

The Formation of Rhythmic Categories and Metric Priming*

Peter Desain¹ and Henkjan Honing^{1,2}

Music, Mind, Machine Group

NICI, University of Nijmegen¹
P.O. Box 9104, 6500 HE Nijmegen
The Netherlands

Music Department / ILLC²
University of Amsterdam
Spuistraat 134, 1012 VB Amsterdam
The Netherlands

Running Head: Formation of Rhythmic Categories

[Published as:

Desain, P. & Honing, H. (2003) The formation of rhythmic categories and metric priming.

Perception. 32(3), 341-365.]

* Both authors contributed equally to this work

[This is Desain & Honing's 75th co-authored paper]

0. Abstract

This paper presents two experiments on categorical rhythm perception. It investigates how listeners perceive discrete rhythmic categories while listening to rhythms performed on a continuous time scale. This is studied by considering the space of all temporal patterns (all possible rhythms made up of three intervals) and how they, in perception, are partitioned into categories, i.e. where the boundaries of these categories are located. This process of categorization is formalized as the mapping from the continuous space of a series of time intervals to a discrete, symbolic domain of integer ratio sequences. The methodological frame work uses concepts from mathematics and psychics (e.g., convexity and entropy) that allow for precise characterizations of the empirical results.

In the first experiment 29 participants performed an identification task with 66 rhythmic stimuli (a systematic sampling of the performance space). The results show that listeners do not just perceive the time intervals between onsets of sounds as placed in a homogeneous continuum. Instead, they can reliably identify rhythmic categories, as a chronotopic time clumping map reveals. In a second experiment the effect of metric priming was studied by presenting the same stimuli but preceded with a duple or triple meter subdivision. It is shown that presenting patterns in the context of a meter has a large effect on rhythmic categorization: the presence of a specific musical meter primes the perception of specific rhythmic patterns.

1. Introduction

Time, as a subjective structuring of events in music, is quite different from the concept of time as duration in the physical realm (Michon & Jackson, 1985; Jones, 1990; Large & Jones, 1999). Listeners to music do not perceive the time intervals between onsets of sounds as being placed in a homogeneous continuum. Instead, they identify rhythmic pattern, categories that function as a reference relative to which deviations from strict mechanical timing can be appreciated (Desain & Honing, 1992; Clarke, 1999; Desain & Windsor, 2000). In fact, temporal patterns in music combine two representations of time that are essentially different: the discrete rhythmic durations as symbolized by, for example, the notes in a musical score, and the continuous timing variations that characterize an expressive musical performance. In music performance, timing information is added to the nominal durations of the note categories, based on the interpretation of the musician (see Figure 1a). These deviations are experienced as the expressive character of the performed rhythm (e.g., ‘anticipated’ or ‘swinging’) and they are to a large extent related to the structure of the music, such as musical phrases, meter or rhythmic structure. These structural and expressive properties of musical rhythm have been studied extensively (Palmer, 1997; Clarke, 1999; Gabrielsson, 1999). In the perception of music (see Figure 1b), the listener separates this temporal information into rhythmic categories and expressive timing. For example, listeners will recognize the performed rhythm shown in Figure 1a as the one notated in Figure 1b (see Footnote 1). Next to the recognition of these discrete rhythmic durations, a listener still perceives the expressive timing of the performed rhythm. Even untrained listeners appreciate the ‘against the beat’ quality of the triplets (Vos & Handel, 1987) and the natural slowing down at the end of the group of sixteenth notes (Todd, 1992). And one can argue that, to be able to perceive a long note as a local slowing down of tempo, one needs a reference, which by itself suggests the existence of rhythmic categories.

<<Insert Figure 1 around here>>

The first question to be addressed in this paper concerns the mapping of the continuous space of performed temporal patterns to the discrete, symbolic space of rhythmic categories (the mapping from Figure 1a to 1b). This issue has been treated in a

relatively small number of studies on categorical rhythm perception (Clarke, 1987; Schulze, 1989). For a few special cases a sharp transitional boundary between rhythmic categories has been found, with a high sensitivity for differences around this boundary. These studies will be reviewed in more detail in Section 2.5. The second question to consider deals with the factors that influence this mapping, in particular the metrical context in which the performed patterns are presented. With regard to this influence of meter, Clarke (1987) has presented some evidence showing that the position of the boundary between rhythmic categories may shift depending on the metrical context in which the rhythm is presented. In both Clarke's and Schulze's study, however, only few categories were investigated. In contrast, in this study, we systematically consider a large set of temporal patterns and the effect of meter on their identification.

In this paper, that is a result of inter-disciplinary research, we use a number of concepts and analysis tools that are relatively new to rhythm perception research. Some of them have already been of use in other fields of perception, for instance, the use of maps in the study of color vision (Le Grand, 1968). Others stem from physics (e.g., the concept of entropy) or mathematics (e.g., the notion of abstract spaces and of convexity). In introducing their use we hope to demonstrate how they help to arrive at the abstract levels of understanding needed to build general theories of human rhythm perception, which is a beautiful, and beautifully complex, process.

2. Rhythmic Categories and Expressive Timing

2.1. Categorization

We use the term categorization to describe the cognitive process of extracting the discrete rhythmic categories from the continuous signal. The continuous time intervals in music performance are not just categorized into symbolic categories: the categories themselves have structure, as they relate to each other as rational numbers. They can be represented and coded by small integers, each signifying a multiple of a small symbolic duration. Thus, because of the suspected context dependency, we will consider categorization of temporal sequences as a whole, not of individual performed time intervals. We expect the series of categories that result from this cognitive process to exhibit structural regularities.

In the rhythm perception literature there are different approaches to this phenomenon. Fraisse (1982), for instance, stressed the importance of low integer ratio's (like 1:1 and 1:2) in the perception of rhythm, ratio's to which non-integer rhythms will migrate. Other authors (e.g., Nakajima, 1987) suggest that categorization is in fact a mapping from a single continuous time dimension into categorical intervals, independent of context. Still other research (e.g., Clarke, 1987) investigated whether this categorization might in fact be a result of categorical perception (we will return to this topic in Section 2.2). However, the implication of true categorical perception is that expressive timing would be barely detectable, which clearly is not the case (Clarke, 1999). In contrast, it has been argued (e.g., Clarke, 2000) that categorization is not simply a mapping from a continuous variable to a discrete one (loosing important continuous information in the process), but that both types of information (i.e. the rhythmic category and expressive timing) are available at the same time. But we will argue that expressive timing is only perceivable because there is categorization, the categorization functioning as a reference relative to which timing deviations are perceived. In this view both types of information are available to the listener, with categorization determining the expressive timing perceived.

In the categorization of a time interval it has been shown that neighboring intervals play an important role (Desain & Honing, 1991). As an example, consider the two marked () time intervals in Figure 1. While the first note is played slightly shorter than the second, the former is commonly interpreted as 1/3 times longer (i.e., the first as part of a triplet, the second as a sixteenth note). This typical example illustrates that categorization is a cognitive process that cannot be described by some simple round-off procedure: the local temporal context has to be taken into account to explain the perceived duration of the time interval. Furthermore, not only is there dependency on context in rhythm perception; without context categorization is hardly possible. Sternberg, Knoll & Zukofsky (1982) showed that even experienced musicians have difficulty in recognizing and reproducing relatively simple ratios like 1:5 or 3:4 if they are presented in isolation. While in music practice these ratios (and even much more complex ones) are quite common, one always needs some context to be able to recognize and perform them well.

2.2. Categorical Rhythm Perception

Categorical rhythm perception (Clarke, 1999) has been studied by presenting interpolations between different rhythmic patterns to listeners in both an identification task (probing the recognition of categories) and a discrimination task (testing for increased sensitivity near category boundaries). As such it applies the paradigm developed for categorical perception (Harnad, 1987) as used, for instance, in the domains of speech (Repp, 1984) and color perception (Saunders & van Brakel, 1997). While categorical perception experiments in these domains have sometimes been used to determine whether this process is innate or learned, it was shown by Clarke (1987) that, at least for the rhythms studied, it is open to top-down cognitive influence. Clarke (1987) describes an experiment in which a short musical sequence was presented in two different metrical contexts (2/8 and 3/8 meter, i.e. a duple and a triple meter), with the two notes at the end of the sequence systematically varied between the ratios 1:1 and 1:2 (see Figure 2a) (see Footnote 2). The participants performed an identification task in which they had to identify the rhythm as belonging to type 1:1 or type 1:2, and a discrimination task in which they judged whether a pair of rhythms was same or different. The resulting identification function showed a strong change in slope at the category boundary between the two rhythms and the discrimination function has a strong peak in the same position, which, as such, is clear evidence for categorical perception. Schulze (1989) did a follow-up study addressing some of the methodological problems of Clarke's study, the main point being that the forced-choice paradigm used in the identification task steered the participant's responses towards the available categories. Schulze therefore used a somewhat different experimental setup in which he trained two participants with a set of interpolated rhythms, and asked them to give a graded identification response (i.e., as many response categories as stimulus types). The discrimination function was derived indirectly from these responses. He investigated interpolations between four different 'prototypical' rhythms (see Figure 2b). The patterns were not preceded by a metrical context, as in Clarke (1987). However, he found similar effects, although weaker for some rhythmic patterns. Furthermore, he showed that categorical rhythm perception is open to perceptual learning (cf. Livingstone, Andrews & Harnad, 1998), the participants being able to distinguish more categories after an intensive training period prior to the experimental trials.

<<Insert Figure 2 around here>>

These studies provide evidence for the categorical perception of rhythmic patterns. Their local character leaves open the question to what extent the results can be generalized for the large variety of patterns occurring in real music, and how the shape and position of these categories are affected by meter (a boundary shift was only shown for the rhythmic categories 1-1 and 2-1). In our study, we will therefore conduct a systematic categorization study and use a large set of temporal patterns as stimuli, patterns of four note onsets and a fixed total duration of one second (sampled on a fine temporal grid). These will be presented without metrical context (Experiment 1) and in three metrical contexts (no meter, duple and triple meter; Experiment 2), as the participant engages in an identification task with semi-open responses.

To study the process of categorization as described above, we need a formal and computational framework. For this we will introduce the abstract notion of rhythm spaces and conceptualize categorization as a mapping operation between them. Consequently, these notions will be used in the analyses of the experimental results.

2.3. Representing Temporal Patterns: The Performance Space

Instead of studying an arbitrarily chosen set of rhythmic patterns, we consider the space of all possible performances of n time intervals. In this n -dimensional space every point represents a different temporal pattern. This infinite set contains both musical and unmusical rhythms, and captures, in principle, all possible performances of all rhythms of $n+1$ onsets. We will therefore refer to this set as the performance space. Restricting ourselves to four onsets, any pattern can be represented in a three-dimensional space (see Figure 3a), with the three axes representing the three inter-onset intervals (IOIs). All patterns that add up to a fixed total duration form a diagonal triangular slice in such a space. Looking from above, towards the origin, the triangle can be presented as a ternary plot (see Figure 3b).

<<Insert Figure 3 around here>>

This ternary plot, that depicts dependent three dimensional data, still allows for the determination of the coordinates of the points: a specific performance can be located by reading the grid along the direction of the tick marks at the axes, with interval-size (or note duration) depicted clockwise and interval-number depicted counter-clockwise. For example, the point labeled A corresponds to the pattern (shown in Figure 3c), its first IOI being 0.25, the second being 0.5 and the last IOI 0.25 seconds (see Footnote 3). We will use this chronotopic map to represent our stimuli and the way they are perceived.

As an example, and for comparison with the empirical data discussed in Section 2.2., Figure 4 shows the stimuli used in Schulze (1989) depicted as chronotopic map. The black dots identify the stimulus patterns, and interpolations between their mechanical versions are marked by crosses. The category boundaries are indicated with a gray line. The gray area is the hypothetical shape of the rhythmic category A. However, we can only base its contours on two measurements (the boundaries between A and B, and A and C), to few to infer its shape.

<<Insert Figure 4 around here>>

2.4. Representing Rhythmic Categories: The Score Space

The syntactic aspects of rhythm can be formally described by considering rhythm to be the result of a metrical grammar (Longuet-Higgins, 1978). Such a grammar describes a rhythm as accommodated by a hierarchical tree of duple or triple subdivisions. The resulting metrical tree specifies an important recurrent time interval, the bar, and the way it is subdivided recursively. These subdivisions define several levels of regular isochronous pulse trains, so-called beats. The notion of tactus refers to the most perceptually salient level of the meter, the level at which musicians keep timing by counting.

However, we have to note that meter is just one structuring factor in rhythm (cf. London, 2001), and unfortunately a complete grammar (or formal theory) of rhythm does not exist. For our purposes, however, it is enough to consider the space of all possible music notations.

