

Perceiving the initial note: Quantitative models of how listeners parse cyclical auditory patterns

Minhong Yu¹ · Laura Getz¹ · Michael Kubovy¹

Published online: 3 September 2015
© The Psychonomic Society, Inc. 2015

Abstract In this paper, we explore the rules followed by the auditory system in grouping temporal patterns. Imagine the following cyclical pattern (which we call an “auditory necklace”—AN for short—because those patterns are best visualized as beads arranged on a circle) consisting of notes (1s) and rests (0s): ...1110011011100110 It is perceived either as repeating 11100110 or as repeating 11011100. We devised a method to explore the temporal segmentation of ANs. In two experiments, while an AN was played, a circular array of icons appeared on the screen. At the time of each event (i.e., note or rest), one icon was highlighted; the highlight moved cyclically around the circular array. The participants were asked to click on the icon that corresponded to the note they perceived as the starting point, or clasp, of the AN. The best account of the segmentation of our ANs is based on Garner’s (1974) run and gap principles. An important feature of our probabilistic model is the way in which it combines the effects of run length and gap length: additively. This result is an auditory analogue of Kubovy and van den Berg’s (2008) discovery of the additivity of the effects of two visual grouping principles (proximity and similarity) conjointly applied to the same stimulus.

Keywords Perceptual organization · Temporal Processing · Audition

Introduction

The temporal parsing of auditory patterns is a form of temporal grouping. The twin problems of *temporal grouping* and *meter* are the main puzzles of auditory temporal organization (Drake, 1998; Lerdahl & Jackendoff, 1983).

The perception of *meter* involves the extraction of the *pulse* of a rhythmic sequence (Cooper & Meyer, 1960)—the rate at which we tap our foot to the sound of music (Drake, 1998). Once the listener hears the pulse of a sequence, its pattern of strongly and weakly accented beats causes a hierarchical structure to be perceived (Essens, 1986). Meter is an emergent property of rhythmic organization, just as symmetry is an emergent property of visual organization (Handel, 1998).

Grouping refers to the segmentation of a sequence of sounds into units on the basis of its duration, pitch, intensity, or timbre (Bregman, 1990; Handel, 1989). Whereas the perception of meter is a learned top-down process (Drake et al. 2000; Jones, 1976; Large & Jones, 1999), grouping is a bottom-up process (Handel, 1998): sensitivity to rhythmic grouping is immediate (Hébert & Cuddy, 2002) and it is seen in infants as young as 3 months old (Demany et al. 1977).

Research on auditory grouping falls into three classes: (a) *The perception of accents*: how the perceived accent pattern of two- or three-note rhythms is affected by the loudness and duration of its notes (Povel & Okkerman, 1981; Woodrow, 1911); (b) *Auditory scene analysis*: how listeners separate parallel temporal patterns into their component

✉ Laura Getz
lauragetz@virginia.edu

Minhong Yu
minhongyu@virginia.edu

Michael Kubovy
kubovy@virginia.edu

¹ Department of Psychology, University of Virginia,
PO Box 400400, Charlottesville, VA 22904-4400, USA

streams (Bregman, 1990); and (c) *Parsing*: how listeners determine the starting point of cyclical rhythmic patterns (Garner, 1974; Preusser et al. 1970; Royer & Garner, 1966; 1970).

In this study we have two goals. First, we undertake to quantify the principles that govern the parsing of ambiguous cyclical rhythm patterns. Secondly, we wish to compare these principles to principles of visual grouping, because our understanding of perceptual organization is by and large based on studies of visual stimuli (Kubovy et al. 1998; Kubovy & van den Berg, 2008; Peterson & Gibson, 1994; Peterson & Lampignano, 2003).

Using ambiguous dot lattices as a tool, Kubovy and his colleagues (Kubovy et al. 1998; Kubovy & Wagemans, 1995) found that visual grouping by proximity was lawful and proposed a probabilistic model to account for this regularity. Furthermore, Kubovy and van den Berg (2008) investigated how the strengths of two grouping principles combined to determine visual grouping. They found that the effects of grouping by proximity and grouping by similarity were additive. Thus, when two visual grouping principles are conjointly applied to a visual stimulus, “the whole is equal to the sum of its parts.”

In the following sections, we describe our stimuli (auditory necklaces) and present several models that could predict their grouping structure. We then describe two studies in which we confront these models with empirical data.

Auditory necklaces

We call the auditory patterns in our studies *auditory necklaces* (a concept borrowed from combinatorics; Ruskey 2011) because they are best visualized as beads arranged on a circle. Figure 1 shows two common representations of a binary auditory necklace (AN) of length 8, where a red bead stands for a note, and a grey bead stands for a rest. This AN can also be represented as a single string of binary digits, where 1 stands for a note, and 0 stands for a rest; i.e., 11100110.

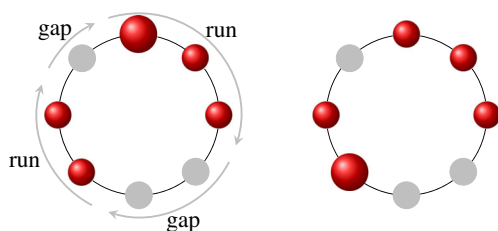


Fig. 1 Two common representations of the same AN: 11100110. Time proceeds in a clockwise direction within each necklace. In each representation, we have made the note perceived as the starting point (i.e., the clasp) larger

Following Garner’s terminology, in our ANs a *block* is a sequence of consecutive identical events (be they notes or rests). When a block consists of notes, it is a *run*; when it consists of rests it is a *gap*. For example, Fig. 1 depicts a four-block AN with two runs (111 and 11) and two gaps (00 and 0).

The question raised by Garner and pursued here is the following: If an AN is played cyclically so that it has no perceptible initial note, which note do listeners choose as the beginning of the pattern? In our parlance, can we predict which note will be perceived as the *clasp* of the AN? In theory listeners could conceivably perceive any note as the clasp, but as we explain later, the clasp is most likely to be the first note of a run.

The ANs we use in this study are ambiguous. In Fig. 1, we illustrate the two ways the AN 11100110 is typically heard: in each panel the clasp is indicated by a larger bead. For example, if one perceives this AN as if it were the pattern 11100110 repeating itself, then we say that underlined 1 is its clasp.

