Journal of Experimental Psychology:
Human Perception and Performance
1993, Vol. 19, No. 3, 641–660

0096-1523/93/$3.00

Auditory Perception of Fractal Contours

Mark A. Schmuckler and David L. Gilden

A series of experiments examined auditory contour formation, investigating listeners' sensitivities to a family of random fractals known as fractional Brownian noises. Experiments 1A and 1B looked at identification of contours when 3 different noises were portrayed using variations in the pitch, duration, or loudness of successive notes of a sequence. Listeners could categorize pitch and loudness encodings, but not duration mappings. Experiment 2 looked at the effect of simultaneous presentation of pitch and loudness information, finding that these dimensions combined additively to increase identification of the noise distributions. Experiment 3 looked at discrimination of pitch contours as a function of changing fractal dimension. Discrimination curves approximated an inverted U shape, a finding that is not understandable in terms of sensitivity to differences in fractal dimension per se, nor in terms of "tuned" perceptual sensitivity to statistical regularities of the environment.

Our environment consists of a wide variety of sounds. One of the critical tasks undertaken by the auditory system is to organize this complex array into a series of unified, discrete objects; this process of auditory organization can be thought of as *auditory scene analysis* (Bregman, 1990). One of the crucial functions of the process of auditory scene analysis is to decide what aspects of the auditory environment go with what—in many ways a problem of auditory localization. A critical component of an auditory signal in this process of scene analysis involves the formation and recognition of a particular signal's contour. This article is concerned with people's ability to form auditory contours and to discriminate among contours under a variety of different situations.

The importance of contour information has been widely recognized by researchers interested in two of the most complex exemplars of auditory signals: speech and music. Within speech, contour is a major component of the "prosody" of an utterance (Kuhl, 1987) and can carry important information, such as the distinction between a declarative and interrogative sentence. Contour is also a fundamental attribute of other complex auditory sequences, specifically musical passages. For example, contour appears to be a particularly salient aspect for short-term memory of a passage (DeWitt & Crowder, 1986; Dowling, 1978; Dowling & Bartlett, 1981; Dowling & Fujitani, 1981). Similarly, there is evidence that familiar melodies can be recognized solely on the basis of their contour information (Dowling & Fujitani, 1971, Experiment 2; Dowling & Hollombe, 1977; White, 1960).

In all of this work, our understanding of contour formation is limited to the role of contour during recognition of complex auditory objects. Perceptual sensitivity to the relational information underlying contour is therefore bound to the role contour plays within these domains. As such, little is known about contour formation generally; under what circumstances, for example, can our perceptual systems recover the global relational information that underlies a contour?

To explore basic questions of auditory contour formation, we first need to define a set of auditory sources in which contour information is preeminent. As a start, we must first decide what it means for a contour to exist in perception. The notion of contour is subtle, involving informal conceptions from gestalt theory that have not been made precise. Consider an array of dots laid out horizontally, with a constant spacing between them. In the informal language of gestalt, the perception of a contour in this display, a straight, horizontal line, is an emergent feature of the arrangement of the dots. The line does not exist distally but arises in perception through organization. This arrangement can be considered analogous to an auditory sequence, in which the dots (and their placement on the page) represent a series of note events, each having a specific pitch, loudness, duration, timbre, and so forth.

One characteristic of this arrangement is that there are a multiplicity of scales arising from the dot size and the interdot separations. The distal support for collinearity arises from chaining features in the three-point correlation function (see Gilden, Bertenthal, & Othman, 1990, for a discussion of n-point image statistics). Perceptually, seeing the line involves perceiving the chain structure over the range of scales extant in the dot distribution. Now, although the notion of scale may be made precise, the sense in which scales are integrated remains informal. In this article we will attempt to develop a more rigorous notion of scale integration by constructing contours that are scale-free and hence featureless. A set of sources that are featureless in the sense dis-

Mark A. Schmuckler, Division of Life Sciences, University of Toronto, Scarborough, Ontario, Canada; David L. Gilden, Department of Psychology, Vanderbilt University.

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Mark A. Schmuckler and National Institute for Child and Human Development Grant HD07036 to David L. Gilden.

We thank Stephen Handel and two anonymous reviewers for their comments regarding this work.

Correspondence concerning this article should be addressed to Mark A. Schmuckler, Division of Life Sciences, University of Toronto, Scarborough, Ontario, Canada M1C 1A4. Electronic mail may be sent to marksch@lake.scar.utoronto.ca.

cussed here is a family of random fractals known as fractional Brownian noises, or *scaling noises* (Mandelbrot, 1977).