As the note durations in Common Music Notation (CMN) can become very fine (1/32nd notes are in common use), and the ways to combine them are manifold (as was

briefly discussed above), the space of all possible rhythmic sequences that can be notated in CMN is enormous, even when considering a brief rhythmic pattern. We will refer to this set of all possible discrete rhythmic notations as score space (see footnote 4).

2.5. Categorization as Mapping from Performance Space to Score Space

Categorization can now be described as a mapping from a performance space into a score space. Such a transformation implies the partitioning performance space into a set of equivalence classes, all points in performance space mapping to the same score belong to the same class. Thus, although rhythmic categories are named or labeled by a sequence of integers, they are characterized by a region, an area in performance space, as is depicted in Figure 5.

<<Insert Figure 5 around here>>

This brings us to the question how these regions or areas in performance space can be characterized. We will do this using concepts from abstract algebra and topology (see, e.g., Fraleigh, 1976 for an introduction). A first characterization could be whether these areas are connected: no two separate regions in performance space form part of the same category (in score space). In that case all renditions of a certain rhythmic category will be enclosed by the same area, supporting the notion that a single performance region represents a certain rhythmic category. More specifically, they may be convex: no interpolation between two performances of the same score will be perceived as a different rhythm. In that case the shape of the areas becomes necessarily bounded by straight lines, they are polygons that partition the performance space (cf. Cemgil, Desain & Kappen, 2000).

Further issues that can be addressed in this way are the size of the areas and their shape. Are they larger for certain, possibly simpler rhythms? Are they symmetrical, or are the boundaries of expressive timing different for some intervals.

Finally, as a last example of the interesting issues that can be resolved when approaching categorization as a mapping between spaces, we focus on quantization. We use the term to refer to categorization, as well as the transformation of the category back into a mechanical performance (i.e., both arrows in Figure 5). The location of the mechanical performance need not necessarily lie inside the boundaries of its own region.

In that case it is no fixpoint: quantizing an already quantized performance yields another category (in the examples given in Figure 5, this holds for 2-1-1 but not for 3-1-4).

Before applying these concepts to our empirical data we will discuss our hypotheses and experimental setup.

2.6. Hypotheses

In this study we, first, expect to find evidence for the formation of rhythmic categories, confirming in a systematic way the research described above. We hope to be able not only to observe the category boundaries, but also to investigate the size of the category – its size and shape representing the amount of expressive variation that is allowed while still being identified as belonging to that rhythmic category. Second, we expect to find an effect of metrical context (as was shown by Clarke, 1987 for two rhythmic categories). If categorical rhythm perception is indeed open to top-down cognitive influence the categories might change in size and shape depending on the metrical context the rhythm is presented in.

Theories of meter perception predict an easier identification, coding and recall for rhythmic patterns that are metrical, a notion that is, however, defined differently in the various models. Using the definition that metrical rhythms are those patterns that have onsets in important metrical positions (Palmer & Krumhansl, 1990), contain subjective accents that align with a beat or clock (Povel & Essens, 1985), minimize syncopation (Longuet-Higgins & Lee, 1984), or couple well with hierarchically arranged oscillators (Large & Palmer, in press), it can be hypothesized that temporal patterns that are in accordance with the metrical context in which they are presented (i.e. they induce the same meter) are more easily identified, priming the responses. We therefore expect increased consistency of responses when a metrical context is presented. Furthermore, a preference for a rhythmic interpretation of patterns in a duple meter (Drake, 1993) would show up as a similarity between the duple and no meter conditions.

Next to the available empirical data, well-performing models of quantization (e.g., Longuet-Higgins, 1976; Desain & Honing, 1989) can be taken as theoretical, indirect indications for the shape of rhythmic categories (see Footnote 5). These models were shown to be quite accurate in describing the process of obtaining discrete rhythmic values from continuously varying note durations in a performed rhythm (Desain & Honing, 1992). A visualization of the behavior of, for example, a connectionist quantizer that is

based on a relaxation method (see Desain & Honing, 1991, p. 161) reveals that for a complete set of rhythms (a rhythm space as discussed above) there are larger areas around small integer ratios and smaller ones for more complex patterns, indicating more ‘tolerance’ or freedom in expressive interpretation for rhythms made up of small integers. Interpreting the watersheds between these so-called basins of attraction as the boundaries between categories, it predicts different shapes and sizes for the different rhythmic categories (see Footnote 6). Furthermore, this research suggests that categorization is actually facilitated by the temporal context: a different metrical context can result in a different quantization for the same performed temporal pattern (as the example in Figure 1 shows on a local scale with duple and triple divisions of the quarter note: two almost identical time intervals can give rise to two very different subjective durations). The Longuet-Higgins (1976) model arrives at a rhythmic interpretation by recursively subdividing a time interval in a duple or triple way until every onset in the pattern is close to the start or end of such a subdivision. This computational process can also be characterized by interpreting decision boundaries as the borders of the rhythmic categories. As such, both computational models address the same domain of cognitive functioning as we are trying to probe in our participants.

We will now, in more detail, describe the first identification experiment, designed to investigate this process of categorization, and the second experiment to pinpoint the influence of metrical context.

3. Experiment 1

3.1 Methods

3.1.1. Participants. The twenty-nine participants of Experiment 1 were highly trained professional musicians and advanced conservatory students from Dutch conservatories and from the Kyoto City University of the Arts in Japan. They had received between 7 and 17 years of musical training and were paid for their participation.

3.1.2. Apparatus. The sounds were presented through headphones (Sennheiser HD 445) on a Yamaha MU-90R synthesizer using General MIDI percussion sounds. The participants could adjust the loudness of the stimuli to a comfortable listening level. The synthesizer was driven by the POCO system (Honing, 1990; Desain & Honing, 1992) via

the OMS MIDI driver, running on an Apple Macintosh G3. The same computer collected the responses via a CMN interface (see Section 3.1.4).

3.1.3. Stimulus construction. The stimuli used in the experiment are a subset of all temporal patterns made up of four onsets adding up to a total duration of one second. Since this performance space is still infinite, a subspace of patterns needs to be sampled. In this study we use a temporal grid with a unit of $1/19^{\text{th}}$ of a second (about 0.053 s). A sampling unit of $1/19^{\text{th}}$ of a second was used (0.053 s) (see Footnote 7), with a minimum inter-onset interval of three units ($3/19^{\text{th}}$, i.e. 0.158 s). Thus IOIs vary between 0.158 s and 0.684 s in steps of 0.053 s. The choice of a prime unit prevents the induction of a fine metrical subdivision by the sampling itself. Furthermore, a minimum IOI was used to remain in the domain of musical note durations. In our experiments we choose a minimum duration of three time units, i.e. $3/19^{\text{th}}$ of a second (i.e. 0.158 s). The sampling needs to balance the size of the set with a fine enough resolution to still have a good representation of the continuous space of all possible performed rhythms. The 66 stimulus patterns that are used in the experiments are shown in Figure 6.

<<Insert Figure 6 around here>>

Each pattern was embedded in a context consisting of a fixed sequence of eight one-second time intervals used to induce the bar-level of a meter. We explicitly controlled for this since rhythmic patterns are known to induce a beat or pulse (Povel & Essens, 1985). The bar was marked by a ‘low bongo’ percussion sound (10 ms attack, 15 ms decay to 6 dB below peak-level) on a Yamaha MU-90R synthesizer at MIDI velocity 64 (see Footnote 8). Within this metrical framework the stimulus patterns were embedded being repeated three times (see Figure 7). The onsets were marked by a ‘high woodblock’ percussion sound (1 ms attack to 5 dB above bar markers, 10 ms decay) at MIDI velocity 76.

<<Insert Figure 7 around here>>

3.1.4. Response. For the identification task a specially designed computer interface (see Figure 8) was used, on which rhythms in CMN could be entered. Its design was guided by the responses obtained in a pilot study with an open response format on

music paper. The computer interface allowed for a large set of time signatures, notes and rests (ranging from a whole to a 32th duration), dots, ties, and grouping in duplets, triplets and quintuplets. Clicking buttons with the mouse allowed CMN to be formed. A correction button was provided to undo actions of button presses. An OK-button signaled the participants confirmed response. An underlying grammar checked for appropriate music notations (i.e. rhythms of three intervals that add up to one bar, starting with a time signature), disabling buttons that would lead to a wrong notation. For example, after selecting a time signature at the start of the bar all time signature buttons were disabled, and after selecting a second dot after a note the dot button was disabled until the next note was entered (see footnote 9). The grammar disallowed only ill-formed responses, supporting millions of different rhythmic patterns, which were subsequently reduced to approximately ten thousand classes each represented by a unique integer sequence. The rhythmic responses were stored along with other data, like the number of times the correction button was used.

<<Insert Figure 8 around here>>

3.1.5. Procedure. The participants performed an identification task in which they were asked to notate the presented stimulus using the computer interface described above. The presentation of the stimuli and entering of the responses was self-paced, but each complete stimulus (i.e. three repetitions of the rhythm embedded in a context of eight bars, see Figure 7) could only be listened to once. Each stimulus was presented once, in random order. Participants were instructed to think of the stimulus as if played by a percussionist, and to notate the score they thought was most likely used by the percussionist, this to prevent them from using extremely complex notation to write out the perceived expressive timing. This task was familiar to the participants as it is part of the standard sofège training at the music conservatory. The experiment took 45 minutes to complete, short breaks were allowed. One single participant was presented the stimuli multiple times, re-randomized in each of the six sessions, which took place with several days in between. Her responses were analyzed separately.

3.1.6. Data preparation. All responses (i.e. music notations) were converted to an irreducible integer representation, with several music notation variants leading to the same integer ratio representation. Response proportions (distribution of responses) were

calculated for each stimulus. This constitutes the measurement data from which the various maps used in this paper are constructed, as is illustrated in Figure 9. The bottom of this figure shows the performance space with the 66 stimuli used (cf. Figure 6). For three arbitrarily chosen stimuli the response proportions for the rhythms are shown (i.e. the three bars). The leftmost bar represents a stimulus that elicited only one response, the right most bar is an example where there is a clear maximal response but alternatives were chosen as well by some participants. The middle bar is an example of the absence of a winning maximal response. The first map that is constructed from this data (see Figure 9a) depicts maximal response proportions. This categorization map, or time clumping map, shows 1-1-1 (red) for the left-most bar, 1-1-2 (green) for the right-most bar, and none (white) for the middle stimulus since there is no response given significantly more often than any other for this performance. The second map (see Figure 9b) shows the agreement in responses. This entropy map shows a relatively high value for the middle bar (red), as many different responses were given for that stimulus. For the other two stimuli there is much more agreement and the entropy is low (blue) (see Footnote 10). The construction of the maps as continuous ternary contour plots was based on interpolated values in the stimulus space and visualized using JMP (version 3.2.2 by SAS). The details of the calculations used for the two maps are provided below.

3.2. Analytical Methods

3.2.1. Measure of consistency. To determine the consistency within and between participants in the identification task we use a measure of uncertainty. Stemming from information theory, and used in physics as well as psychology (Garner, 1975), the amount of noise in a signal can be expressed as the number of bits needed to encode the information. We assume a multinomial distribution of responses. For a set of responses to the same stimulus, with P_i denoting the response proportion for response i , and n being the number of measurements (i.e., the number of participants times the number of repeated presentations), the Shannon entropy E is defined as:

$$E = -\sum_{i=1}^n P_i \log_2 P_i$$

For example, if each of eight participants responds with a different notation to a single presentation of the same stimulus (or one participant responds differently on each of eight repeated trials) the entropy is three bits (much noise, low concordance). If the same response is given by all participants for that specific stimulus (or by one participant on all repeated trials), the entropy is 0 (high consistency).

We normalize this entropy measure by the maximally possible entropy, given n , the number of measurements, to obtain a measure which can be compared across conditions with a different number of measurements:

$$E_r = \frac{E}{\log_2 n}$$

Since the relative entropy E_r is defined for each stimulus, the average relative entropy of a complete stimulus set is used when comparing the overall consistency of different groups of participants or of different conditions. Correlations between relative entropy distributions over the stimuli are used when comparing patterns of consistency of different groups of participants or of different conditions.

3.2.2. Confidence areas. For one stimulus the response proportions for the various rhythmic patterns stem from a multinomial distribution. The rhythm that exhibits the largest response proportion was considered the winning category. However, one has to consider the statistical reliability of the statement that the probability of this category is indeed larger than any other, given the number of trials and the distribution, which is in general hard to do. However, realizing that usually most proportions are close to zero and only two responses compete, the significance of their difference was tested as if they were independent. This approximation made it possible to indicate where in the stimulus space there was more than 90% confidence that one proportion was indeed the largest.