The AN we have considered so far have both runs and gaps. When an AN has gaps, we say that it is *sparse* (Fig. 2a). In this work we will also study *dense* ANs. Consider two complementary ANs whose runs fit into the gaps of the other, and whose notes differ in some respect (pitch or loudness, for example). For example, $N = 11100110$ and $M = 00022002$ are complementary. When they are combined, we get a *dense* AN (Fig. 2b): 11122112, which has no rests and hence no gaps.

Approaches to the perceptual organization of auditory necklaces

We now compare four approaches to the prediction of the clasp of an AN. We believe that only after we have a quantitative model of this form of grouping, can we undertake a search for a mechanism. The [Appendix](#) provides details on the computation of the metrics involved in each approach.

Run and gap principles

In their seminal work, Garner and his colleagues (Garner, 1974; Preusser et al. 1970; Royer & Garner, 1966; 1970) proposed two principles for the segmentation of sparse ANs: (a) the *run principle*, according to which the clasp is perceived as the first note of the longest run, and (b) the *gap principle*, according to which the clasp is perceived as the first note following the longest gap. For example, the run principle predicts that 11100110 will be heard as 11100110 whereas the gap principle predicts that it will be heard as 11011100.

Fig. 2 Two types of auditory necklace. Red and blue beads represent different notes, and grey discs represent rests



(a) A *sparse auditory necklace*: 11100110 consists of two runs, two gaps, and has two potential clasps.

(b) A *dense auditory necklace*: 11122112 consists of four runs and has four potential clasps.

They also discussed the organization of dense ANs. The organization of 11122112 depends on (a) the selection of 11100110 or 00022002 as the figure, while relegating the other to the background, and (b) the run and gap principles operating on each of the complementary sparse ANs.

Garner and colleagues conjectured that if the two principles are in agreement, the clasp is stable and emerges readily, but if they disagree, the clasp is ambiguous and takes longer to emerge. To test this hypothesis, they asked participants to report the perceived organization of ANs by pressing a key in synchrony with the pattern or notating the pattern. Although these procedures recorded the participants' impressions faithfully, they were inefficient: each trial took too long. The amount of data collected was thus too small to allow quantitative modeling, though descriptive findings did support their predictions.

***E* measure**

MacGregor (1985) made the first attempt to quantify the likely location of the clasp using a transformation of the run and gap principles based on block sizes and their relative positions within the pattern. Block size is measured as the number of elements in each block (i.e., run or gap). Relative position (i.e., the “enclosure”) of each block is defined as the number of blocks from that block to the closest end block. He proposed the measure (*E*) given by the sum of the cross products of block size (r_i) and the enclosure of each block (e_i) ($E = \sum r_i \cdot e_i$).

MacGregor (1985) predicted that the organization with the lowest *E* value would be perceived most often because it is the least complex pattern organization. Indeed, using a variety of patterns from previous studies, he found an inverse relationship between the *E*-value and frequency of selection.

Local surprise

Boker and Kubovy (1998) developed a measure called *local surprise* based on information theory, and they used it to

predict the segmentation of sparse ANs. The local surprise value is a measure of the predictability of a event at the current position within a given pattern. For example, in 1110000, the first note is less predictable than the second because the event that precedes the first note is a rest whereas the second note is preceded by another note. The third note is more predictable than the second because it is preceded by two notes whereas the second is only preceded by one. Boker & Kubovy conjectured that the less predictable a note, the more likely it is to be perceived as the clasp (which is in line with research on subjective accents, e.g., Cooper and Meyer 1960).

In their experiments, they asked participants to strike a key on a synthesizer keyboard at the moment they heard the clasp. This allowed them to collect voluminous data. They modeled their data using local surprise as well as the gap and run principles. The local surprise measure provided a better model fit than the run and the gap principles.

Although a productive means of data collection, this method has two drawbacks: (a) Participants had trouble synchronizing their responses with the tones; their responses often preceded or followed the note. To analyze the data, Boker & Kubovy had to decide which note a response aimed for, introducing noise into the data and complicating the analysis. (b) The task confounds the contributions of motor control and perception.

Predictive power

van der Vaart (2009) improved the original Boker and Kubovy algorithm by adding a new measure, *predictive power*, to predict the segmentation of ANs. Whereas local surprise only considers *backward information*, the information before a point in an AN, predictive power considers *forward information*, the information after a point. For example, in 1110000, the second note has larger predictive power than the third because the event that follows the second note is also a note, whereas the third note is followed by a rest. The first note has even larger predictive power than the second note because it is followed by two notes whereas the second is only followed by one.

Preliminary results from a study using methods similar to Boker and Kubovy (1998) suggest that integrating local surprise ratings and predictive power ratings may result in a better-fitting model than local surprise alone.

Our studies

We devised a new method that (a) allows participants easily and quickly to report the clasp, thus allowing us to obtain enough data to build quantitative models; (b) does not require participants to synchronize their taps with the clasp so that the data reflects perception alone.

At the beginning of each trial, a circular array of n icons (where n = the length of the AN) appeared on the screen (Fig. 3). The computer randomly assigned icons to positions around the circle, and randomly associated the top icon with one of the events (a note or a rest) of the AN. While the AN was played (over headphones), a square highlighted the corresponding icon and moved clockwise as each note or rest played. The participants were instructed to click at any time on the icon corresponding to the note they perceived as the clasp.

We created 49 ambiguous ANs and asked two questions: (a) Which of the approaches described above best accounts for the temporal organization in these patterns? and (b) When two temporal grouping principles are conjointly applied to a stimulus, is their conjoint effect equal to the sum of their separate effects? In Study 1, we explored the segmentation of four-block *sparse* ANs and in Study 2 we explored the segmentation of four-block *dense* ANs.