Fractional Brownian noises contain two properties that make them of interest for our research. Specifically, these noises have random phase spectra and are self-affine. The first property differentiates fractional Brownian noises from other types of nonrandom contour. The property of self-affinity means essentially that an arbitrary magnification of any portion of the noise reproduces its global statistical structure.

An example may clarify what it means for a noise to be self-affine. If a tape recording of a self-affine noise is played, say, at a faster speed than at which it was recorded, an appropriate change in the overall volume would nullify this difference. Noises that are not self-affine will sound altered in pitch no matter how one adjusts the volume control. Similarly, a photograph of a scaling contour would not reveal the distance or magnification at which the photograph was taken. A change in distance or magnification can be offset by cropping the photo.

The importance of self-affinity is ultimately related to what it means for a contour to be perceived. The perception of a contour must involve, at least, the integration of the multiple scales that define a stimulus. Although we are unsure as to what integration means, we do know what a scale is, and self-affine stimuli are those in which all scales are on an equal footing. In other words, in a self-affine noise, every scale carries the exact same signal information under magnification.

When a signal contains a feature, this feature must exist on some scale or group of scales. A self-affine noise does not have a unique structure that exists on a finite set of scales; therefore it is featureless. The dot pattern discussed earlier is not featureless because the constant interdot separation defines a unique scale. In order for a scale to carry a structure not shared by all scales, the signal must contain some sort of periodic or rhythmic structure. The existence of this structure and the way it is differentiated from self-affine structure is best described in terms of the power spectrum of the signal. Only one power spectrum is consistent with the requirement of self-affinity; the spectrum must be a power law. For definiteness, we will denote by β the exponent in the power law for the power spectrum, $P = f^{-\beta}$.

Scaling noises do not have structural descriptions that refer to any single scale. Instead, these noises can only be described globally in terms of the slope of the power spectrum; no other information exists. The fact that the slope identifies the noise requires that discriminations be made on the basis of how each particular element appears in the composite matrix of the entire signal, that is, how the point-to-point fluctuation is embedded in overall trend.

Figure 1 shows three examples of scaling noises. The simplest example, seen at the top of Figure 1, is white noise. When one hears white noise, such as static on a radio, it sounds like a dull hiss regardless of the intensity or speed at which it is heard. For white noise, $\beta = 0$. A more complicated scaling noise is seen in the bottom of Figure 1 and is called brown noise (Gardner, 1978). Brown noise consists of num-

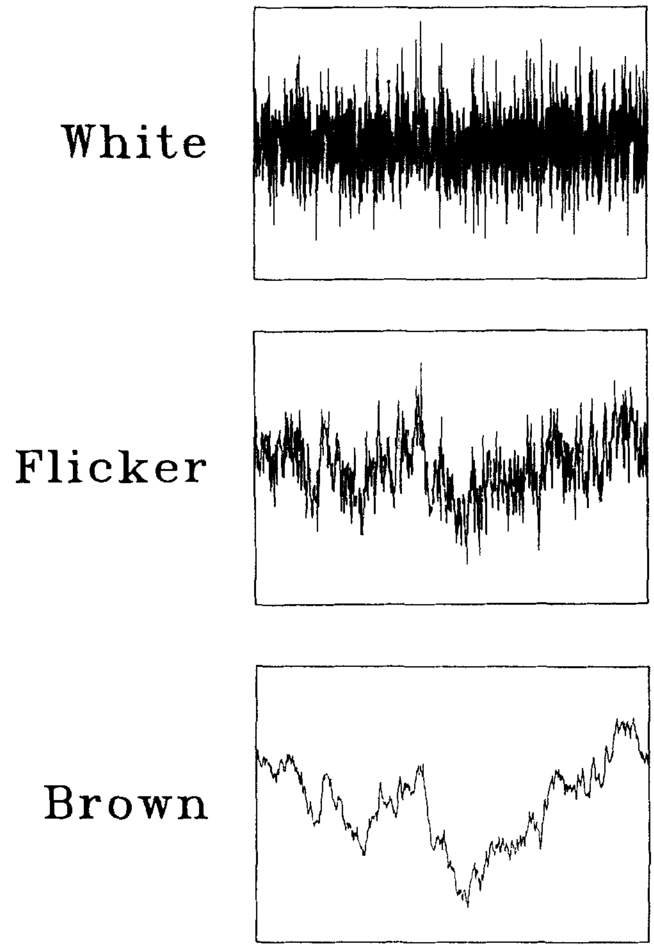

Figure 1. Graphic representations of white, flicker, and brown noise distributions.