3.3. Results

The 29 participants responded with 123 different categories. Of those responses, 21 rhythms were used only once by one participant. The pattern 1-1-1 was used most often as a response, in about 10% of all trials (in Section 3.3.3, Table 1 a complete

overview will be given). In presenting the instruction, the nature of the identification task was difficult to make clear to some participants. Few professional musicians were lured into using extremely complex notation, not uncommon to the modern music repertoire. The time to get acquainted with the instruction and interface was minimal, and making use of the response interface proved easy. In the post-experiment interview most participants stated that the task was tiring but not too hard to complete. None of the participants complained about the range of responses allowed by the user interface, not even when explicitly asked.

3.3.1. Is the task difficult? We first considered the difficulty of the task and the consistency of responses. To study this, one participant participated in this experiment six times. The relative entropy of the response distribution per stimulus for this participant is shown in Figure 10a. Here the blue areas indicate rhythms for which the same response was always given for the same stimulus (low entropy/high consistency), the red areas indicate patterns in which almost every presentation of the same stimulus elicited a different response (high entropy/low consistency). Whether the areas of high entropy represent unmusical performed rhythms or just highly ambiguous ones cannot be decided here, but both these high entropy (high uncertainty) and low entropy (high consistency) areas appear roughly similar when visualizing the relative entropy of a large group of participants ($N=29$) that received each stimulus once (see Figure 10b). The correlation between the two entropy distributions is .47 ($p<.001$), indicating a response consistency within one participant which is analogous to the concordance between the participants.

A further validation of the interpretation of entropy as a measure of the difficulty of the identification task, as opposed to an interpretation as the size of inter-subject differences in the perception of a performance, can be obtained by considering the number of corrections the participants used before arriving at the chosen response. The correction button was used on average 1.8 times per response. The correlation of the number of corrections over the stimuli with the entropy was .55 ($p<.001$), suggesting that difficulty of finding a proper response accounts indeed for a large part of the amount of entropy.

The areas of low entropy seem to focus around relatively simple rhythms like 1-1-1, 2-1-1, 1-2-1, and 1-1-2. For example, one of the most consistently identified stimuli is [.474, .263, .263] (in seconds) which was identified by 28 of the 29 participants as 2-1-1. The single participant also chose the same response six out of six times. The

stimulus that prompted the widest range of responses was [.211, .368, .421]. This pattern gave rise to 17 different rhythms of which both 1-2-3 and 1-3-2 were most common (14%) and many complex rhythms were chosen once. The single participant chose three different patterns in the six repeated trials of that same stimulus.

In conclusion we may say that the entropy patterns suggest that the task can be considered as not too difficult, especially in specific areas around relatively simple rhythms (i.e. rhythms with low integer ratios). Thus meaningful analyses of the rhythmic categories identified come within reach.

3.3.2. Which rhythmic categories are identified where? To study this question we examine the partitioning of the performance space according to the rhythms that are most frequently identified, considering all stimuli that attract the same response rhythm most often as belonging to the same category. Since we did not present participants with two pre-selected response categories as used in standard categorical perception investigations (Harnad, 1987) but used semi-open response categories (cf. Schulze, 1989), we cannot use the conventional operational definition of the boundary of a category (i.e. examining the shapes of the identification and discrimination functions) (see Footnote 11). Also the commonly used confusion matrix approach to the derivation of a discrimination measure is not applicable here since correct response categories are not known. Figure 11 shows the maximally identified rhythms. This time clumping map reveals the apparent coagulation in rhythm perception. Colors represent rhythmic categories. Their music notation and integer representation is shown in the legend, which lists them in order of response proportion. Gray lines represent the boundaries between categories. Darker shades of color indicate a larger proportion of participants who choose this identification (darkest shade marks 100% agreement among participants). The white areas are an indication that there is less than 90% statistical confidence that one rhythmic category is identified most often.

As the reverse of perceptual categorization, which assigns areas in performance space to the same score, these categories also form constraints on expressive musical performance. Each area signifies the freedom for the musician in choosing expressive timing, for were the performer to push the amount of expressive timing too much from the center of the category, passing across a boundary, the audience would perceive another rhythm.

3.3.3. Analyses. This representation (see Figure 11) allows us to investigate the topological aspects (like partitioning, shape and position) of the responses and characteristics like symmetry and permutation.

We first consider the range of the categorization transformation: the set of scores that are identified as the categorization of a performance. Out of 123 different responses, there are twelve rhythms that are identified in the time clumping as achieving a maximal response. They cover 60% of the responses (see Table 1). There are large differences in the sizes of these categories: simple patterns seem to account for a larger area of performance space. This is also reflected in the underlying response proportions. In Table 1 the response proportions (over all 66 stimuli and all 29 participants of Experiment 1) of these patterns are listed, ordered by their response proportion. Only those patterns are listed that become the maximal response somewhere in the performance space.

<<Insert Table 1 around here>>

Informally one can state that the response proportions reflect the simplicity of the pattern: patterns made up of low integer ratios (e.g., 1-1-1 and 1-2-1) take a large proportion of responses, and patterns like 5-4-7 or 1-5-3 are almost never given as response (in this case less than .001). However, since it still all but clear which structural and perceptual factors contribute to perceived complexity of a rhythm (cf. Tanguiane, 1993; Shmulevich & Povel, 2000), we will leave a discussion on the relation between our data and these theories to a separate occasion.

Second, note that the set of rhythmic sequences that appear in the time clumping map is not closed under permutation. Closed under permutation would mean that all orderings of 1-2-3 occur (i.e. 1-2-3, 2-3-1, 1-3-2, 3-1-2, 3-2-1, and 2-1-3). However, it turns out that certain patterns (e.g., 2-3-1 and 3-1-2) are present, but re-orderings of the intervals in the sequence (e.g., 1-2-3) are not. In that sense rhythmic sequences in score space are not represented symmetrically in the responses. Only simple patterns, such as 1-1-2 appear in all possible orders. As an overall measure of this asymmetry of rhythmic sequences we consider the correlations between the 62 response proportions of patterns consisting of three different note durations (represented by A, B and C) and their permutations. They are listed in Table 2. Note that a reflection in time, i.e. C-B-A, though perceptually very different, is quite similar to the original rhythm A-B-C in the amount it

is being identified from performances. A reason for this could be that the tree structure needed to metrically encode A-B-C is the same for its reversal C-B-A, attracting a similar response proportion, while other permutations of A-B-C might have a different metrical encoding, accounting for more differences between the amount of responses.

<<Insert Table 2 around here>>

Third, we consider the partitioning of the performance space into areas. As it turns out there is only one area for each rhythmic category. This means that in the representation adopted in this paper the rhythmic categories are connected. Thus, for any two performances perceived as instances of the same rhythm, there exists a path of arbitrarily fine interpolations connecting them such that all performances on this possibly curved path will be perceived as the same rhythm.

Fourth, the area for each category appears to be almost convex, which means that there is even a straight-line segment connecting two points within the same category that will lie completely inside that category. The map itself is therefore close to an idealized one with convex tiles in which the areas are replaced by convex polygons. We will refer to this situation as quasi-convex. To measure the closeness to convexity we compare each area to its convex hull (Cormen, Leiserson & Rivest, 1990), the smallest convex polygon circumscribing it. The boundaries of a convex area must necessarily be straight, as any bend would create a non-convex area on one side of the line.

Convexity of the categories is measured by the proportion of the surface area of the convex hull that is taken up by the category itself. It turns out that the categories cover on average 91% of the surface area of their convex hull, with a minimum of 71% and a maximum of 97%. Thus, though one has to keep in mind that we average over participants (see footnote 12), one could say that rhythmic categories are indeed convex. This has an important theoretical implication: with the categories being convex, a simple attractor model could, in principle, suffice to describe the behavior observed; non-convexity would have ruled out these models. Furthermore, it implies that the boundaries between neighboring categories, which are all convex, can only be formed by straight line segments (see footnote 13), important characteristic for modeling these data.

Fifth, we investigated whether mechanical performances of a rhythm are fixpoints of the quantization transformation. If this is the case, quantization can be applied twice

without changing the outcome. In other words, any performance region will contain its rhythmic category mapped back to the performance space (see example 2-1-1 in Figure 5). This turns out to be the case for all categories that attract a maximal response (i.e., all rhythms mentioned in Table 1).

Sixth, we consider the position of the rhythmic category with regard to a prototype of the category: the mechanical performance. Judging this by the boundary, in general the amount of allowed expressive timing deviation is distributed unevenly around the mechanical performance (marked with a black cross in Figure 10) of the rhythmic category, usually with a slightly longer third interval. The same effect shows in the position of the areas marked with the darkest shade (i.e. the region where all participants identified a certain stimulus as the same rhythm). These regions also are not centered around the mechanical performance. This is most easily seen in the four patterns marked with a black cross in Figure 10. For these rhythmic categories a non-mechanical stimulus attracted more responses than the mechanical one. This suggests that rhythmic categories may be labeled, but are not necessarily best prototyped, as a pattern of integers (see Footnote 14). They are actually shaped such as to reflect common expressive timing patterns in music performance. This also means that simple rounding methods can never model the process of rhythmic categorization accurately.

Yet another way to estimate this effect is to judge the position of the centroids of the response distributions for the main rhythmic categories (see Footnote 15). Table 3 lists the performance centroids (in seconds) of the sensitive regions for the main rhythms and the differences (in seconds) with a mechanical rendition. The distance from the mechanical performance reflects expected patterns of expressive timing for the various rhythms. Some regularities can be seen easily, for example, a tendency to slow down at the end of a rhythm. Most of the IOIs are significantly different ($p < .001$) from their mechanical version when considered as marginal normal distributions, except the first interval of 2-1-3 and the last of 3-1-4, though they are all quite small relative to the sampling interval of 0.053 s. The just noticeable difference (JND) of timing in an isochronous sequence of intervals of 0.333 s is about 0.006 s (Michon, 1964; Friberg & Sundberg, 1995) and can be compared with the 0.030 s timing deviation in the last note of the 1-1-1 pattern. Since a generative theory on the relation of rhythmic structure and expressive timing does not exist, we have no model to test against this data. However, the data support our argument that rhythm in score space is not well represented by its

mechanical rendition in performance space, the performance centroid (interpreted as the most communicative rendition of a category) being a more likely candidate.

<<Insert Table 3 around here>>

It is important to note that because responses arise in competition, the centroids cannot be directly interpreted as the means of prior distributions of the timing of the rhythmic prototypes. Nevertheless, their position, in conformity to a well-known rule of music performance (i.e. slowing down at the end, cf. Palmer, 1997), is as expected.

The results presented above concern the categorization of rhythmical patterns independent of context. As previous research (Clarke, 1987) has shown that there might be an effect of meter on rhythmic identification, we conducted a second experiment in which we presented the same stimuli in different metrical contexts. This allows us to address the question whether the presence of a meter influences rhythmic identification.

<<Insert Color Figures 9, 10, 11 and 13 around here>>

4. Experiment 2

4.1. Method

4.1.1. Participants. Eleven participants took part in the experiment. All participants of Experiment 2 had taken part in Experiment 1 as well.

4.1.2. Apparatus. Equipment was the same as in Experiment 1.

4.1.3. Stimulus construction. The stimuli used in the experiment were similar to the stimuli in Experiment 1. The empty bars were filled with a subdivision depending on the context condition. The two metrical context conditions (see Figure 12b and 12c) were created by filling the empty bars with a triple or duple subdivision (see Footnote 16) The beats were marked by a ‘low bongo’ percussion sound (10 ms attack to 8 dB below bar markers, 15 ms decay) at MIDI velocity 38. The control condition (see Figure 12a) contained no subdivisions, exactly as in Experiment 1.

<<Insert Figure 12 around here>>

4.1.4. Response. The same interface was used for collecting the responses as in Experiment 1.

4.1.5. Procedure. The procedure and instruction was identical to those in Experiment 1. The stimuli were presented in random order, blocked by context condition. The control condition was presented first, and the duple and triple condition were presented in a random order.

4.2. Analytical Methods

Data processing and analysis were conducted similar to Experiment 1. For comparing significance of differences between response proportions in multinomial distributions we had to use an approximation: differences between probabilities of the same response in different conditions were tested individually, as if stemming from two binomial distributions.

4.3. Results

In total there were 158 different responses over all three conditions, but only 112 of those were used more than once. The duple meter condition yielded 78 different responses, 90 different rhythmic patterns without metrical context, and 113 in triple context.

