

In Rosenbaum, D. A. and Collyer, C. E. (1998) *Timing of Behavior: Neural, Computational, and Psychological Perspectives* Cambridge, MA: MIT Press. pp 109–123

The Perception of Segmentation in Sequences: Local Information Provides the Building Blocks for Global Structure.

Steven M. Boker

Department of Psychology
The University of Notre Dame
Notre Dame, Indiana 46556

Michael Kubovy

Department of Psychology
The University of Virginia
Charlottesville, Virginia 22903

Abstract

We propose that sequences of auditory events may be thought of as being partitioned according to boundaries constructed at times of maximum surprise. A strong version of this hypothesis states that this same mechanism for segmentation of sequences should hold in different modalities and at widely varying time scales, from events measured in years to events measured in milliseconds. One way of quantifying surprise is via information theoretic predictions. The results of an experiment are presented which test the plausibility of the local information segmentation hypothesis using simple repeating auditory rhythmic sequences. Local estimators of information content within an auditory sequence are used to construct predictors of perceived segmentation. These predictors are fit to results of the experiment by using a structural equation model and are compared to Garner's Run Principle and Gap Principle (Garner, 1974). The information theoretic model is found to be a significantly better predictor of the experimental results than the Run-Gap model.

Introduction

To understand the perception and representation of time, we must understand the organization of temporal events. The simplest event is an impulse, an instantaneous burst

of activity preceded and followed by intervals of no activity. A train of impulses manifests the purest organization in time, an organization entirely determined by the relationship of the intervals between the impulses; its organization is purely that of relative durations. For this reason, psychologists interested in the perception and representation of time have turned to sequences of impulses as the most elemental temporal stimuli. The work presented here uses simple repeating sequences of auditory impulses to examine the way that people segment a continuous stream of auditory input into the perception of an auditory object: a rhythm.

We propose that temporal distance in auditory perception can be quantified using a metric derived from information theory (Shannon & Weaver, 1949). When two events share a great deal of information, then after perceiving the first event we are not particularly surprised by the second event. When two events share little information between them, we are surprised by the second event and the first event and second event are likely to be perceived as a segmentation boundary in time. The experiments presented here use the perception of segmentation in simple auditory rhythms to test this proposition.

Perception of Isochronous Sequences

Garner and his colleagues studied subjects who were presented with a variety of visual and auditory cyclic temporal patterns and asked them to indicate which element started the pattern (Royer & Garner, 1966; Garner & Gottwald, 1968; Garner, 1974). Royer and Garner (1966) began by presenting isochronous auditory sequences selected from the set of all the patterns of length eight, and excluding those that were reducible to shorter cycles (such as ABAAABAA or ABABABAB) or could be derived from each other merely by changing a pattern's starting point. In the first experiment all the intervals were filled with an A or a B tone. This turned out to be a more complicated paradigm than necessary, and once Garner and Gottwald (1968) began to use a filled/not-filled interval paradigm, they discovered subjects reported having perceived a segmentation that had a simple structure.

Garner and his colleagues found that the subjects' response latency was correlated with the variance of their choices, and so they proposed that each pattern had a difficulty level that was correlated with its organizational ambiguity. They proposed that two principles govern the organization of single-pitch, uniform intensity cyclic sequences: the *run principle* and the *gap principle*. According to the run principle, a cyclic sequence will appear to begin at the beginning of the longest run of filled intervals. According to the gap principle, a pattern will appear to end at the completion of the longest gap of unfilled intervals. The strength of the prediction made by the run or gap principle is proportional to the length of the longest run or gap.

In some patterns the run and the gap principles coincide, predicting the same per-

We would like to thank John Nesselroade for his friendship and thoughtful advice during the development of this project, Jay Friedenberg for helping to carry the experiments to a successful conclusion, and Jack McArdle for helping to achieve the stimulus-response simplicity of the structural models. We would also like to acknowledge Charles Collyer and an anonymous reviewer whose careful and helpful comments substantially improved this chapter. This research was supported by NIMH grant number 5-R01 MH47317 to the University of Virginia, Michael Kubovy, P.I. and also supported by the Institute for Developmental and Health Research Methodology at the University of Virginia.

ceived starting point to a pattern, and in others they are in conflict, predicting different starting points. When they are in conflict, the stronger of the two principles (as defined by the relative length of the longest run and the longest gap) dominates the other (Garner & Gottwald, 1968).