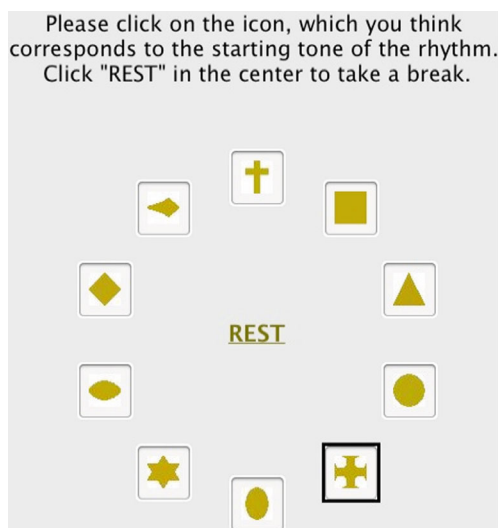


Fig. 3 Screen shot of the display for a ten-event AN. At the moment depicted, the *cross* is the highlighted note/rest

Study 1: Sparse Auditory Necklaces

Method

Participants Ten students from the University of Virginia volunteered. All reported normal or corrected-to-normal vision and normal hearing. We excluded one of them because of a misunderstanding of the instructions.

Stimuli In this experiment, we used four-block sparse ANs. Each AN contained two runs (A and B) and two gaps (A and B). Gap A precedes run A and gap B precedes run B. We generated 49 sparse ANs by crossing seven run ratios (the ratio between the lengths of run A and run B: {1 : 3, 1 : 2, 2 : 3, 1 : 1, 3 : 2, 2 : 1, 3 : 1}) with seven gap ratios (the length ratio between gap A and gap B: {1 : 3, 1 : 2, 2 : 3, 1 : 1, 3 : 2, 2 : 1, 3 : 1}).

Table 1 lists these ANs. Only 25 of the 49 ANs are unique (e.g., 10001110 and 11101000 are rotations of the same pattern), but we treated them as different ANs because we assigned run A, run B, gap A and gap B differently for each pattern (e.g., 1 is run A and 111 is run B in the first example, whereas 111 is run A and 1 is run B in the second). This allowed us to fully cross the seven levels of Run ratio and the seven levels of Gap ratio.

The notes were 440-Hz pure tones lasting 50 ms (including 5 ms linear fade-in and 5 ms linear fade-out). The stimulus-onset asynchrony (SOA) between successive events was 200 ms. To eliminate the bias of selecting the first heard note as the clasp, each pattern was played very fast (SOA = 60 ms) at first and decelerated to a steady SOA = 200 ms after 20 events (Fig. 4). The first played event was randomly selected for each trial.

For visual stimuli, we used ten icons with different shapes (Fig. 3). On each trial, the program randomly selected n icons (corresponding to the length of the auditory necklace) and randomly arranged them on the circumference of a circle. The size of each icon was 55 × 55 pixels and the radius of the circle was 150 pixels.

Design Each participant completed 25 blocks of trials. Each block contained the 49 ANs in random order. It took about 5 hours to complete the experiment. The participants were allowed to divide the experiment into as many sessions as they wished. They were required to complete a block before pausing the experiment: after each block they could choose to continue to the next block, or to quit and later pick up where they left off.

Procedure At the beginning of each trial, the screen showed the circular array of icons (Fig. 3). Through headphones, the participants heard the ANs with the first 20 events decelerating. While the pattern was playing, a square

Table 1 Stimuli in experiment 1

Gap Ratio	Run Ratio						
	1/3	1/2	2/3	1/1	3/2	2/1	3/1
1/3	10001110	1000110	110001110	11000110	111000110	1100010	11100010
1/2	1001110	100110	11001110	1100110	11100110	110010	1110010
2/3	100011100	10001100	1100011100	110001100	1110001100	11000100	111000100
1/1	10011100	1001100	110011100	11001100	111001100	1100100	11100100
3/2	100111000	10011000	1100111000	110011000	1110011000	11001000	111001000
2/1	1011100	101100	11011100	1101100	11101100	110100	1110100
3/1	10111000	1011000	110111000	11011000	111011000	1101000	11101000

highlighted the icon corresponding to the currently playing note/rest in a clockwise direction. The participants were asked to click on the icon corresponding to the tone they heard as the beginning of the pattern. If they clicked on an icon corresponding to a rest (as opposed to a note), the program asked them to choose again because we assumed that a clasp cannot coincide with a rest. They could click the REST button in the center of the display anytime to take a break.

Results and discussion

Responses to the first note of a run The median proportion of error responses (i.e., choosing icons corresponding to rests) for the nine participants was 1.1 % (ranging from 0.1 % to 7.9 %). Among the remaining responses, the median proportion of responses to the first note of a run was 99.8 % (ranging from 90.7 % to 100 %).

Figure 5 shows the frequency of responses to the first note of a run (labeled A and B) compared to other responses for one auditory necklace pattern: 1110001100. The errors and the responses not to the first note of a run were either to the note before or after the first note of a run, and may be due to momentary lapses of attention.

Interpersonal concordance To test the extent to which our participants’ choices of clasp were in agreement with one another, we used the R (R Development Core Team, 2013) package irr (Gamer et al. 2010) to compute the

Kendall coefficient of concordance W (where $0 \leq W \leq 1$), corrected for ties, across participants and within stimuli. We found high agreement among our participants’ clasp selections: $W_t = 0.78$ ($p \approx 0$).

Statistical model selection We excluded all trials in which participants did not choose the first note of a run. We could thus classify participants’ responses using a binomial response variable (clasp selection was either at the start of run A or run B). This allowed us to model our data using mixed-effects logistic regression. All of our generalized linear mixed-effects models (GLMMs) were computed using the package lme4 (Bates et al. 2014).