4.3.1. Is there an effect of meter? A meter can be defined as consisting of at least two levels of temporal structuring, for instance, a bar subdivided in two or three equal beats (Martin, 1972). We focus here only on the topmost two levels of metrical structure. This was studied by presenting the same rhythmic patterns preceded by a bar subdivided in two or three equal beats (duple and triple meter, respectively). As the participants were forced to choose a meter in which to notate their response, it is worthwhile inspecting these choices. In the condition without a metrical context, 84% of the responses were notated in a duple meter (like 4/4 or 6/8). There is no significant difference with the results of Experiment 1 in which 81% of the responses were notated in a duple meter. The amount of duple meter responses increased to 99% when duple meter was indeed presented, a highly significant increase ($p < .001$). In the case of a triple meter context, only 5% of the responses were notated in duple meter, a highly significant decrease ($p < .001$), the rest used a triple meter (like 3/4 or 9/8). These results suggest a successful manipulation of metrical context in the three conditions.

Let us first focus on the agreement between participants. The average relative entropy over all stimuli is .62 for the no meter condition, .60 for duple meter and .68 for triple meter. So, while slightly less consistent in the triple meter condition, the overall consistency seems not to be affected by the absence or presence of a metrical context. This indicates that the difficulty of the task is, in general, not influenced by the availability of a metrical context. However, correlating the entropy distributions over the stimuli leads us to a different conclusion. Whilst the correlations between the entropy distributions are all significant ($p < .05$ for the lowest correlation) they are still quite low (.64 between no and duple meter, .28 between no and triple meter, and .23 between triple and duple meter). Given that the mean entropies are about the same for the three conditions, we conclude that the pattern of entropy distribution across the stimuli must be quite different in the three conditions. In other words, even though the conditions are of the same overall difficulty, the process of rhythmic identification is helped by presenting an appropriate meter and hindered by another one.

We may even assume that the appropriate meter would be the meter that participants construct mentally while listening to the stimulus in the no meter condition. To test this hypothesis we defined for each stimulus the notion of appropriate meter as the meter that yields the highest agreement among subjects, or, in other words, the metrical context that made the transcription task less equivocal. If a stimulus is embedded in its most appropriate meter the average entropy is .54, as compared to the .62 when no meter is provided. Similar figures (.60 and .68) arise when duple or triple meter is provided. So we can conclude that presenting the stimuli in a metrical context does not, by itself, ease the identification task. But if the context is appropriate, it seems to help identification, as judged by the agreement among subjects.

Moving from the amount of agreement to chosen responses themselves, we can now present the effect of metric priming on the empirical response proportions.

The time clumping maps in Figure 13 show the effect of meter on rhythmic identification in the three metrical contexts. Response proportions in the no meter condition (Figure 13a) correlate very highly (.94, $p < .001$) with the responses in the duple meter condition (Figure 13b). This is as expected, since listeners tend to prefer a duple interpretation of rhythms when no metrical context is given (Longuet-Higgins, 1976; Drake, 1993). Notably, for the twelve rhythms that are shown in the time clumping map, only for three rhythms there is a significant difference in response proportions between

the duple and the no meter condition. Table 4 lists the response proportions over all (66) stimuli and all (11) participants in the no meter condition, plus the difference between the conditions, only listing those patterns that become the maximal response somewhere in the performance space. The table is ordered according to the size of the differential effect of triple vs. duple meter (right most column), the topmost response gaining most when presented in a triple context (approximate significance of the shifts in response proportions is indicated by asterisks). This suggests a similarity of the duple meter, when primed, to the duple meter as preferentially induced in the mind of the listener when absent in the stimulus.

<<Insert Table 4 around here>>

However, behavior in the triple meter condition (Figure 13c) differs significantly more ($p < .01$) than duple meter from the no meter condition as the correlation is .70 (as compared to .94 between duple and no meter). And, as expected, it is even more different from the duple meter condition (correlation .53). This is reflected in the significance of the differences between response proportions (as were shown in Table 4): only for two rhythms these shifts are not significant. Pattern 1-1-1, we can assume, is such a simple and well known subdivision that even adding a duple meter, in which case it appears syncopated, does not make it less readily identifiable. The other pattern that is indifferent to metrical context, 2-1-3, is an interesting one. It is ambiguous in the sense that it fits both meters well. Thus, adding either duple or triple meter makes it not harder to identify, it just achieves another metrical interpretation. This is reflected in the relative position of the 2-1-3 area in performance space, in duple meter it is centered around [.341, .179, .480], exhibiting a long first note, while a triple context yields a longer second note. In this case [.329, .196, .476] is the performance centroid, apparently the most communicative rendition of the rhythm 2-1-3 in a triple meter context.

Thus, while an ambiguous rhythm may receive the same response proportions irrespective of meter, the expected expressive timing may differ. This is an important point for any theory of rhythm perception, as the interaction between the processes of rhythmic categorization and beat induction need to be made clear. Apart from the two patterns 1-1-1 and 2-1-3, most rhythms are very sensitive to metric priming as exhibited by their response proportions. The largest change occurs for 1-2-1, which, albeit

syncopated in duple meter, is readily identifiable in that meter. In the triple meter case this pattern becomes almost impossible to recognize. The largest change in the opposite direction is exhibited by rhythm 1-3-2. This pattern becomes easier to recognize when primed by triple meter, primed by duple meter it almost disappears. The results presented show the large overall effect of metric priming considering the response proportions over all of the stimulus space.

Furthermore, the effect of meter also completely changes the tiling of the performance space (see Figure 13). Some simple rhythmic categories that were identified quite often for a specific meter-less stimulus, and that were even more salient in the context of a duple meter, disappear completely in triple meter. To give an example: the stimulus that is perceived most differently in the two meters is [.210, .474, .316]. In duple meter it is heard mostly as 1-2-1 (64% of the responses; with another 10% identifying it as 1-3-2), in triple meter it is interpreted mostly as 1-3-2 (36% of the responses, not one participant identifying it as 1-2-1).

So, in summary we conclude that there is a strong effect of meter on rhythmic identification.

5. Discussion

Psychological research focusing on cognitive processes often restricts itself to the study of the effects of a few conditions, when it is faced with the sheer enormity of the set of possible stimuli that may be used. In this paper we have tried to directly address this difficulty in more encompassing way. By carefully formalizing restricted spaces of stimuli and responses and systematically probing them, it turned out to be possible to contain the combinatorial explosion of possibilities and yet arrive at general statements on the cognitive process itself, instead of being constrained to ad-hoc statements about the differences between a few stimuli. Concerning the categorization of temporal patterns we have concluded from our study that the space of performances is partitioned into a small set of connected rhythmic regions. As a consequence a model that chooses the smallest distance between performance and a set of rhythmic categories (nearest neighbor) might be appropriate. This set is not closed under sequential permutation. The categories differ in size, according to the complexity of the rhythm, and their location is not centered around the position of a mechanical rendition of the category. The shape of a rhythmic category in performance space is quasi-convex. This tells us something about

the fabric of the underlying rhythm space and the processes that access it, as convex compartments occur quite often in nature, like the surfaces between soap bubbles. These mechanisms are commonly formalized by non-symbolic methods like differential equations, complex dynamics, relaxation networks and attractors. Thus such formalisms may be the best candidates for a successful modeling of rhythmic categorization.

The fact that one is now able to formalize the rhythmic categories facilitates research on this abstract level towards a computational model that captures behavior so fundamental to rhythm perception. Both its importance, and the complexity of the modeling task are demonstrated by two facts. First, without rhythmic categorization there would be no reference against which to judge the expressive duration of a note: one would not be able to appreciate the difference between a deadpan and an expressive performance. And secondly, too much and too strict categorization would cause a loss of timing information and the difference between a deadpan and an expressive performance would not even be noticeable.

Turning to the effect of metric priming, Clarke (1987) had already demonstrated an effect of meter on the position of the category boundary for a few rhythmic patterns. In this investigation, we have shown the effect for a systematic set of temporal patterns and their clumping into rhythmic categories. The results for the different contexts had to be subsequently interpreted as a large and robust effect of metric priming. Thus rhythmic categorization depends on a pre-established cognitive framework of time structuring, a main finding of this investigation. In the absence of a metrical context a duple meter was clearly preferred, as the responses were similar to those with a duple meter prime. The effect of priming can be partly explained on the basis of the temporal structure of the responses: the location of note onsets on a temporal grid. This is in accordance with theories that stress the importance of the role of meter in the mental representation of rhythm (Longuet-Higgins, 1976; Povel & Essens, 1985), sometimes even so far as to state that rhythm only exists under a metrical interpretation.

The question of the relation between the rhythmic categorization process and the meter induction processes is still open. Are these processes modular and conducted in sequential order, or are they better understood as one integrated process? Most meter-induction studies (Longuet-Higgins & Lee, 1984; Povel & Essens, 1985; Parncutt, 1994) (see footnote 17) have used readily categorized rhythmic time intervals as input, with the exception of Longuet-Higgins (1976) who presents an algorithm that induces meter while

categorizing the rhythm. Our finding that the rhythmic categories depend on meter can only demonstrate that the categorization process is open to induced meter. However, since the areas of timing that define a category for the ambiguous rhythms, which fit both meters equally well, were shown to be systematically different, the timing information itself might provide a cue for meter induction. This communication of meter by expressive timing (cf. Sloboda, 1983) suggests that meter induction and rhythmic categorization are closely interrelated processes. Furthermore, we found that, while presenting one or the other meter did not increase participant agreement as compared to presenting no meter, presenting an appropriate meter for each response did increase the conformance of the responses. Thus presence of a metrical context eases the formation process of rhythmic categories, which again is in agreement with a model such as the one formulated by Longuet-Higgins (1976; 1979) that, while categorizing, grafts a representation of musical rhythm onto the framework of a pre-established meter.

As meter is not the only contextual aspect that may influence rhythm identification, we suggest that other aspects of rhythm, such as global tempo (Handel, 1993), loudness accents, articulation and melodic structure (Tekman, 1997), remain worthy of investigation (Honing, 2002). The methodology that was developed and the results that were presented can contribute to a fuller understanding and validation of the computational models of quantization and the theoretical accounts of rhythm perception (Longuet-Higgins, 1976; Desain & Honing, 1991; Large & Kolen, 1995) which are still in urgent need of empirical support (Desain, Honing, Van Thienen & Windsor, 1998). And, in the larger scheme, it appears that the topic of categorization has a much wider relevance, as it reflects the transition from sub-symbolic to symbolic mental representations and thus forms a bridge from perceptual processes to cognitive ones, with rhythm perception being an intriguing domain of where these levels of representation meet (see footnote 18).

References

Cemgil, T., Desain, P. & Kappen, B. (2000). Rhythm Quantization for Transcription. *Computer Music Journal* 24(2), 60-76.

- Clarke, E. F. (1987). Categorical rhythm perception: an ecological perspective. In A. Gabrielsson (Ed.): *Action and Perception in Rhythm and Music*, Royal Swedish Academy of Music, 55, 19-33.
- Clarke, E. F. (1995). Expression in performance: generativity, perception and semiosis. In J. Rink (ed.), *The Practice of Performance: Studies in Musical Interpretation*. Cambridge: CUP, pp. 21-54.
- Clarke, E. F. (2000). Categorical rhythm perception and event perception. In *Proceedings of the International Music Perception and Cognition Conference*. Keele University, Department of Psychology. (CD-Rom).
- Clarke, E.F. (1999). Rhythm and Timing in Music. In D. Deutsch (Ed.), *Psychology of Music, 2nd Edition* (pp. 473-500). New York: Academic Press.
- Cooper, G. & Meyer, L.B. (1960). *The rhythmic structure of music*. Chicago, IL: University of Chicago Press.
- Cormen, T. H., Leiserson, C. E., & Rivest, R. L. (1990) *Introduction to Algorithms*. Cambridge, MA: MIT Press.
- Desain, P. & Honing, H. (1989). Quantization of musical time: a connectionist approach. *Computer Music Journal* 13(3), 56-66.
- Desain, P. & Honing, H. (1991). Quantization of musical time: a connectionist approach. In P.M. Todd, & D.G. Loy, (Eds.) *Music and Connectionism* (pp. 150-167). Cambridge: MIT Press.
- Desain, P. & Honing, H. (1992) *Music, Mind and Machine: Studies in Computer Music, Music Cognition and Artificial Intelligence*. Amsterdam: Thesis Publishers.
- Desain, P. & Honing, H. (2001, In press). Modeling the Effect of Meter in Rhythmic Categorization: Preliminary Results. *Journal of Music Perception and Cognition*, 7.

Desain, P. & Windsor, L. (Eds.) (2000) In: *Rhythm Perception and Production*. Lisse: Swets and Zeitlinger.

Desain, P., Honing, H., Van Thienen, H. & Windsor, L.W. (1998). Computational Modeling of Music Cognition: Problem or Solution? *Music Perception* 16 (1), 151-166.

