this paper. As always, Henkjan Honing prevented me from getting sidetracked from the main issues. Without him, I would have had less than half the fun in contriving these ideas.