GLMMs, which use maximum-likelihood estimation, have many advantages over traditional repeated-measures analysis of variance, which use ordinary least-squares. In addition to providing estimates of fixed effects, they allow us to predict subject-by-subject variations in model parameters (called random effects). Furthermore, GLMMs do not rest on many of the assumptions required by traditional analyses, such as quasi- F tests, by-subjects analyses, combined

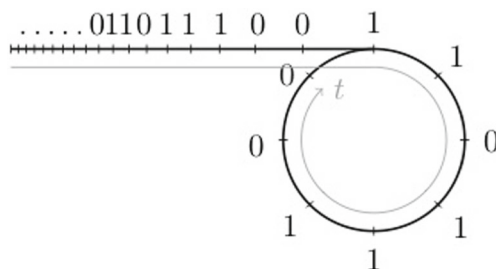


Fig. 4 Decelerating into 11011100

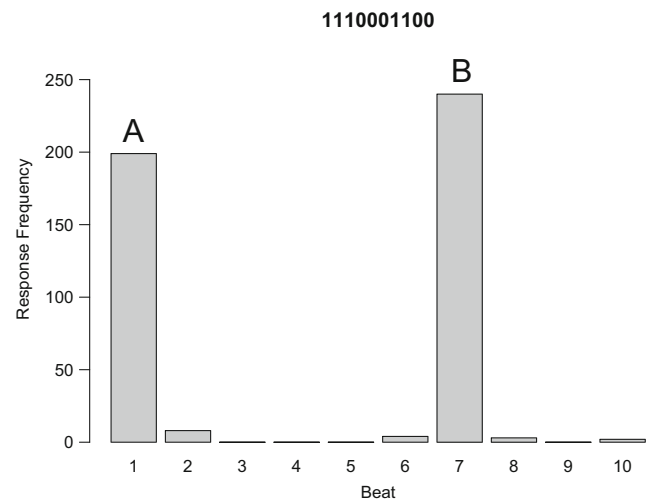


Fig. 5 Frequency of responses to 1110001100 (A11000B100)

Table 2 Study 1: Model comparison of the five models

	K	ΔAIC_c	weight	$R^2_{\text{marg.}}$	$R^2_{\text{cond.}}$
Run/gap ratio (additive)	9	0.00	0.71	0.465	0.678
Run/gap ratio (non-additive)	10	1.78	0.29	0.465	0.678
Predictive power	9	103.27	0.00	0.470	0.677
Local surprise	5	1513.55	0.00	0.435	0.546
McGregor <i>E</i>	5	1907.66	0.00	0.444	0.484

by-subjects and by-items analyses, and random regression (Baayen et al. 2008).

For each of our five GLMMs (run/gap additive, run/gap non-additive, local surprise, predictive power, MacGregor *E*), we treated the predictors derived from the approaches described above (see Appendix for formulas) as fixed effects and the subject-by-subject variation of the intercept and subject-by-subject variation of the slope of the predictors as random effects. To compare the five candidate models, we used a method of model comparison based on the Akaike's Information Criterion (AIC), which offers a principled balance between goodness-of-fit and parsimony (Burnham et al. 2011; Bozdogan 1987, for introductory presentations). Because the probability of overfitting can be substantial when using AIC (Claeskens & Hjort, 2008), we used AICc—which penalizes extra parameters more heavily than does AIC—as recommended by Anderson and Burnham (2002).

Whereas AICc is an appropriate method for model comparison and selection, it tells us nothing about the absolute model fit of a model. To give us an idea of this fit, we computed two types of R^2 for GLMMs using the MuMIn package (Bartoń, 2014). The first, called the marginal R^2 ($R^2_{\text{marg.}}$), estimates the proportion of variance accounted for by the fixed effects only, whereas the second, called the conditional R^2 ($R^2_{\text{cond.}}$), estimates the proportion of variance accounted for by the fixed and random effects taken together (Johnson, 2014; Nakagawa & Schielzeth, 2013).

Best-fitting model Table 2 compares the five models. The best two models are the two versions of the run and gap approach, in which the two predictors are the ratio of the gap lengths and the ratio of the run lengths. The first of these does not include an interaction between the predictors; the second does. The marginal R^2 (which is identical for the top two models) shows a good deal of variance explained by the fixed effects ($R^2_{\text{marg.}} = 0.465$). When adding in the variance of the random effects, an additional $\approx 20\%$ of variance is explained ($R^2_{\text{cond.}} = 0.678$).

The first two models are competitive: a ΔAIC_c of 1.78 implies an evidence ratio (or Bayes factor, see Anderson 2008, Section 4.4) of 2.44, which Jeffreys (1961, p. 432) considers “barely worth mentioning.” There is, however,

no question that the evidence in favor of the first model is immeasurably stronger than the evidence in favor of the third, fourth, and fifth models. An AICc difference of 103 implies an evidence ratio on the order of 10^{22} , which is far beyond what Jeffreys considers “decisive.”

When multiple models are competitive, we are faced with *model uncertainty*. The consensus in the statistical literature is that the best way to deal with such a situation is to construct a compromise model by a process called *model averaging* (Anderson, 2008; Claeskens and Hjort, 2008; Ginestet, 2009; Grueber et al. 2011; Richards et al. 2011; Symonds & Moussalli, 2011).

Table 3 shows the coefficients of the averaged model, their standard error, and a 95 % CI¹. Because the interaction coefficient is almost zero (i.e., -0.03) and the confidence interval straddles zero (95 % CI: $-0.15, 0.09$), we are inclined to favor additive effects of the run and gap principles. Furthermore, with the same scale, the gap principle is more important than the run principle (i.e., participants use the gap principle more often than the run principle in making their clasp selection).

Figure 6 shows the predictions of the averaged model. In this figure, in addition to the data points and their confidence intervals, we plot the predictions of an additive model. Two features of this plot require some clarification.

First, the proportions are plotted on an unevenly spaced *y*-axis. This is because these are binomial data fit using logistic regression. There are two ways to plot the predictions of a logistic regression. One, which we did not use here, is to plot the predicted proportions on a linear *y*-axis, which produces seven sigmoid (i.e., S-shaped) functions, one for each level of gap ratio. We chose to plot the proportions on a log-odds scale (resulting in unevenly spaced proportions on the *y*-axis), which produces seven linear functions.

Second, we plot lines that represent the predictions of an additive model. We did this because it effectively shows that for the most part, the data deviate from the additive model only when the run ratio and gap ratio are 1/3 or 3/1.

¹If the confidence interval for an estimated parameter does not straddle zero, this estimate may be considered significant at $\alpha < 0.05$.

Table 3 Study 1: Coefficient estimates, standard error, and the lower and upper limits of 95 % of the model (in log-odds) resulting from the averaging of the two competitive models in Fig. 2

	Estimate	SE	Lower	Upper
(Intercept)	0.01	0.03	−0.05	0.07
Run ratio	0.76	0.46	−0.14	1.66
Gap ratio	2.89	0.50	1.92	3.87
Run ratio*Gap ratio	−0.03	0.06	−0.15	0.09

Figure 6 shows that as the run ratio and gap ratio increase, the growth of the probability of choosing run A as the clasp approximates a linear function.