We performed an experiment that involved the measurement of segmentation in cyclic rhythmic stimuli. We had subjects listen to cyclic rhythms and we asked them to tap a key each time they perceived the pattern to begin. We measured the velocity and timing of these responses and used them as dependent variables to test two models: (1) a model based on the run and gap principles; (2) a model which made predictions of segmentation based on information theory.

Information Theory and Boundary Segmentation

Shannon (1949) formulated a measure of the information contained in a communication, a measure that can be thought of as an extension of Boltzmann's entropy to mixtures of molecules. If x_i is the i th letter in an N letter alphabet and x_i is transmitted with probability $p(x_i)$, then on the average, the information transmitted by one letter is

$$I = - \sum_{i=1}^N p(x_i) \log_2 p(x_i) ,$$

where the Shannon entropy, I , is information measured in bits.

Garner and Hake (1951), Attneave (1954) and others noted the similarity between the formulation of Shannon entropy and Fechner's psychophysical function

$$R = a_i \log S ,$$

where S is magnitude of a stimulus, R is the magnitude of a response and a_i is a constant for individual i . These researchers measured the amount of information contained in stimuli by measuring the responses of subjects perceiving the stimuli. Barlow (1961) proposed that information was extracted from the environment by a process of "reduction of redundancy". Redundancy can be thought of as the amount of information shared between features of a stimulus — as two features share more information, the presence one of the features becomes more redundant. Barlow suggested that as information from the environment is transmitted from a perceptual organ through the nervous system, at each stage redundant information is extracted and discarded. Thus the perceptual signal is reduced to the minimum required to efficiently transmit the information from the environment to the brain.

Now suppose that the events comprising a stimulus consist of cyclical patterns of auditory pulses. If a perceptual system can find a way to combine these events so as to minimize the amount of information, then the system has succeeded in reducing the redundancy in the auditory signal. The resultant perception will be that of a rhythm. By measuring the information shared between possible combinations of the auditory pulses, we can make probabilistic predictions about how these elements are likely to be combined in the perception of a rhythmic stream.

Redundancy can be stated in terms of a ratio of entropies (Barlow, 1961), and has been explored by Redlich (Redlich, 1993) as an active mechanism in visual perception. Redlich proposed a measure of redundancy that we have adapted to rhythmic stimuli.

Consider a measure M to be a cyclic pattern of length v consisting of the ordered set $M = \{m_1, m_2, m_3, \dots, m_v\}$. Each element m_i in the measure can take on a value of 1 if that element is sounded or 0 if that element is not sounded. We will call these elements *beats*, and beats with a value of 1 we will call *notes*. Thus we have an “alphabet” consisting of the set of notes $N = \{n_0, n_1\}$ where $n_0 = 0$ and $n_1 = 1$. Given a random element drawn from M we can calculate $p(n)$, the probability that a note n will occur, as the number of total occurrences of the note n within the measure M divided by the length of the measure v .

We define a feature x of size s within a measure M of length v to be a contiguous ordered subset of the cyclic ordered set M in which the element m_v is considered to be followed by the element m_1 . The set $X^{(s)}$ will be defined to be all of the possible features of size s . Since in our case each feature x is composed of s notes from the set N where the elements of N are 0 and 1, then the set $X^{(s)}$ must map to the binary numbers of length s bits. For instance suppose that $s = 2$. Then the set of features of size 2 would consist of $X^{(2)} = \{\{0, 0\}, \{0, 1\}, \{1, 0\}, \{1, 1\}\}$.

We can now define the probability $p(x)$ that a feature x of size s occurs at a random position i in the measure M to be

$$p(x) = t/T, \quad (1)$$

where t denotes the total occurrences of the feature x within the measure M starting at position i and T denotes the total possible unique features of size s in the cyclic ordered set M starting at a position i . The total possible unique features of size s within a cyclic ordered set M of length v is calculated as $T = kv/s$ where k is the smallest integer that will result in T being an integer. For example, if the length of the measure M is $v = 4$ and the size of the feature x is $s_x = 2$, then there are only two possible unique features of size 2 in the ordered set M starting at some arbitrary point i in the set. However for the same measure M of length $v = 4$, if the size of the feature x is $s_x = 3$, then there are four possible unique features of size 3 starting at an arbitrary point i in the set (see Figure 1).