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians and adult non-musicians, *Perception & Psychophysics*, 53 (1), 25-33.

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

Fraleigh, J. B. (1976). *A first course in abstract algebra*. Reading, MA: Addison-Wesley.

Friberg, A. & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America*, 98(5), 2524-2531.

Gabrielsson, A. (1999). The Performance of Music. In D. Deutsch (Ed.), *Psychology of Music*, 2nd Edition (pp. 501-602) New York: Academic Press.

Garner, W.R. (1975) *Uncertainty and Structure as Psychological Concepts*. New York: Wiley.

Handel, S. (1993). The effect of tempo and tone duration on rhythm discrimination. *Perception & Psychophysics*, 54(3), 370-382.

Harnad, S. (1987). *Categorical Perception. The Groundwork of Cognition* Cambridge: Cambridge University Press.

Honing, H. (1990). POCO: an environment for analysing, modifying, and generating expression in music. In *Proceedings of the International Computer Music Conference*. 364-368. San Francisco: Computer Music Association.

Honing, H. (2002) Structure and interpretation of rhythm and timing. *Tijdschrift voor Muziektheorie* [Journal of Music Theory], 7(3), 227-232.

- Jones, M.R. (1990). Musical events and models of musical time. In R.A. Block (Ed.), *Cognitive models of psychological time* (pp. 207-240). Hillsdale, NJ: Erlbaum.
- Le Grand, Y. (1968) *Light, Color and Vision, 2nd Edition*. London: Chapman and Hall.
- Large, E. W. & Kolen, J.F. (1995). Resonance and the perception of musical meter. *Connection Science*, 6, 177-208.
- Large, E.W. & Jones, M.R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review* 10(1), 119-159.
- Lerdahl, F. & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- Livingstone, K.R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 732-753.
- London, J. (2001). Entry on 'Rhythm.' *The New Grove Dictionary of Music and Musicians, 2nd Edition*. London, Macmillan.
- Longuet-Higgins, H. C. (1976). The perception of melodies. *Nature*, 263, 646-653. (Reprinted in Longuet-Higgins, 1987)
- Longuet-Higgins, H.C. & Lee, C.S. (1984). The rhythmic interpretation of monophonic music. *Music Perception*, 1(4), 424—441. (Reprinted in Longuet-Higgins, 1987)
- Longuet-Higgins, H. C. (1978). The grammar of music. *Interdisciplinary Science Reviews* 3(2), 148-156. (Reprinted in Longuet-Higgins, 1987)
- Longuet-Higgins, H.C. (1979). The perception of music. *Proceedings of the Royal Society of London. B* 205, 307-322. (Reprinted in Longuet-Higgins, 1987)
- Longuet-Higgins, H.C. (1987). *Mental Processes*. Cambridge, Mass.:MIT Press.

- Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behaviour. *Psychological Review*, 79, 487-509.
- Michon, J. A. (1964). Studies on subjective duration 1. Differential sensitivity on the perception of repeated temporal intervals. *Acta Psychologica*, 22, 441-450.
- Michon, J.A. & Jackson, J.L. (1985). *Time, Mind, and Behavior*. Berlin: Springer.
- Nakajima, Y. (1987). A model of empty duration perception. *Perception*, 16, 485-520.
- Palmer, C. & Krumhansl, C.L. (1990). Mental representations of musical meter. *Journal of Experimental Psychology: Human Perception and performance*, 16(4), 728-741.
- Palmer, C. (1997) Music Performance. *Annual Review of Psychology*, 48, 115-138.
- Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, 11, 409-464.
- Povel, D.J. & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2(4):411-440.
- Repp, B.H. (1984). Categorical Perception: Issues, Methods, Findings. In N.J. Lass (Ed.), *Speech and Language, Advances in Basic Research and Practice 10*, (pp. 242-335). New York: Academic Press.
- Saunders, B.A.C. & van Brakel, J. (1997). Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences*, 20, 167-228.
- Schulze, H.H. (1989). Categorical perception of rhythmic patterns. *Psychological Research*, 51, 10-15.
- Shmulevich, I. and Povel, D. (2000). Complexity Measures of Musical Rhythms. In P. Desain, and W.L. Windsor (Eds.), *Rhythm Perception and Production* (pp. 239-244). Lisse: Swets & Zeitlinger.

- Sloboda, J.A.(1983). The communication of musical meter in piano performance. *Quarterly Journal of Experimental Psychology*, 35, 377-396.
- Sternberg, S., Knoll, R.L. & Zukofsky, P. (1982). Timing by Skilled Musicians. In D. Deutsch (Ed.), *Psychology of Music* (pp. 181-239). New York: Academic Press.
- Tanguine, A.S. (1993). *Artificial perception and music recognition*. Berlin: Springer-Verlag.
- Tekman, H.G. (1997). Interactions of perceived intensity, duration, and pitch in pure tone sequences. *Music Perception*, 14, 281-194.
- Timmers, R. & Honing, H. (2002). Issuing an empirical musicology of performance. In Timmers, R., *Freedom and constraints in timing and ornamentation. Investigations of music performance*. (pp. 19-49). Maastricht: Shaker Publishing.
- Todd, N.P.M. (1992). The dynamics of dynamics, a model of musical expression. *Journal of the Acoustical Society of America*, 91, 3540-3550.
- Vos, P. & Handel, S. (1987). Playing triplets: facts and preferences. In A. Gabrielsson (Ed.): *Action and Perception in Rhythm and Music*, Royal Swedish Academy of Music, 55, 35-47.

Acknowledgements

This research was funded by the Netherlands Organization for Scientific Research (NWO). Support was also provided by the Canon Foundation. Experiments were conducted by Rinus Aarts, Chris Jansen and Maki Sadakata, at the Nijmegen Institute for Cognition and information (NICI) and at Kyoto City University of the Arts, Japan. We are grateful to John A. Michon, Mari Riess Jones, and two anonymous reviewers for their thoughtful comments on an earlier draft, greatly improving the current presentation. Finally, we like to thank Huub van Thienen for implementing the CMN user-interface, Kathleen Jencks for correcting our English, and Torsten Anders, Maki Sadakata and Paul Trilsbeek for constructing the web demo available at www.nici.kun.nl/mmm/time.html.

Figure Captions

Figure 1. Example of the two representations of time present in music: a performed rhythm in continuous time (**a**) and the perceived rhythmic interpretation in discrete, symbolic time (**b**). (An audio example is available at www.nici.kun.nl/mmm/time.html).

Figure 2. Stimuli used by (**a**) Clarke (1987) and (**b**) Schulze (1989) (see text for details).

Figure 3. Performance space of all 3-interval temporal patterns adding up to one second duration (**a**), ternary plot (**b**), and two example patterns (**c**) (see text for details).

Figure 4. Stimuli and results of Schulze (1989) depicted as a chronotopic map (NB zooming-in on part of the performance space, indicated by the diagram at the right). The gray lines indicate the perceived category boundaries. The gray area is the hypothetical shape of the rhythmic category **A**. The dots identify the (interpolated) rhythms, crosses mark the mechanical ones (cf. Figure 2b).

Figure 5. Hypothetical performance regions and their mapping to score space (dark gray arrows). And the mapping from score space to performance space (light gray arrows) with crosses indicating the mechanical rendition of the rhythmic category.

Figure 6. The sampling of the performance space as used in the experiments.

Figure 7. The stimulus pattern as used in Experiment 1: Each line represents the onset of a percussive sound, the gray lines represent the metrical context against which the to be identified pattern (black lines) is repeated three times.

Figure 8. Music notation interface as used for the identification task.

Figure 9. Construction of a time clumping map (**a**) and an entropy map (**b**) from the responses (see text for details).

Figure 10. Entropy maps: Entropy as a measure for consistency for the space of all possible 3-interval temporal patterns (in seconds), **(a)** within participant ($N=1$, 6 repetitions) and **(b)** between participants ($N=29$). Blue indicates low entropy/high consistency areas, red represents high entropy/high uncertainty areas.

Figure 11. Time clumping map: transforming continuous time intervals (physical time) into rhythmic categories (perceived time). Colors represent the winning rhythmic categories (as identified in the Rhythms legend at the right) with darker shades indicating a higher proportion of participants identifying it. Gray lines are category boundaries. In the white areas there is less than 90% statistical confidence that there is one rhythmic category identified most often.

Figure 12. The stimulus pattern as used in the three conditions: control (no meter), duple meter, and triple meter. Each line represents the onset of a percussive sound, the gray lines represent the metrical context against which the to be identified pattern (black lines) is repeated three times.

Figure 13. Time clumping maps: the effect of duple **(b)** and triple **(c)** meter as compared to the no meter condition **(a)**. Note the slight growth of the areas around simple rhythms like 1-1-2 and 1-2-1 in the duple meter condition, and their complete disappearance in the triple meter condition. (See www.nici.kun.nl/mmm/time.html for an animation showing the changing categories.)

Tables

Table 1Response Proportions of Experiment 1

Response	Proportion	Cumulative
1-1-1	.094	.094
1-2-1	.080	.174
2-1-1	.064	.238
1-1-2	.060	.298
2-3-1	.046	.344
4-3-1	.043	.387
3-1-2	.041	.428
4-1-1	.039	.467
1-3-2	.038	.505
3-1-4	.031	.536
1-1-4	.031	.567
2-1-3	.028	.595

Table 2**Correlations Between Response Proportions of the Pattern A-B-C and its Permutations**

Permutation	Correlation with A-B-C
B-A-C	0.97
C-B-A	0.94
B-C-A	0.88
C-A-B	0.76
A-C-B	0.75

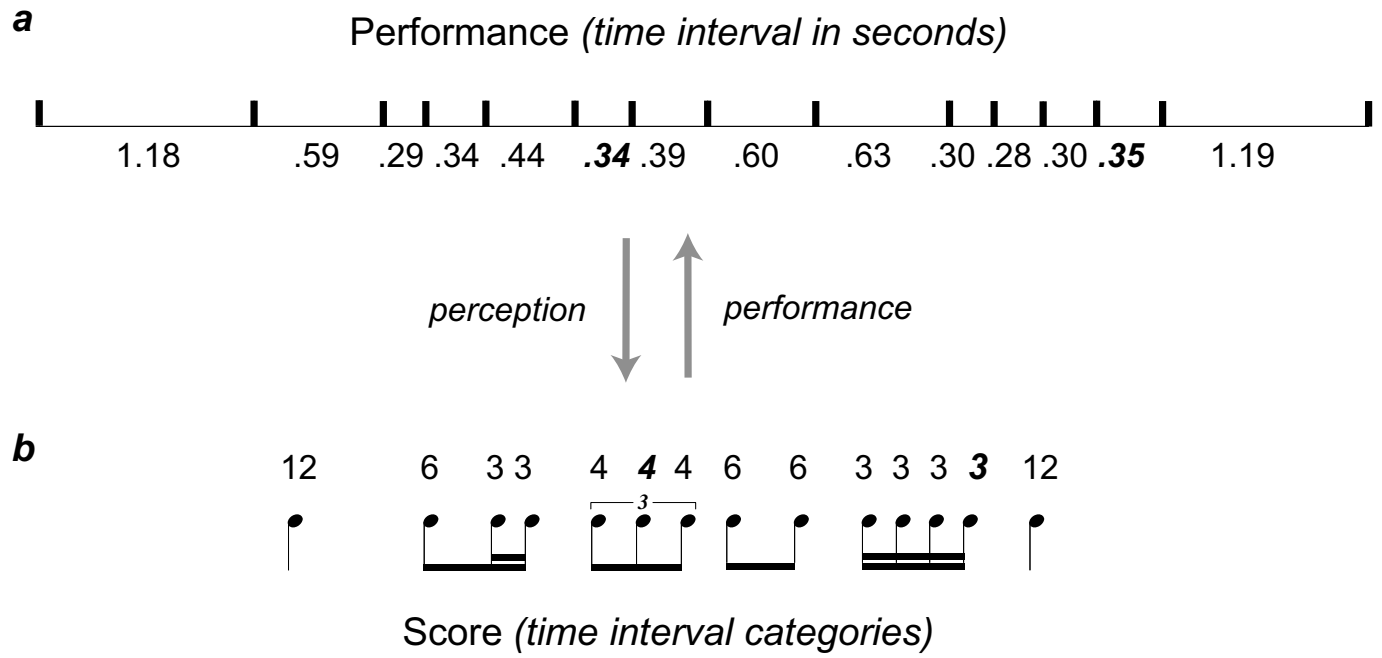
Table 3**Performance Centroids and their Difference with a Mechanical Rendition**