bers produced by a random walk, with successive positions highly correlated. For Brown noise, $\beta = 2$. Finally, the middle of Figure 1 presents a third scaling noise; this noise has been called *flicker* or *one-over-f* noise (Mandelbrot, 1977; Voss & Clarke, 1978). Flicker noise is moderately correlated, producing both general trend (as occurs in Brown noise) and point-to-point fluctuation (as in white noise). Flicker noise represents a value halfway between white and brown noise, with $\beta = 1$.

Now we can understand why fractional Brownian noises comprise the ideal stimulus for studying contour formation. They are the only stimulus having scale integration as the sole attribute of their structure. Although other stimuli may lead to the perception of contour, these contours have other pieces of information bound to a particular scale that could serve as the basis of discrimination. On the other hand, fractional Brownian noises can only be discriminated on the basis of how the different scales are integrated, or put differently, by differentiating between slopes of the power spectrum. Discrimination made on the basis of integration is to distinguish on the basis of contour per se. In the experiments that follow, we exploit this property of fractional Brownian noises to examine listeners' sensitivity to contour information in auditory sequences.

Our primary motivation in studying fractional Brownian noises is due to this property of being "pure" contour information. Secondarily, and related to this notion, is the fact that fractional Brownian noises and fractal analyses more generally have been found to describe a variety of visual (Keller, Crownover, & Chen, 1987), auditory (Voss & Clarke, 1978), and motor phenomena (Schmidt, Beek, Treffner, & Turvey, 1991). For our purposes, the most relevant work involves a series of analyses by Voss and Clarke (1978), who performed spectral density analyses on a range of speech and musical sources, including Scott Joplin rags, jazz and blues, rock and roll, the Bach Brandenburg concertos, and a news and talk show. The results of analyses of frequency and power changes indicated that all of these sources could be described in terms of fractional Brownian noises. Even more interestingly, the exponent of the power law of the power spectrum approximating all of the sources was one over the frequency, or flicker noise.

The fact that many natural phenomena are describable in terms of fractal structure suggests that fractal structure might also be relevant psychologically. Can listeners, for example, distinguish between auditory contours on the basis of their fractal dimension? Do fractals have some sort of "privileged" status in perception, given that a great deal of perceptual information can be described in this way? The experiments in this article attempt, as an underlying recurrent theme, to address this question.

Unfortunately, little psychological research has examined the perception of fractal information in its own right. Moreover, what little work has been done has examined primarily visual perception. For example, Cutting and Garvin (1987) had subjects rate the complexity of a set of fractal curves varying in fractional dimension, recursion, and the number of segments in the initiator. These ratings were then predicted from dummy codings of the factors just described. Although recursion depth was the most reliable predictor of these viewers' judgments, for those stimuli with the greatest recursion depth, fractal dimension predicted these ratings quite well. Other research on the perception of fractals has focused on the perceived roughness of fractal curves and textures (Marshak, 1986; Pentland, 1984, 1985; both cited in Cutting & Garvin, 1987). Knill, Field, and Kersten (1990) had observers discriminate among two-dimensional images in which the graininess determined the fractal dimension of these displays. People's ability to discriminate between fractal images varied as a function of fractal dimension, with maximal discrimination found for images with a fractal dimension on the order of 2.5.

Experiment 1A: Contour Formation in Pitch, Loudness, and Duration

Our first step in answering questions concerning contour formation and the perception of fractal structure involves determining whether listeners can perceive differences among fractional Brownian noises. Although one can argue that, theoretically, fractal noises comprise an ideal stimulus for studying contour perception, it is not clear that listeners are sensitive to this structure. A simple test would be to present listeners with auditory sequences differing in their fractal properties (i.e., in the slope of their power spectra) and see whether these noises can be reliably discriminated. To start, we arbitrarily choose noises differing by a constant amount in terms of their fractal dimension (as measured by the slope of the power spectra) and examine listeners' abilities to distinguish between these noises. Although we have no way of determining whether this difference in fractal dimension between noises is perceptually equivalent, it is, at least, physically comparable.