We can now calculate R_c , the total redundancy in the cyclic pattern due to correlations between the notes as

$$R_c = 1 - \frac{H(X)/\bar{s}}{H(N)} \quad (2)$$

where $H(N)$ is the entropy in the original notes,

$$H(N) = \sum_{n \in N} p(n) \log_2 p(n), \quad (3)$$

and $H(X)/\bar{s}$ is the entropy per beat of a repeated rhythmic pattern, where

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x), \quad (4)$$

and \bar{s} is the average of the lengths of the derived features s_x

$$\bar{s} = \sum_{x \in X} p(x) s_x. \quad (5)$$

To understand segmentation, we must calculate *local* information content rather than global redundancy (which does not distinguish between different starting points for the

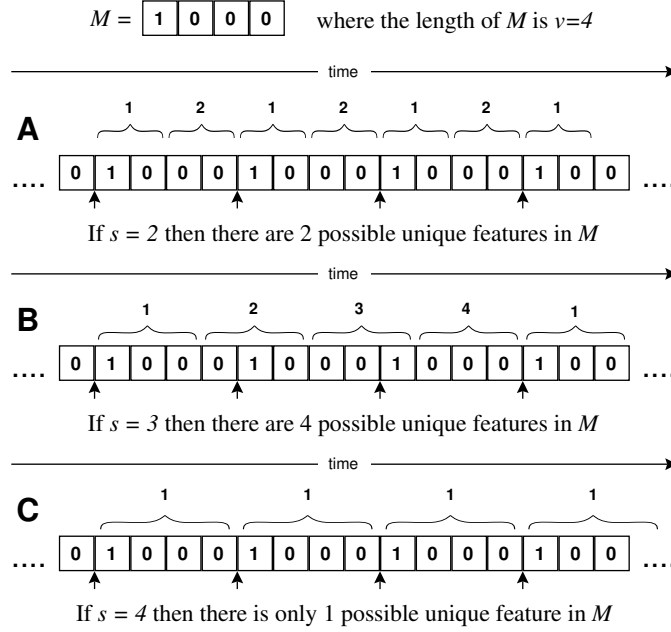


Figure 1. Total unique features of size s in a cyclic measure M of length $v = 4$ calculated by finding the smallest number of measures for which the feature size is an integral divisor. (A) $s = 2$, (B) $s = 3$, (C) $s = 4$.

pattern). Suppose we have a measure M of length v and a set of features X of size s as defined above. We now wish to calculate the local information at position i within the measure M given features of size s . Let r be the number of successive repetitions of a feature x of size s immediately prior to position i as shown in Figure 2. We calculate the local information at position i if feature x does reoccur immediately following position i as

$$\begin{aligned}
 L_s(x) &= \frac{H(x)}{H(!x)} \\
 &= \frac{-r/(r+1) \log_2(r/(r+1))}{-1/(r+1) \log_2(1/(r+1))} \\
 &= \frac{r \log_2(r/(r+1))}{\log_2(1/(r+1))}, \tag{6}
 \end{aligned}$$

where $!x$ stands for all of the features excluding x . On the other hand, if x does not occur immediately following position i then the local information is calculated as

$$\begin{aligned}
 L_s(x) &= \frac{H(!x)}{H(x)} \\
 &= \frac{\log_2(1/(r+1))}{r \log_2(r/(r+1))}. \tag{7}
 \end{aligned}$$

In this way, we can calculate a value for the local “surprise” which is generated immediately following position i within the measure M .

a synthesizer keyboard in synchrony with the perceived starting point. We asked them to continue to strike the key at the perceived beginning of each cycle (not necessarily occurring at the same point in the cycle) until they had settled on a stable starting point. They then pressed a mouse button to end the trial.