Pattern	Performance Center	Distance from Mechanical
1-1-1	[.324, .312, .363]	[-.009, -.021, +.030]
1-2-1	[.255, .476, .269]	[+.005, -.024, +.019]
2-1-1	[.479, .244, .277]	[-.021, -.006, +.027]
1-1-2	[.246, .263, .491]	[-.004, +.013, -.009]
2-3-1	[.350, .432, .218]	[+.017, -.068, +.051]
4-3-1	[.469, .346, .184]	[-.031, -.029, +.059]
3-1-2	[.462, .179, .358]	[-.038, +.013, +.025]
4-1-1	[.592, .183, .225]	[-.075, +.016, +.058]
1-3-2	[.184, .465, .351]	[+.018, -.035, +.018]
3-1-4	[.330, .170, .500]	[-.045, +.045, -.001]
1-1-4	[.181, .192, .628]	[+.014, +.025, -.039]
2-1-3	[.333, .189, .478]	[-.000, +.022, -.022]

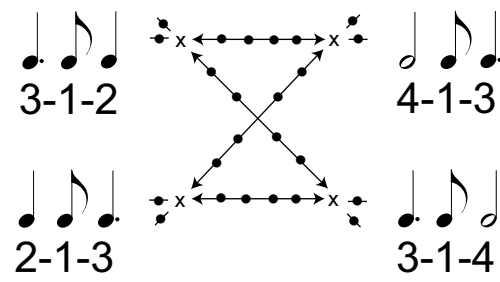
Table 4**Response Proportions in the No Meter Condition and the Differences Between Conditions**

Response	No meter	Duple meter		Triple meter		Triple	
		vs no meter		vs no meter		vs duple meter	
1-3-2	.035	-.030	***	.036	**	.066	***
1-1-4	.029	-.011		.029	**	.040	***
2-3-1	.037	-.009		.026	*	.035	***
4-1-1	.036	-.006		.029	**	.035	***
3-1-2	.036	-.004		.019	*	.023	*
2-1-3	.024	.001		.014		.013	
1-1-1	.084	.004		-.012		-.016	
4-3-1	.040	-.004		-.029	***	-.025	***
3-1-4	.026	.016	*	-.018	**	-.034	***
1-1-2	.063	.008		-.039	***	-.047	***
2-1-1	.066	.008		-.048	***	-.056	***
1-2-1	.074	.039	**	-.056	***	-.095	***

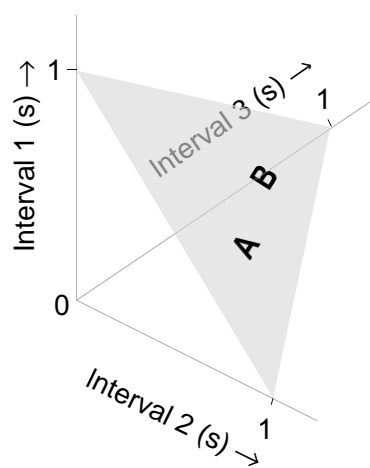
* $p < .05$. ** $p < .01$. *** $p < .001$.



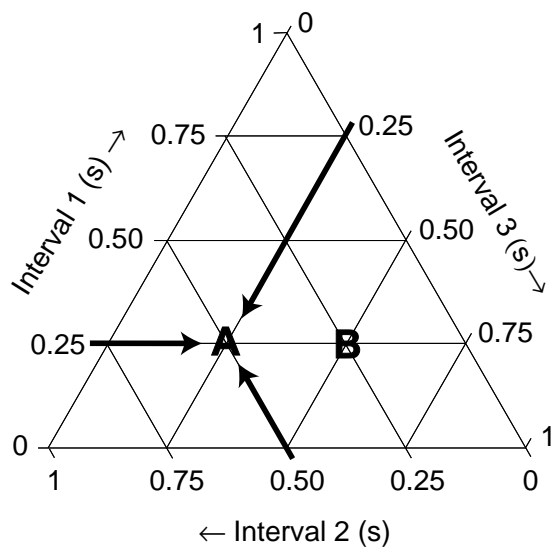




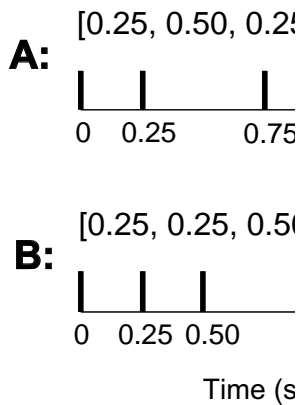
a



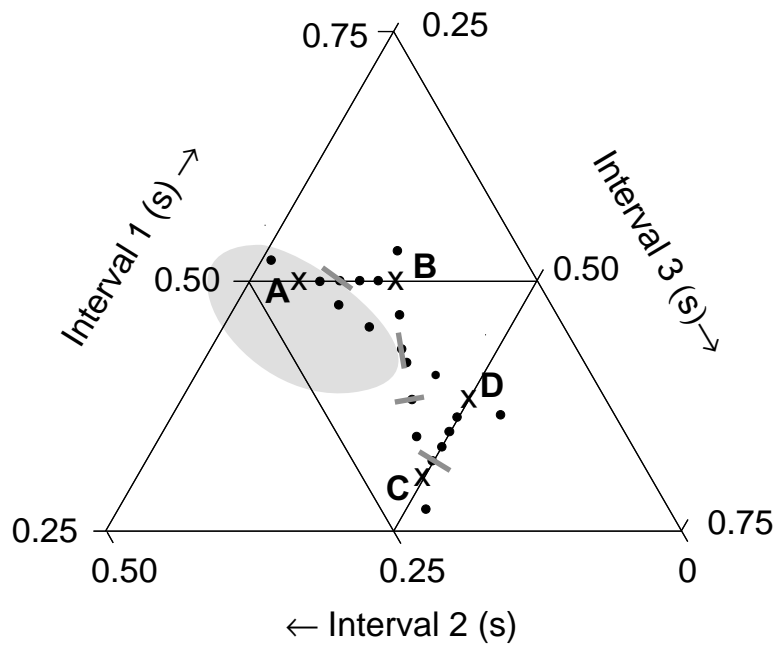
b



c



Schulze, 1989 (N=2)



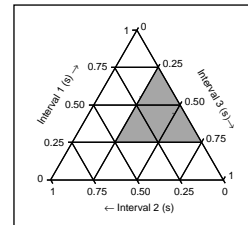
Rhythms:

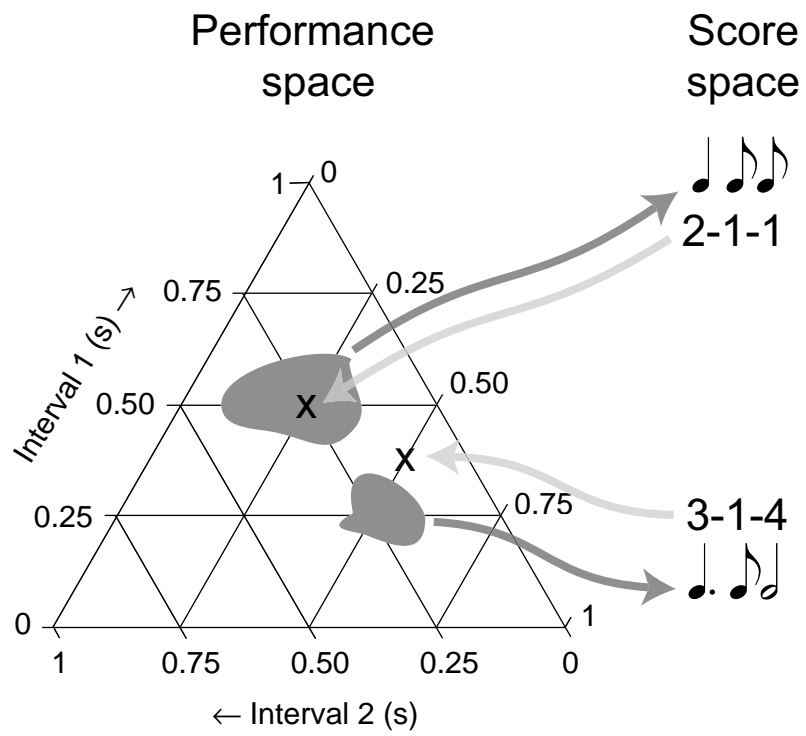
A: 3-1-2

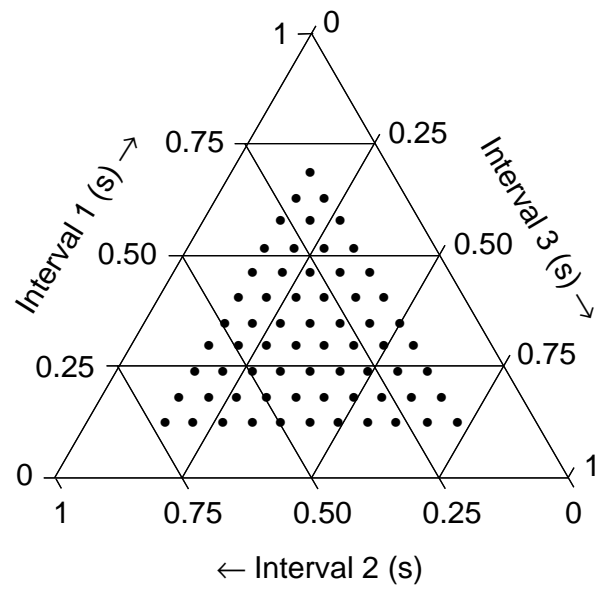
B: 4-1-3

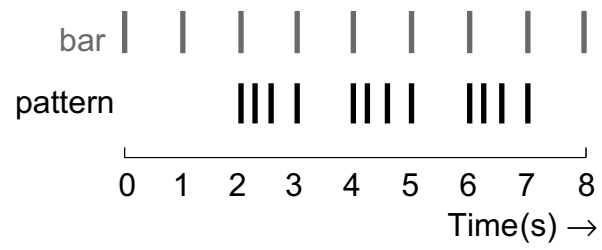
C: 2-1-3

D: 3-1-4









Time signature

2/16 2/8 2/4 2/2 4/16 4/8 4/4 6/16 6/8 6/4

3/16 3/8 3/4 3/2 9/16 9/8

Tuplets

┌ 2

2 ─┐

┌ 3

3 ─┐

┌ 5

5 ─┐







Extra







•

⌒

⏟

Notes and Rests




     

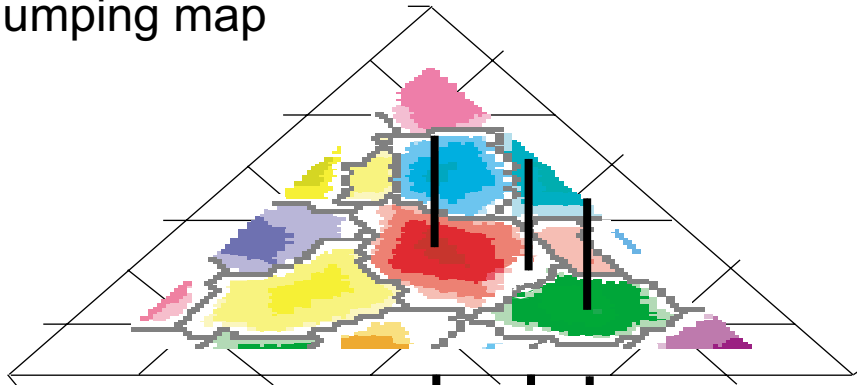
Control

Undo

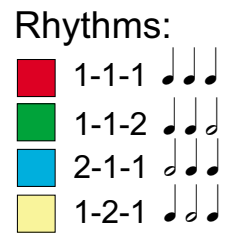
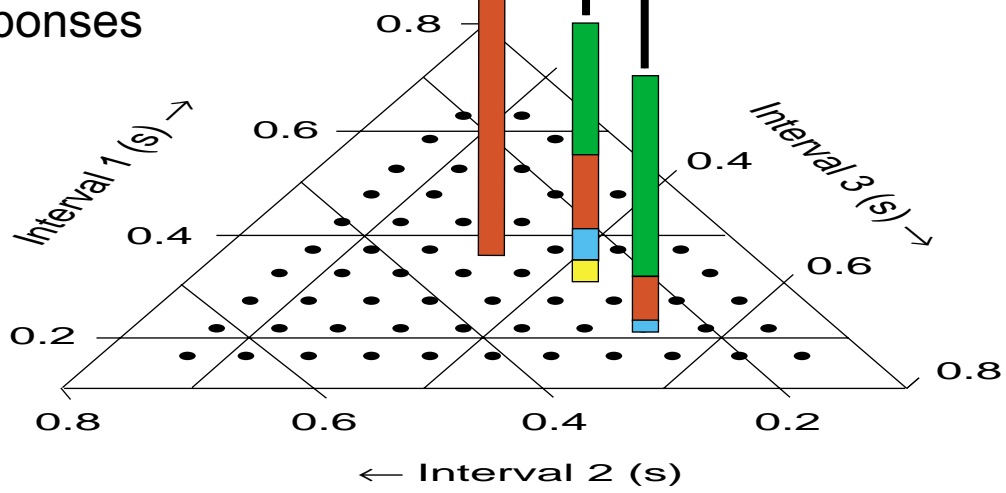
OK