Fractional Brownian noises can be represented as a set of random numbers possessing certain mathematical properties. To produce a perceptible contour, it is necessary to map these random numbers into a perceptually extensive dimension, such as pitch or loudness in audition or height or brightness in vision. Two issues arise out of this transformation. The first involves the choice of perceptual dimension into which this information is encoded, whereas the second involves the consequences of this transformation on the mathematical properties of the sequence. The first of these issues will be discussed here; the second will be taken up in the discussion.

Contour is potentially applicable to any number of dimensions. In audition, contour is most routinely thought of in terms of rises and falls in pitch. However, other auditory dimensions, such as loudness and duration, are candidates for conveying such information, Voss and Clarke (1978), for example, found evidence of fractal structure in both frequency (pitch) and power (loudness) of their sample, and subsequently produced auditory sequences in which both pitch and duration of successive notes were determined in accordance with a fractal noise structure.

Given that contour information can be multiply represented by different auditory dimensions, we can question whether listeners' sensitivity to this information is equivalent across attributes. Does the perception of contour vary depending on a particular mode of representation? To answer this question, we examined listeners' abilities to recognize different power spectra when instantiated in pitch, loudness, and duration changes.

A final concern involves issues related to perceptual learning. Specifically, does discrimination of fractional Brownian noises depend on previous experience with complex auditory passages, which might occur through formal musical training? Throughout the musical cognition literature, one finds that listeners differ in their sensitivity to complex musical structure as a function of the level of musical training they have received.

Method

Subjects

The final sample of subjects consisted of 18 adult listeners recruited from the University of Virginia community.[1] These listeners

[1] Seven additional subjects participated in the experiment, but

either volunteered their services or participated in partial fulfillment of course requirements for an introductory psychology course. Subjects were divided into groups of musically trained and musically untrained listeners. Each group consisted of nine listeners, with the musically trained group having an average of 7.3 years of formal instruction and the untrained listeners having 0.8 years of formal instruction. All of the listeners reported normal hearing, and none had absolute pitch.

Apparatus and Stimulus Materials

All stimuli were generated on-line by a Yamaha DX7 synthesizer controlled by an IBM-PC computer, using a Roland MPU-401 MIDI interface. The listeners, who were seated directly in front of a computer, heard the stimulus passages over Sennheiser HD414SL headphones plugged directly into the synthesizer.

The harmonic structure of the voices used by the DX7 are complex, consisting of six sine-wave generators running simultaneously. The timbre of these stimuli approximated that of a piano. Each tone had a 15-ms rise to peak amplitude, a gradual decay over the length of the tone until its release, and an 80-ms fall to zero amplitude.

Conditions

There were three conditions in this study, corresponding to different mappings of the noise distributions into musical attributes. For each condition, noise distributions (a sequence of random numbers) determined changes in the pitch, loudness, or duration of tones in a passage; these will be referred to as the *pitch*, *loudness*, and *duration* conditions respectively. Three noise distributions were used for generating these sequences; these noise distributions were white, flicker, and brown noise sources (see Figure 1). Generation of the auditory sequences occurred by binning the continuous noise distributions into discrete levels of pitch, loudness, or duration. For each condition, 10 sets of 100 random numbers from each noise distribution produced passages 100 notes in length. To ensure comparable overall ranges between the sequences, the break points between the bins were scaled individually for each sequence of 100 notes. Thus, the maximum number in each sample of 100 corresponded to the highest level of pitch, loudness, or duration, whereas the minimum of the sequence corresponded to the lowest level.

The pitch condition consisted of frequency changes, with the loudness and duration of each tone in the sequence remaining constant. Fourteen different pitch levels were used, corresponding to two octaves of a C major scale, beginning on C4 (middle C, 260 Hz) and ending on B5. The duration of each tone was 200 ms, and the loudness was approximately 84 dBC.

The loudness condition consisted of amplitude changes, with the pitch and duration of each note in the sequence held constant. Twelve different loudness levels were chosen, ranging from approximately 66.5 dBC to 93.5 dBC, in roughly 2.5–3-dB steps.[2] The pitch of each tone in this sequence was C4, and the duration was always 200 ms.

The duration condition consisted of changes in tone length, with the pitch and loudness of the sequence remaining constant. There were 14 durations, ranging from 100 ms to 750 ms, in 50-ms steps. There were no silent intervals between tones. The pitch of each tone was C4, whereas the loudness was approximately 84 dBC.

their data were removed because of equipment error, failure to correctly perform the task, and so forth.