Each stimulus was a cyclic repetition of a fixed number of *beats*: equal intervals of time which could either be silent or begin with a percussive sound. Thus a stimulus could be represented by a binary number where each binary digit represents a beat: a zero representing silence and a one representing sound. The set of stimuli for Experiment 1 consisted of all of the unique rhythmic patterns of length 8 or less, 115 patterns in all. The set of rhythmic stimuli in Experiment 2 consisted of a random sample of 115 stimuli drawn half from the unique patterns of length 8 or less and half from the unique patterns of length 12.

We used two methods to minimize the likelihood that subjects would associate the first beat of the stimulus with the perceived beginning of the cycle: (1) we began each trial at a very fast tempo (10 ms per beat) and within 2 seconds we slowed it to a steady beat length of 250 ms, and (2) the stimulus was started at a random point in the cycle.

The percussive sound we used was a synthesized musical cowbell produced by a wavetable synthesis module (Roland MT-32 MIDI) and delivered to the subjects binaurally via headphones (Sennheiser HD-414-SL). The subjects responded by striking a key on a Kawai K5 Digital Synthesizer keyboard. We measured three variables for each keypress: (1) the time of response in milliseconds relative to the beginning of the presentation of the stimulus; (2) the time of response in milliseconds relative to the beginning of current repetition of the pattern as represented internally by the computer software; and (3) the velocity of the response as an integer between 1 and 127.

Models

Run-Gap Predictions

We used a latent variable structural equation model to test the goodness of fit of Garner’s “run-gap” heuristic predictions to the data gathered from the two experiments. Figure 3 shows a path model of Garner’s run-gap heuristics. The predictor variables are *Run*, the length of subsequent run, and *Gap*, the length of the previous gap. The latent variable is *S*, the perceived structure of the rhythmic pattern. The measured outcome variables are *RB*, the response within the beat; *A*, the accuracy of the response; and *V*, the velocity of the response.

We constructed the *Run* variable as follows: if the current stimulus beat value is 1 and the previous stimulus beat value is 0, then $Run = 1 +$ (the number of stimulus beats following the current stimulus beat before another stimulus beat with a value of 0 is encountered). We constructed the *Gap* variable similarly: if the current stimulus beat value is 1 and the previous stimulus beat value is 0, then $Gap = 1 +$ (the number of stimulus beats preceding the current stimulus beat before a stimulus beat with a value of 1 is encountered).

We coded the outcome variables as follows. If a keypress occurred within ± 0.5 beat of the onset of the current stimulus beat, then we coded *RB* as 1, otherwise as 0. If *RB* was 1, we coded the MIDI velocity of the keypress (range 1–127) as *V*. If *RB* was 1, we coded

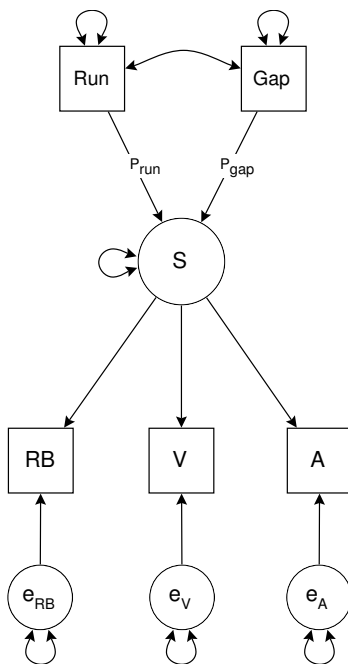


Figure 3. Path diagram showing a structural equation model of Garner's basic run-gap theory.

A as

$$A = 1 - \frac{2|t_k - t_b|}{b} \quad (8)$$

where t_k is the elapsed time to the keypress, t_b is the elapsed time to the onset of the stimulus beat, and b is the duration of the beat. This means that the accuracy, A , of the keypress was 1 if the keypress began simultaneously with the onset of the stimulus beat, and the accuracy was 0 if the keypress occurred halfway between two stimulus beats.

Local Information Predictions

We constructed a similar latent variable model to test the fit of the local information content predictions to the data from Experiments 1 and 2 (see Figure 4). We coded the outcome variables in the same manner as the Run-Gap Model. We calculated the predictor variables in terms of redundancy for features of a particular size.