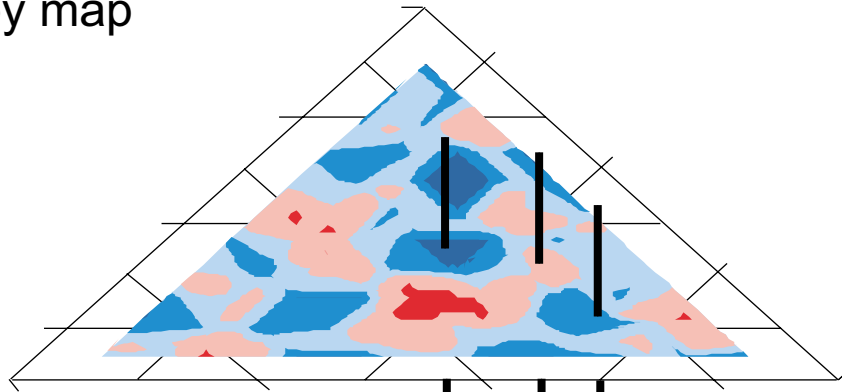
Time clumping map



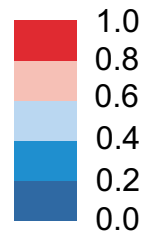
Responses



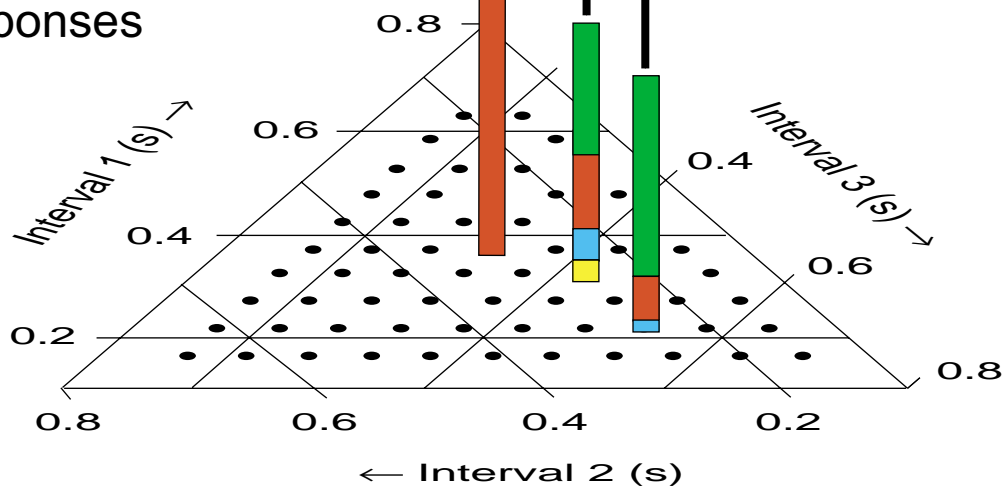
Entropy map



Relative Entropy:



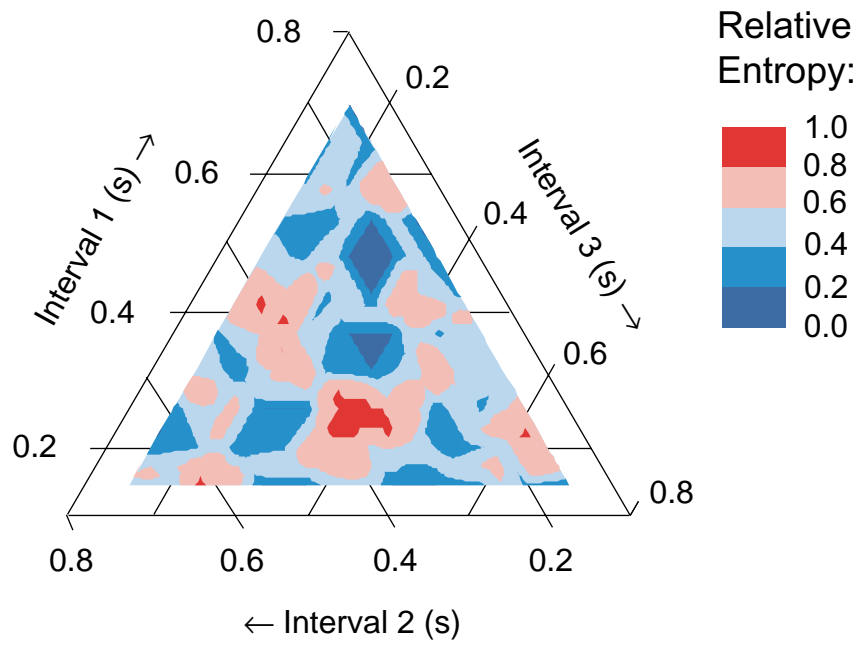
Responses



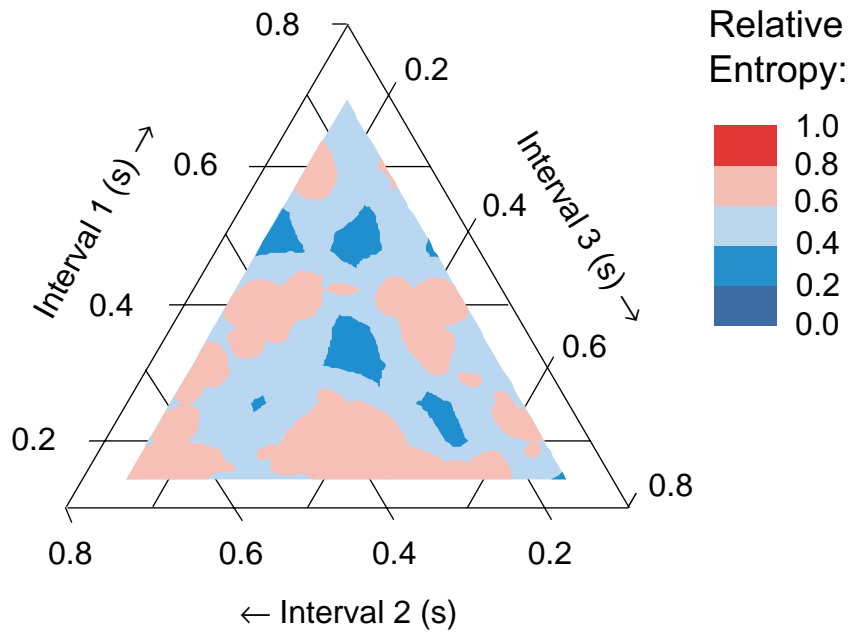
Rhythms:



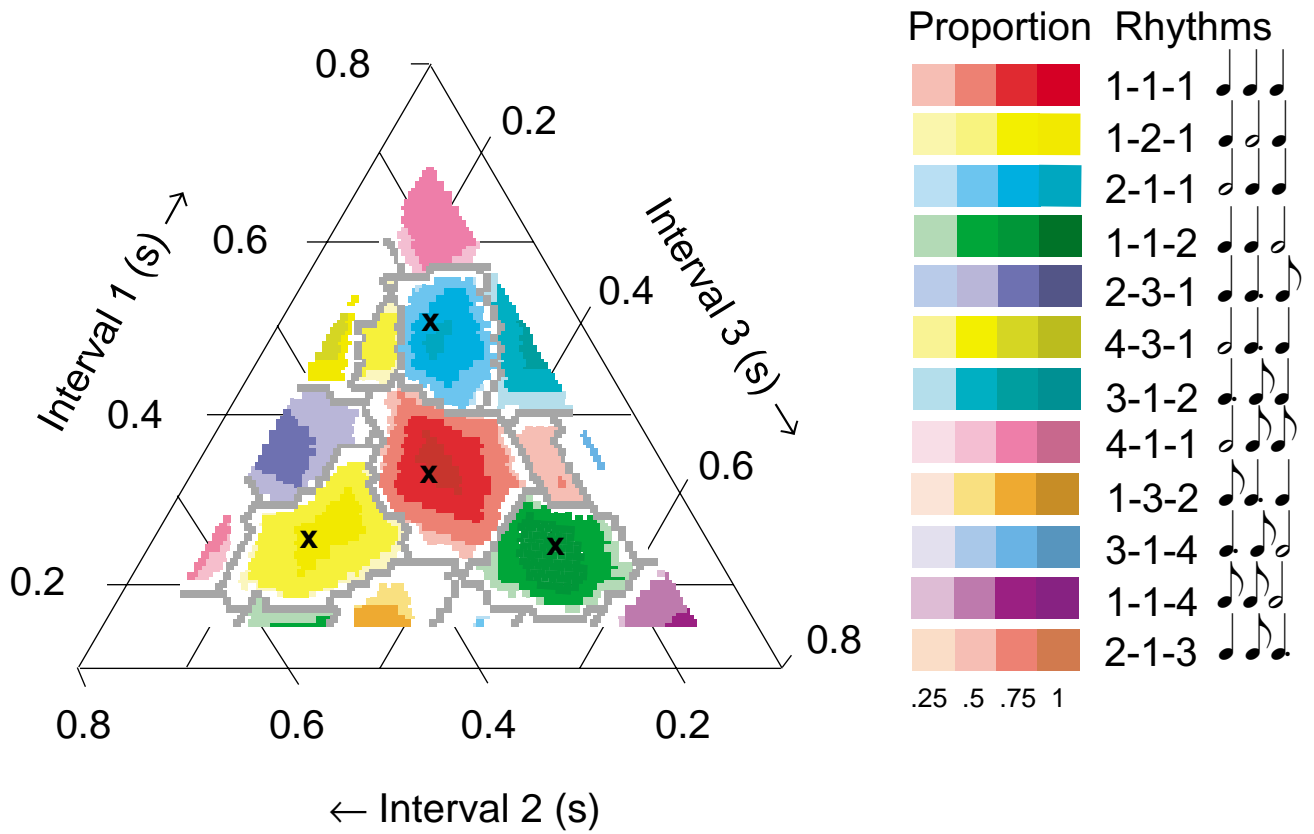
Within Participant (N=1, 6 repeats)