Conclusions

First, the data show that participants organized the notes in each run as a perceptual unit and almost always perceived the first note of a run as the clasp, which replicates Preusser et al. (1970) and Royer and Garner (1966; 1970).

Second, although the MacGregor (1985), Boker and Kubovy (1998), and van der Vaart (2009) predictors are mathematically and conceptually more sophisticated, they did not fit the data nearly as well as the Garner models.

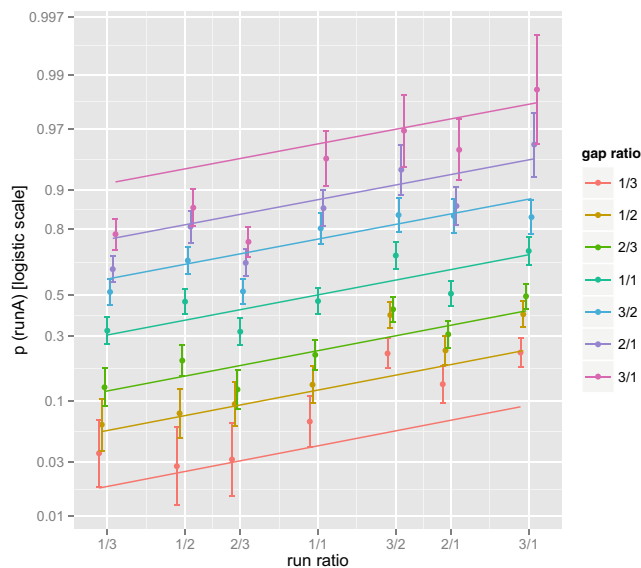


Fig. 6 Study 1: Predictions of the model averaged using the additive and non-additive run and gap models. The individual data points accompanied by 95 % confidence intervals were produced assuming that run ratio and gap ratio are categorical predictors in a 7×7 crossed design. The fitted lines represent the averaged model, assuming that the run ratio and gap ratio are continuous predictors. Note that because we averaged a model with and without the interaction, in some cases the fitted lines do not precisely match the data points

There are several possible reasons for the poor showing of these sophisticated models:

- The E measure takes both run and gap principles into account, but it combines them by summing them with equal weight, which results in a loss of information. This is relevant because the gap principle was found to be more important than the run principle in the current study.
- The predictive power algorithm also takes both runs and gaps into consideration, which may be why it fares better than the local surprise model that only considers gaps. Nonetheless, both are but transformations of run and gap lengths. The complexity of these transformations and their sophisticated rationale appears not to produce better fits than the simple measures of relative run and gap lengths.
- The local surprise model was designed to deal with data involving perception *and* motor skills. The location of the clasp was only one of the three response variables they used. The other two were the temporal accuracy of the tap and its strength (which may reflect the participant's confidence). The latter two do not involve perception.
- Boker and Kubovy (1998) sampled a number of eight-event ANs without considering number of blocks, whereas we used only ANs with four blocks to make the stimuli bistable. Therefore, many ANs they used were more complex than ours. The complexity of the stimuli may favor their information theory model. This remains an open question.

Third, the run length ratio and gap length ratio were found to additively predict the auditory organization. This is in line with what Kubovy and van den Berg (2008) found in the visual domain.

Study 2: Dense auditory necklaces

Method

Participants We paid 14 students from the University of Virginia \$8 an hour for their participation. They all reported normal or corrected-to-normal vision and normal hearing.

Table 4 Study 2: Model comparison of the five models including responses to LOW clasps

	K	ΔAIC_c	weight	$R^2_{\text{marg.}}$	$R^2_{\text{cond.}}$
Run/gap ratio (additive)	9	0.00	0.73	0.569	0.745
Run/gap ratio (non-additive)	10	2.00	0.27	0.569	0.745
Predictive power	9	58.94	0.00	0.575	0.758
Local surprise	5	492.94	0.00	0.581	0.727
McGregor <i>E</i>	5	4764.14	0.00	0.354	0.412

Stimuli We created four-block dense ANs by filling the rests of the sparse ANs used in Study 1 with notes of another pitch. For example, the AN 11100110 in Study 1 became 11122112 in Study 2. Therefore, in the current experiment, each AN contains two runAs and two runBs, which we called *runA*₁, *runA*₂, *runB*₁, and *runB*₂. We treat the *run1* length ratio (*runA*₁/*runB*₁) as we treated runs in Study 1 and the *run2* length ratio (*runA*₂/*runB*₂) as we treated gaps in study 1. The *run1* length ratio and *run2* length ratio again each have seven levels (1:3, 1:2, 2:3, 1:1, 3:2, 2:1, 3:1). Again, only 25 of the 49 ANs were unique (e.g., 12221112 and 11121222 are the same), but we treated them as different ANs since they have different *runA*₁, *runA*₂, *runB*₁, and *runB*₂.

The notes corresponding to 1s were 440-Hz piano MIDI tones (which we called LOWs) and the notes corresponding to 2s were 880-Hz piano MIDI tones (which we called HIGHs). The visual stimuli and all other aspects of the auditory stimuli were the same as Study 1.

Design and procedure The design and procedure were identical to Study 1 except that the dense ANs contain no rests and therefore participants could choose any note as the potential clasp (i.e., there were no incorrect responses due to the selection of a rest).

Results and discussion

Responses to the first note of a run Three participants chose the first note of a run on fewer than 70 % of the trials. They were excluded from further analysis. The remaining 11 participants chose the first note of a run on more than 80% of the trials. The median proportion of those responses

was 98.1 % (ranging from 83.8 to 99.8 %). As in Study 1, we disregarded trials on which participants did not choose the first note of a run.

Model comparison for responses to LOWs All participants chose LOWs (1s) more than HIGHs (2s) as the clasp, with eight of the 11 participants predominantly choosing LOW clasps (> 80%).