For each beat x in a rhythmic sequence, three local measures of information content were calculated according to Equations 6 and 7: (1) the local information for features of size 1, $L_1(x)$; (2) the local information for features of size 2, $L_2(x)$; and (3) the local information for features of size 3, $L_3(x)$. We used these three variables, each of which has a fixed value for each beat of each rhythmic sequence, to predict where the segmentation would occur.

Results

We first describe the results of the experiment using histograms to show distribution patterns, and then we summarize the results of model fitting.

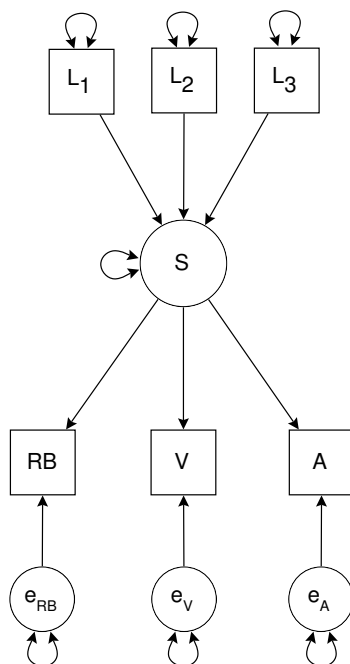


Figure 4. Path diagram of entropy prediction model where the predictors are: L_1 , entropy of features of length 1; L_2 , entropy of features of length 2; and L_3 , entropy of features of length 3.

Response Distributions for Selected Stimulus Patterns.

In this section we present histograms of the response distributions for the stimuli that are pertinent to the theory and models under discussion.

In Figure 5 we show the distributions of responses to two stimuli. For instance, in Figure 5–A we graph the distribution of responses to one stimulus: an eight beat pattern in which four beats are sounded and four are silent. In the label for each histogram we give a binary representation of the stimulus pattern. Elapsed time increases from left to right and intervals between ticks on the abscissa represent 250 ms. We show the starting point of the rhythmic stimulus predicted by the run and gap principles on the far left side of the histogram, and this point is also the first binary digit in the label.

Note that although we numbered the abscissa of Figure 5–A from 0 to 8, there are only 8 beats in the measure [1 1 1 0 1 0 0 0]. It is the eight gaps between the sequence 0 to 8 that represent the durations of the eight beats within this measure. Thus responses that occurred just after the predicted starting point appear at the left edge of the histogram, and the responses which occurred at the end of the duration of the last beat appear at the right edge of the histogram. Since the stimulus is cyclic, responses that occur at the end of the last beat also occur just before the first beat. Therefore, Figure 5–A can be read as a distribution of responses with a mode just before the predicted starting point of the measure.

Figure 5 highlights the inconsistency of the run and gap principles. The responses in Figure 5–A to pattern [1 1 1 0 1 0 0 0] are predicted by the run–gap principles, whereas the

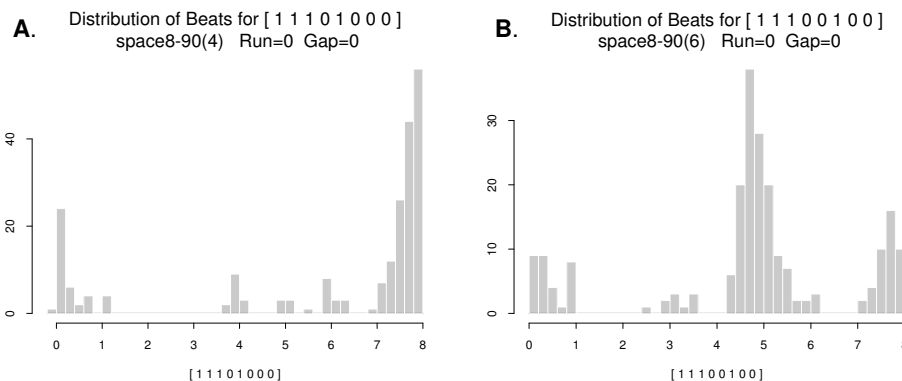


Figure 5. Response distributions for two stimulus patterns. (A) Histogram showing response distributions for one stimulus pattern which elicited responses predicted by the run-gap principle. (B) Histogram showing response distributions for a nearly identical stimulus pattern which elicited responses not predicted by the run-gap principles.

pattern in Figure 5–B, [1 1 1 0 0 1 0 0], is almost identical and yet the responses are not predicted by the run and gap principles. Any theory of organization of rhythmic perception must take into account this type of shift. A number of these inconsistencies appear in our data.