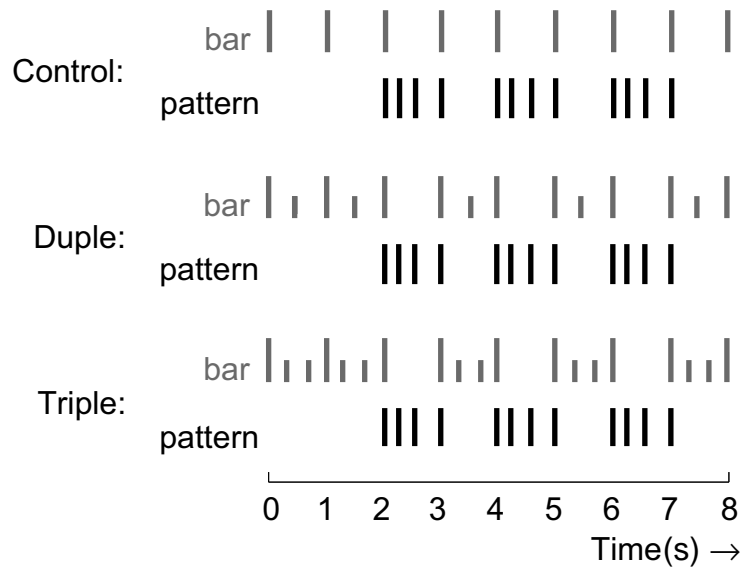


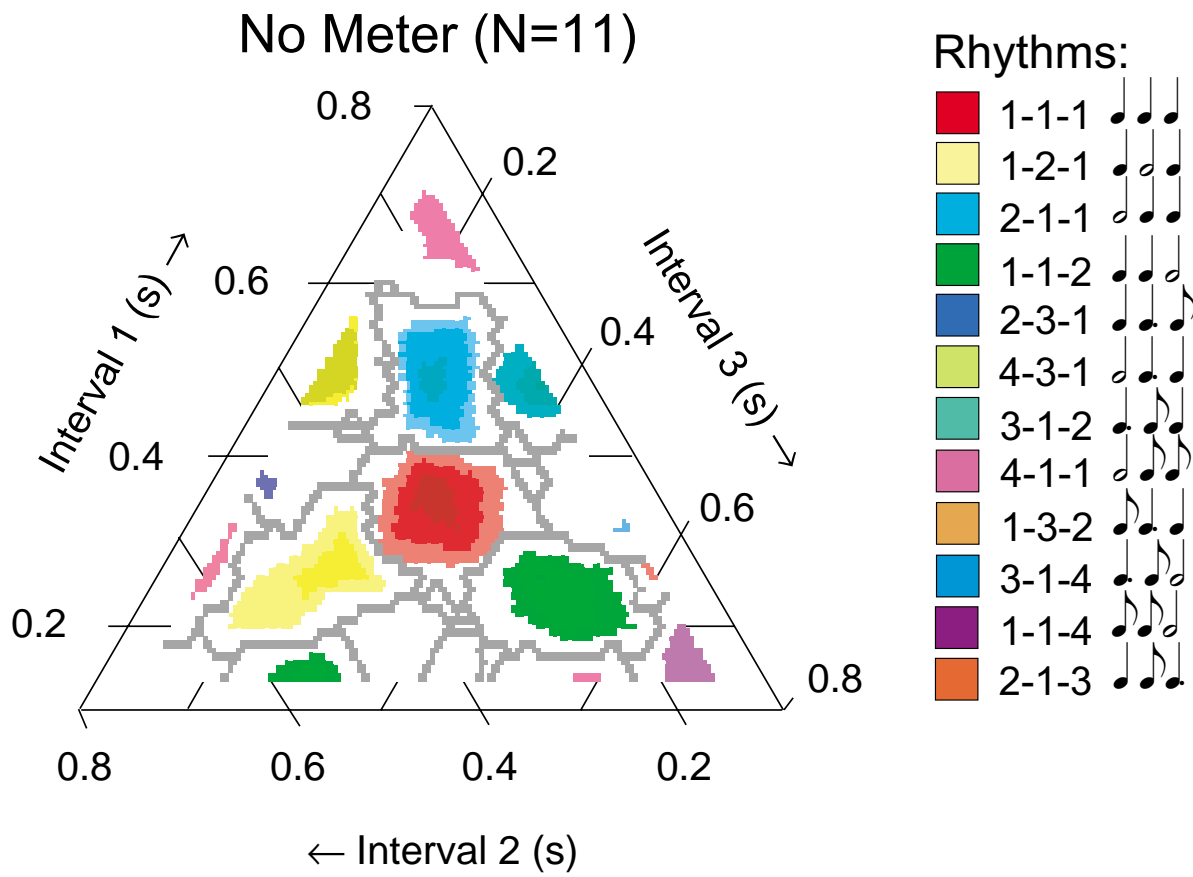
Between Participants (N=29)

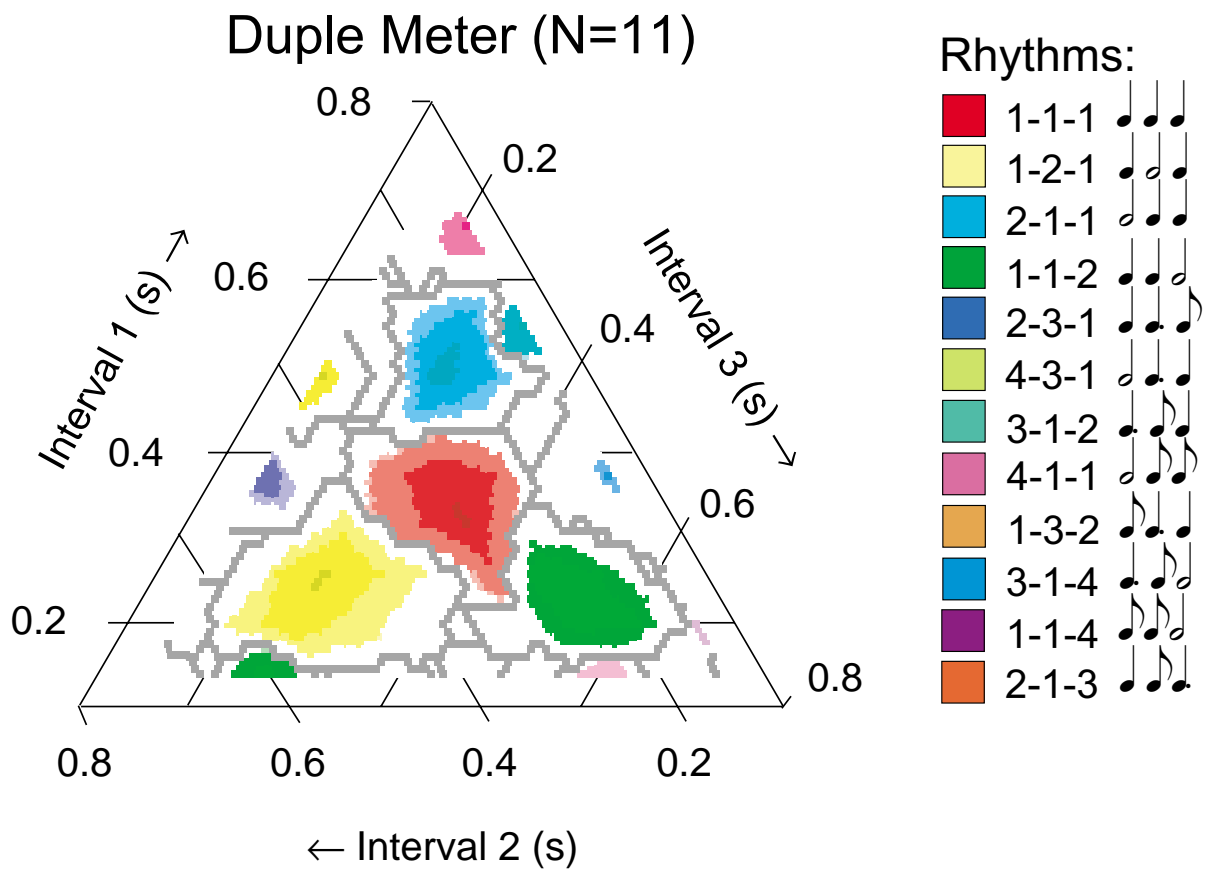


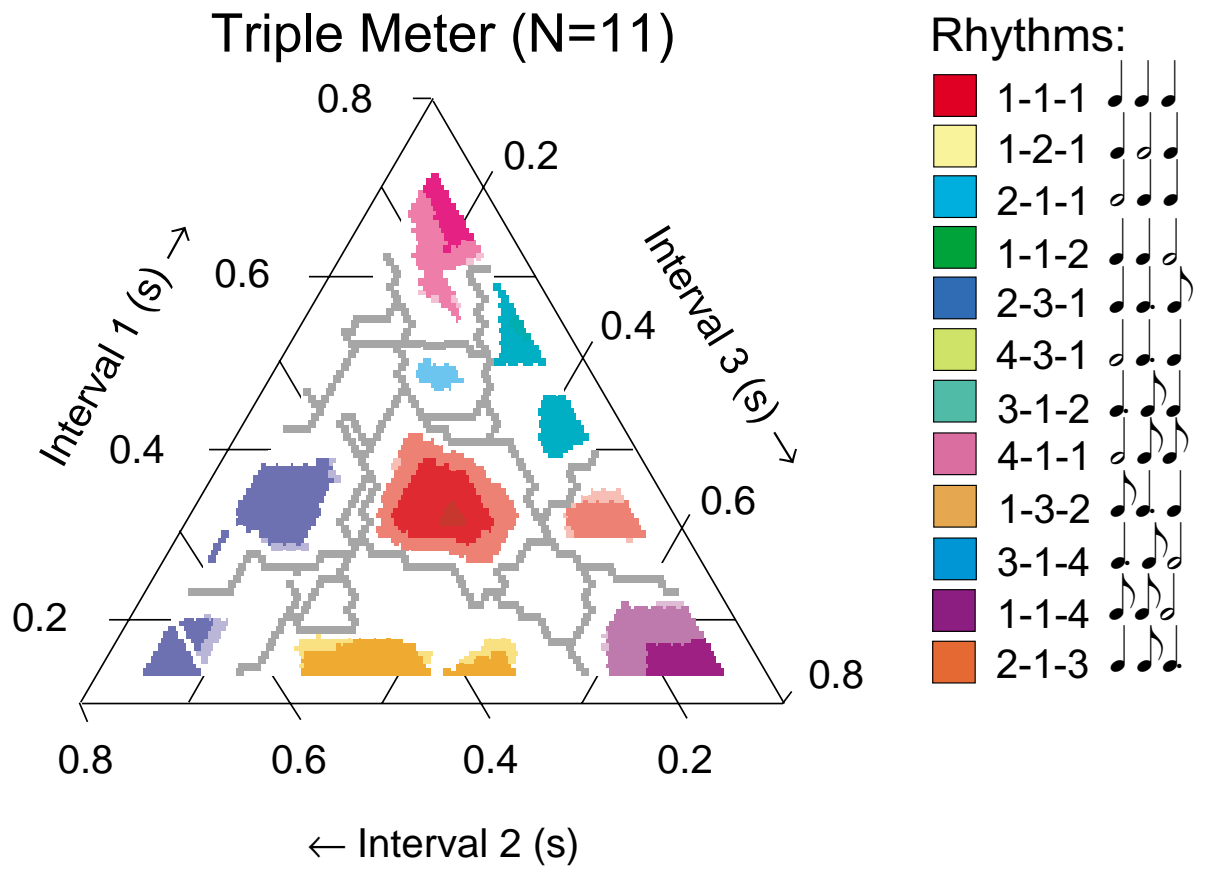
Time Clumping Map (N=29)











Footnotes

* Both authors contributed equally to this work.

1 Note that in Common Music Notation (CMN) several notational variants of the same underlying integer representation are possible: the notation shown is an often encountered one.

2 Note that note symbols are used throughout to represent inter-onset intervals (IOIs), hence the last note is notated as durationless (i.e., marked with a cross).

3 In this paper all stimuli will be represented in the same format, e.g., the pattern A as [.25, .5, .25].

4 However, note that a score space is not a space in the strict mathematical sense (it has no metric). We use the term here loosely to refer to the domain of well formed-sentences in the grammar of CMN.

5 Both quantization models are able to derive the score shown in Figure 1b from the performed rhythm shown in Figure 1a.

6 Relating these predictions to empirical data will be topic for a future paper.

7 Note that IOIs of stimulus patterns are presented not as multiples of $1/19^{\text{th}}$ of a second but are rounded to milliseconds for readability. This causes sometimes a small rounding error, e.g., a pattern summing to slightly less than one second duration. However, in the stimulus preparation and response analysis software a greater precision was used.

8 In this paper we represent performances (stimuli) always as a sequence of inter-onset intervals which makes the representation independent of the actual durations of the short percussive sounds.

9 As another example of the workings of this grammar, consider the partial result as entered by the user in Figure 8: a triplet of a quarter note and an eighth note in 2/4 meter. In this state the buttons representing notes and rests that would exceed the duration of the bar are disabled (e.g., the quarter rest is disabled because still a third note needs to be added).

10 The response proportions shown in the bottom of Figure 9 are slightly idealized for instructive purposes. For detailed information on the actual responses (as depicted in the entropy and time clumping maps later in this paper), see www.nici.kun.nl/mmm/time.html.

11 A check for an increased discrimination sensitivity around the category boundaries that are identified in our experiment will be left for future research.

12 A set of convex individual response distributions might easily combine into a non-convex tiling for the whole population, for instance when a group of expert participants would distinguish a small rhythmic category which lays inside the larger, simpler category of less sophisticated users. It is thus all the more surprising that we find such highly convex tiling in our data, as the converse, non-convex distributions adding up to a convex one is much less likely.

13 That is why the term tiling is appropriate.

14 Cf. the discussion in music performance research as to whether a mechanical performance can be used as norm in measuring expressive timing (Clarke, 1995; Timmers & Honing, 2002). Our findings suggest that a performance centroid (i.e., the most communicative rendition of a category) might be a better norm or reference.

15 Because of boundary effects in the space of presented stimuli they cannot be reliably estimated for the other categories.

16 In the light of the results of Experiment 1, presenting a mechanical subdivision might not be the best way to prime a meter. However, a proper expressive rendition would be hard to construct without creating, e.g., a dependency on musical style. Therefore we resorted to this solution.

17 See Desain & Honing (2001) for a study of the first two models referred to, and their success in explaining the empirical data presented in this paper.

18 For an appendix containing all experimental data, audio examples and animations, see www.nici.kun.nl/mmm/time.html.