Therefore, we fit the five GLMMs (used in Study 1) by including only responses to LOWs as the clasp. We did this by treating the *run1* length ratio as we treated runs in study 1 and the *run2* length ratio as we treated gaps in study 1.

Table 4 compares the five models in terms of AICc. The two run and gap models are again the best and are competitive models. The additive run and gap model is superior to the non-additive Garner model with a ΔAIC_c of 2.00, which implies an evidence ratio of 2.72 (making it a small victory for the additive model). However, the AICc differences between the two run and gap models and other three models show that they are decisively better than other models.

Because the two run/gap ratio models were competitive, we again used the process of model averaging. Table 5 shows the coefficient estimates of the averaged model, their standard error, and a 95 % confidence interval. Both the model comparison results and the confidence interval of the interaction term (95 % CI: -0.16, 0.16) again lead us to favor additive effects of the run and gap principles. Furthermore, with the same scale, the gap principle (here *run2*) was again more important than the run (here *run1*) principle.

Figure 7 shows the predictions of the averaged model. As the run (i.e., *run1*) ratio and gap (i.e., *run2*) ratio increase, the probability of choosing *runA*₁ as the clasp increases

Table 5 Study 2: Coefficient estimates, standard error, and the lower and upper limits of 95 % of the model (in log-odds) resulting from the averaging of the two competitive models in Fig. 4

	Estimate	SE	Lower	Upper
(Intercept)	0.01	0.03	-0.05	0.08
Run (<i>run1</i>) ratio	0.18	0.26	-0.33	0.69
Gap (<i>run2</i>) ratio	-3.72	0.49	-4.69	-2.75
Run ratio*Gap ratio	0.00	0.08	-0.16	0.16

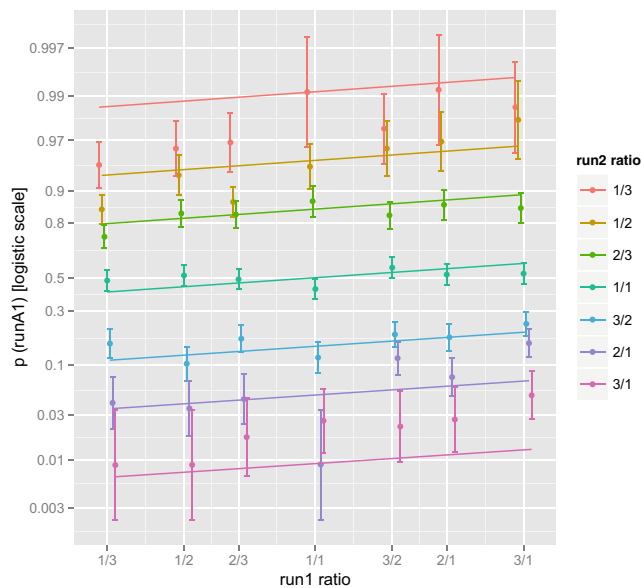


Fig. 7 Study 2: Predictions of the model averaged using the additive and non-additive run and gap models. The individual data points accompanied by 95 % confidence intervals were produced assuming that Run_1 ratio (i.e., $runA_1/runB_1$) and Run_2 ratio (i.e., $runA_2/runB_2$) are categorical predictors in a 7×7 crossed design. The fitted lines represent the averaged model, assuming that Run_1 ratio and Run_2 ratio are continuous predictors

more or less linearly. Additionally, the plotted lines represent the predictions of an additive model; these parallel lines show that again for the most part, the data deviate from the additive model only when the run2 ratio is 1/3 or 3/1. Finally, the regression lines here are flatter than in Study 1. Because gap principle was found to be the stronger predictor of clasp selection in Study 1, the flatness here may be due to the fact that gaps are replaced with $run2s$ in dense ANs.

Statistical models for responses to HIGHS Because most people chose LOW clasps rather than HIGH clasps, none of the five models converged on a solution using the same fixed-effect and random-effect predictors as the previous two analyses. Therefore, it remains an open question whether the pattern of responses would be the same for HIGH as for LOW clasps.

Conclusions

Perceiving the clasp in dense ANs is a two-step process. First, listeners must do a task akin to the figure-ground problem in selecting which notes (LOWs versus HIGHs) will be treated as the figure, while relegating the other set to the background. We found that the majority of participants chose LOWs (1s) more than HIGHs (2s) as the figure. Second, once the figure-ground comparison is complete, participants must choose the clasp based on the run and gap

principles: now $run1s$ are treated as runs in a sparse AN and $run2s$ are treated as gaps.

Using this interpretation, the results for dense ANs (with LOW clasps) replicate the results for sparse ANs in all respects. (a) Participants overwhelmingly perceived the first note of a run as the clasp (Preusser et al. 1970; Royer & Garner, 1966; 1970). (b) The two run and gap models fared much better than the MacGregor (1985), Boker and Kubovy (1998), or van der Vaart (2009) models. (c) People use the gap (i.e., $run2$) principle as a grouping cue more often than they use the run principle. (d) The ratio of run lengths and the ratio of gap lengths additively predicted the auditory organization (a result analogous to Kubovy and van den Berg 2008).

General discussion

We have established three facts:

1. The perceptual grouping of simple, cyclical auditory rhythm patterns can be predicted from two complementary (and not necessarily synergistic) principles, first proposed by Garner and his colleagues. According to the *gap* principle, a cyclically played pattern of sounds appears to start after the longest gap in the sounds. According to the *run* principle, such a pattern appears to start at the beginning of the longest run of sounds.
2. Of these two principles, the gap principle is much stronger than the run principle.
3. To a first approximation, when the two principles imply different parsings of the pattern, they additively affect the probabilities of the two parsings. Thus the generalization to audition of Kubovy and van den Berg's 2008 finding that the conjoined effects of grouping by proximity and grouping by similarity are additive is promising. However, a larger replication is advisable to remove remaining doubts regarding additivity of the run and gap principles.

The generality of our findings is limited in several regards. First, choice of rhythmic grouping may have been affected by our choice of tempo (SOA = 200 ms). Tempo is known to affect discrimination accuracy (Handel, 1992) and the perception of constancy (Handel, 1993) in rhythm patterns. Other work in our lab has also shown that performance of the same rhythm pattern shows systematic differences across tempi (Barton, Getz, & Kubovy, *under review*). It therefore remains an open question how aspects of our stimulus presentation (i.e., tempo, but also frequency and timbre) may change the results.