Model Fitting Results

We fit the two models to the data from Experiment 1 and Experiment 2 using the structural equation modeling procedure in SAS (PROC CALIS) and present the results of fitting the Run-Gap Model and the Local Information Model in Table 1. The estimated parameters and χ^2 statistics are presented side by side for the two models.

Table 1: Comparison of prediction model parameters and χ^2 for models fit to the data from Experiment 1 and Experiment 2.

Experiment 1			
	Null	Run-Gap	Entropy
χ^2	923649	2623	79
DF	10	4	6
N	174781	174781	174781
Experiment 2			
	Null	Run-Gap	Entropy
χ^2	541596	388	59
DF	10	4	6
N	104512	104512	104512

The Run-Gap model has a χ^2 fit statistic of 2623 with 4 degrees of freedom (DF). Although this might seem large, the effective sample size is 174,781 separate stimulus response

pairs which contribute to a null model χ^2 of 923,649 with 10 DF. The Run–Gap model fits much better than the null model. However, the Local Information (Entropy) Model has a χ^2 of only 79 with 6 DF. Although these two models are not nested and so cannot be compared precisely in terms of χ^2 goodness of fit, the difference in χ^2 is so great that it overwhelms the possible loss of accuracy due to the non-nested nature of the comparison.

Table 1 also presents the results of the analysis of Experiment 2. The Run–Gap model has a χ^2 fit statistic of 388 with 4 DF compared to a null model χ^2 of 541,596 with 10 DF. The Local Information (Entropy) Model has a χ^2 of 59 with 6 DF. As in Experiment 1, the two models are not nested, but the difference in χ^2 between the two models is so large that we do not hesitate to prefer the Local Information Model over the Run–Gap model.

Discussion

A local information structural equation model (Entropy Model) based on ideas advanced by Barlow (1961) fits our data much better than does the Run–Gap model. This Entropy Model suggests that perceptual segmentation is based on principles of “minimum redundancy”, or alternately put, “maximum surprise”. By segmenting an incoming signal with respect to maxima of local information, redundancy can be removed from the perceived signal.

Other systems concerned with the efficient transmission and storage of information have been shown to obey the laws of information theory and their physical cousins, the laws of thermodynamics. These laws provide a way of measuring relative organization and disorganization. Entropy not only provides a metric for the process by which the universe becomes more disorganized over time, it also provides a metric for the process by which the brain organizes input over time.

Perceptual systems other than auditory are likely to obey these laws of minimization of redundancy while they discover structure in the environment. There is no reason why higher cognitive areas of the brain would be immune to these laws for the communication of information and organization of structure. Thus one might speculate that abstract constructs such as the syntactic and grammatical structure of language should also behave according to general laws of organization.

The importance of an entropic Weltanschauung should not be underestimated. The systematic observation of human behavior inevitably involves measurement of the perception and expression of the organization of structure. Re–expressing psychometric measurement of perceptual tasks in terms of an information theoretic metric may provide a foundation upon which to build structural models of cognition.

References

- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61, 183–193.
- Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.
- Boker, S. M. (1994). *The perception of structure in simple auditory rhythmic patterns*. Unpublished master’s thesis, University of Virginia.

- Boker, S. M., & Kubovy, M. (1998). Predicting the grouping of rhythmic sequences using local estimators of information content. In D. Rosenthal & H. G. Okuno (Eds.), *Computational auditory scene analysis* (pp. 309–320). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.
- Garner, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Garner, W. R., & Gottwald, R. L. (1968). The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology*, *20*, 97–109.
- Garner, W. R., & Hake, H. W. (1951). The amount of information in absolute judgments. *Psychological Review*, *58*, 446–459.
- Redlich, N. A. (1993). Redundancy reduction as a strategy for unsupervised learning. *Neural Computation*, *5*, 289–304.
- Royer, F. L., & Garner, W. (1966). Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception and Psychophysics*, *1*(1), 41–47.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: The University of Illinois Press.