Second, although our models in Figs. 6 and 7 show an additive relationship whereby increasing the run ratio and gap ratio additively increase the likelihood of choosing run

A as the clasp, we have not determined the limits (i.e., the ceiling and the floor) of the model. It remains an open question at what point the run ratio and gap ratio become so large that they make selection of the clasp unambiguous. It also remains an open question whether more complex ANs will be predicted best by the simple run and gap models or alternatively whether more complex ANs are better explained by more complex models (as found by Boker and Kubovy 1998).

Finally, we should keep in mind that additivity is not inevitable. Temporal grouping and spatial grouping are intertwined in a manner inconsistent with a linear mechanism that could produce additivity (Gepshtein & Kubovy, 2000). Similarly, in a spatial grouping task curvature, density, and proximity were non-additive (Strother & Kubovy, 2012). Kubovy and Yu (2012) conjectured that additive conjoint effects are found when the conjoined grouping principles do not give rise to a new emergent property. This points to the conclusion that our listeners did indeed separate temporal organization into component aspects of grouping and metric structures. Because meter can be thought of as an emergent property of rhythmic grouping (Handel, 1998), it is an open question whether imposing a stronger metric grid onto the rhythms used here would result in non-additivity of the grouping principles.

Acknowledgments This research was supported by National Science Foundation Grant 1027120 (Michael Kubovy, PI) and the Joseph Arwood Memorial Fund, which honors the memory of a young colleague who showed great promise and contributed much during the initial phases of these studies. We thank Bill Epstein, Steven Scheid, and Judith Shatin for helpful comments on previous drafts of this manuscript.

Appendix: Quantification of the approaches

In our experiments, we manipulated the lengths of runs and gaps. Let r_A and r_B be lengths of run A and run B, and g_A and g_B be lengths of gap A and gap B.

For each potential clasp, we first computed the overall strength of the each principle (i.e., run principle, gap principle, E , local surprise, and predictive power). Then we computed the relative strength between the two competing clasps for each principle by taking either the ratio or the difference depending on the attributes of the quantification. We used these relative strengths (and their interactions, if any) as predictors in our probabilistic models.

Run and gap ratios For the two clasps, r_A and r_B represent the strength of the run principle and g_A and g_B represent the strength of the gap principle. We calculated the log-ratio between them (run ratio and gap ratio) as a measure of relative strength. The run ratio and

gap ratio served as predictors in the additive run/gap model. The run ratio, gap ratio, and their interaction (run ratio \times gap ratio) served as predictors in the non-additive run/gap model.

$$\text{Run ratio} = \log \frac{r_A}{r_B} \tag{1}$$

$$\text{Gap ratio} = \log \frac{g_A}{g_B} \tag{2}$$

MacGregor E We calculated the E measures for the two potential clasps (E_A and E_B). We calculated the difference between them (Δ_E) as a measure of relative strength. Δ_E was entered as the predictor in the E model.

$$E_A = r_A \times 1 + g_B \times 2 + r_B \times 2 + g_A \times 1 \tag{3}$$

$$E_B = r_B \times 1 + g_A \times 2 + r_A \times 2 + g_B \times 1 \tag{4}$$

$$\Delta_E = E_A - E_B \tag{5}$$

Local surprise We calculated the local surprise measures for the two potential clasps. We calculated the ratio between the two local surprise values (R_{L_s}) as a measure of relative strength, which served as a predictor in the local surprise model.

$$L_{A1} = \frac{\log_2 \frac{1}{g_A+1}}{g_A \log_2 \frac{g_A}{g_A+1}} \tag{6}$$

$$L_{B1} = \frac{\log_2 \frac{1}{g_B+1}}{g_B \log_2 \frac{g_B}{g_B+1}} \tag{7}$$

$$R_{L_s} = \frac{L_{A1}}{L_{B1}} \tag{8}$$

Predictive power We computed the predictive power measures for the two potential clasps. We calculated the ratio between the two predictive power values (R_{P_s}) as a measure of relative strength. R_{P_s} and R_{L_s} were both entered as additive predictors in the predictive power model.

$$P_{A1} = r_A * \frac{\log_2 \frac{r_A}{r_A+1}}{\log_2 \frac{1}{r_A+1}} \tag{9}$$

$$P_{B1} = r_B * \frac{\log_2 \frac{r_B}{r_B+1}}{\log_2 \frac{1}{r_B+1}} \tag{10}$$

$$R_{P_s} = \frac{P_{A1}}{P_{B1}} \tag{11}$$

Example pattern Using the pattern 1110001100 (see Fig. 5) as an example, the lengths of the runs are 3 and 2, so $r_A = 3$ and $r_B = 2$. The lengths of the gaps are 3 and 2, and because gap A precedes run A and gap B precedes run B, $g_A = 2$ and $g_B = 3$. Therefore:

$$\text{Run ratio} = \log \frac{r_A}{r_B} = \log \frac{3}{2} = 0.176 \quad (12)$$

$$\text{Gap ratio} = \log \frac{g_A}{g_B} = \log \frac{2}{3} = -0.176 \quad (13)$$

$$\Delta_E = E_A - E_B = (3 \times 1 + 3 \times 2 + 2 \times 2 + 2 \times 1) - (2 \times 1 + 2 \times 2 + 3 \times 2 + 3 \times 1) = 0 \quad (14)$$

$$R_{Ls} = \frac{L_{A1}}{L_{B1}} = \frac{\frac{\log_2 \frac{1}{2+1}}{2 * \log_2 \frac{2}{2+1}}}{\frac{\log_2 \frac{1}{3+1}}{3 * \log_2 \frac{3}{3+1}}} = \frac{1.355}{1.606} = 0.844 \quad (15)$$

$$R_{Ps} = \frac{P_{A1}}{P_{B1}} = \frac{3 * \frac{\log_2 \frac{3}{3+1}}{\log_2 \frac{1}{3+1}}}{2 * \frac{\log_2 \frac{2}{2+1}}{\log_2 \frac{1}{2+1}}} = \frac{0.623}{0.738} = 0.844 \quad (16)$$

References

- Anderson, D.R. (2008). *Model Based Inference in the Life Sciences: A Primer on Evidence*. New York: Springer.
- Anderson, D.R., & Burnham, K.P. (2002). Avoiding pitfalls when using information-theoretic methods. *The Journal of Wildlife Management*, 66(3), 912–918.
- Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Bartoń, K. (2014). MuMIn, Multi-model inference [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=MuMIn> (R package version 1.10.5).
- Barton, S., Getz, L., & Kubovy, M. (under review). Systematic variation in rhythm production by tempi.
- Bates, D., Maechler, M., & Bolker, B. (2014). lme4: Linear mixed-effects models using Eigen and S4 [Manuel de logiciel]. Disponible sur, <http://CRAN.R-project.org/package=lme4> (R package version 1.1–7).
- Boker, S.M., & Kubovy, M. (1998). The perception of segmentation in sequences: Local information provides the building blocks for global structure. D.A. Rosenbaum, & C.E. Collyer (Eds.) *Timing of Behavior: Neural, Computational, and Psychological Perspectives*, pp. 109–123. Cambridge: MIT Press.
- Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345–370.
- Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge: MIT Press.
- Burnham, K., Anderson, D., & Huyvaert, K. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65(1), 23–35.
- Claeskens, G., & Hjort, N.L. (2008). *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.
- Cooper, G., & Meyer, L.B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.
- Demany, L., McKenzie, B.A., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature*, 266, 718–719.
- Drake, C. (1998). Psychological processes involved in the temporal organization of complex auditory sequences: Universal and acquired processes. *Music Perception*, 16, 11–26.
- Drake, C., Jones, M.R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, 77(3), 251–288.
- Essens, P.J. (1986). Hierarchical organization of temporal patterns. *Perception & Psychophysics*, 40, 69–73.
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2010). irr: Various coefficients of interrater reliability and agreement [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=irr> (R package version 0.83).
- Garner, W.R. (1974). *The Processing of Information and Structure*. Potomac: Erlbaum.
- Gepshtein, S., & Kubovy, M. (2000). The emergence of visual objects in space-time. *Proceedings of the National Academy of Sciences of the United States of America*, 97(14), 8186–8191.
- Ginestet, C.E. (2009). Model selection and model averaging [review]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(4), 937–937.
- Grueber, C.E., Nakagawa, S., Laws, R.J., & Jamieson, I.G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology*, 24(4), 699–711.
- Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge: MIT Press.
- Handel, S. (1992). The differentiation of rhythmic structure. *Perception & Psychophysics*, 52(5), 497–507.
- Handel, S. (1993). The effect of tempo and tone duration on rhythm discrimination. *Perception & Psychophysics*, 54(3), 370–382.
- Handel, S. (1998). The interplay between metric and figural rhythmic organization. *Journal of Experimental Psychology: Human Perception & Performance*, 24, 1546–1561.
- Hébert, S., & Cuddy, L.L. (2002). Detection of metric structure in auditory figural patterns. *Perception & Psychophysics*, 64(6), 909–818.
- Jeffreys, H. (1961). *Theory of probability*. Oxford: Oxford University Press.
- Johnson, P.C. (2014). Extension of Nakagawa & Schielzeth's R^2 GLMM to random slopes models. *Methods in Ecology and Evolution*, 5(9), 944–946.
- Jones, M.R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, 83, 323–335.
- Kubovy, M., Holcombe, A.O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, 35(1), 71–98.
- Kubovy, M., & van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*, 115, 131–154.
- Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative gestalt theory. *Psychological Science*, 6(4), 225–234.
- Kubovy, M., & Yu, M. (2012). Multistability, cross-modal binding and the additivity of conjoined grouping principles.

- Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591), 954–964. <http://rspb.royalsocietypublishing.org/content/367/1591/954.abstract>
- Large, E.W., & Jones, M.R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106, 119–159.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge: MIT Press.
- MacGregor, J.N. (1985). A measure of temporal patterns. *Perception & Psychophysics*, 38, 97–100.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142.
- Peterson, M.A., & Gibson, B.S. (1994). Must figure-ground organization precede object recognition? An assumption in peril. *Psychological Science*, 5, 253–259.
- Peterson, M.A., & Lampignano, D.W. (2003). Implicit memory for novel figure-ground displays includes a history of cross-border competition. *Journal of Experimental Psychology: Human Perception and Performance*, 29(4), 808–822.
- Povel, D.J., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics*, 30, 565–572.
- Preusser, D., Garner, W.R., & Gottwald, R.L. (1970). Perceptual organization of two-element temporal patterns as a function of their component one-element patterns. *The American Journal of Psychology*, 83(2), 151–170. <http://www.jstor.org/stable/1421321>
- R Development Core Team (2013). R: A language and environment for statistical computing [Computer software manual]. Retrieved from <http://www.R-project.org/> (ISBN 3-900051-07-0).
- Richards, S., Whittingham, M., & Stephens, P. (2011). Model selection and model averaging in behavioural ecology: The utility of the IT-AIC framework. *Behavioral Ecology and Sociobiology*, 65(1), 77–89.
- Royer, F.L., & Garner, W.R. (1966). Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception & Psychophysics*, 1, 41–47.
- Royer, F.L., & Garner, W.R. (1970). Perceptual organization of nine-element auditory temporal patterns. *Perception & Psychophysics*, 7(2), 115–120.
- Ruskey, F. (2011). Information on Necklaces, Lyndon Words, de Bruijn Sequences. In *The (Combinatorial) Object Server* (May 23, 2011 ed.). Retrieved from <http://www.theory.csc.uvic.ca/~cos/info/neck/NecklaceInfo.html>.
- Strother, L., & Kubovy, M. (2012). Structural salience and the non-accidentality of a gestalt. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 827.
- Symonds, M., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, 65(1), 13–21.
- van der Vaart, A. (2009). Information relevant in rhythms. Distinguished major program thesis, University of Virginia, Charlottesville.
- Woodrow, H. (1911). The role of pitch in rhythm. *Psychological Review*, 18(1), 54–77.