Ten exemplars of each of the three noise distributions were generated for each condition, producing 90 trials in all. These trials were blocked according to condition (pitch, loudness, or duration), with the 30 trials for each condition presented randomly for each listener.

Procedure

At the beginning of the study, listeners saw a drawing of the white, flicker, and brown noise distributions (see Figure 1), and the structure of these sequences was explained. White noise was described as containing large excursions from point to point, with little or no overall trend. In contrast, brown noise consisted of mostly overall trend, with few large, point-to-point excursions. Flicker noise was described as being intermediate between the two, consisting of both point-to-point excursions as well as containing overall trend. The visual models of these three distributions were placed on a poster, which was present throughout the entire experiment.

After having the structure of the noise distributions explained to them, listeners were told that the different noise distributions generated auditory passages varying in pitch, loudness, or duration. The listener's task was to decide which noise distribution (white, flicker, or brown) generated the auditory sequence they heard by entering a 1, 2, or 3 into the computer terminal after hearing each sequence. At the start of each condition, listeners were told which attribute would vary in that block, and then they listened to nine practice trials, three examples of each noise distribution. During these practice trials, listeners received feedback about their categorization accuracy, and when incorrect they were told the correct category. The experimenter was present during the initial practice session of the first condition to answer any questions listeners might have as well as assist them in the use of the computer, and so on.

The order of the three conditions was counterbalanced across subjects. Listeners were told that they could take a break between the conditions but were asked not to stop during a block of trials. The entire experimental session lasted approximately 1 hr.

Results

For each listener, the frequency with which they responded *white*, *flicker*, or *brown* to the white, flicker, and brown noise sequences was calculated for each condition. Table 1 lists the percentage responses for the noise distributions for the conditions averaged across listeners. Inspection of Table 1 reveals that listeners were generally successfully in categorizing the different white, flicker, and brown noise sequences, with this ability dependent on the particular mode of representation. When encoded in pitch and loudness dimensions, listeners categorized these noise distributions successfully. In contrast, when mapped into duration changes, categorization performance diminished greatly.

Subsequent analyses aimed at quantitative verification of these differences between encoding dimensions, along with examining categorization performance as a function of musical training. For each listener, chi-squares were calculated to determine whether their categorizations differed significantly from chance performance. The mean chi-squares for

[2] Given the nature of the synthesizer equipment, all loudnesses are approximate. Loudness measurements were performed with a Bruel and Kjaer Type 2203 Precision Sound Level meter.

Table 1
Mean Percentage Categorization Responses for the White, Flicker, and Brown Noise Distributions for the Different Auditory Coding Dimensions

Auditory dimension and noise stimulus	Listener's response				
	White	Flicker	Brown	χ^2	λ_B
Pitch					
White	*66.1*	31.7	2.2	19.9*	0.41
Flicker	21.1	*58.9*	20.0		
Brown	6.7	27.8	*65.6*		
Duration					
White	*52.2*	27.2	10.6	11.2	0.23
Flicker	26.1	*31.7*	42.2		
Brown	18.3	41.1	*40.6*		
Loudness					
White	*75.6*	22.2	2.2	24.1**	0.51
Flicker	22.2	*61.7*	16.1		
Brown	2.8	25.0	*72.2*		

Note. Correct categorizations are in italics.
* $p < .05$. ** $p < .01$.

both pitch and loudness conditions differed significantly from chance, whereas the mean chi-square for the duration condition did not differ from chance (see Table 1).

A 2 × 3 × 3 analysis of variance (ANOVA) was performed, using these chi-squares as the dependent variable. This analysis used the between-subject factors of training (trained, untrained) and order of conditions and the within-subject factor of coding dimension (pitch, loudness, and duration). This analysis revealed main effects for order, $F(2, 12) = 5.4$, $p < .05$, and coding dimension, $F(2, 24) = 20.2$, $p < .001$. There was no effect for training, $F(1, 12) = 0.08$, *ns*, nor were there significant interactions.[3] Multiple comparisons between all pairs of mean chi-squares for the three coding dimensions, using Bonferroni corrections, found that whereas pitch and loudness conditions did not differ from each other, both differed from the duration condition (both *ps* < .01).

The preceding analyses examined whether categorization of the different encodings differed from chance; they did not test the accuracy of these categorizations. Did listeners correctly associate the white noise sequences with the "white noise category," and so on. To examine this question, an *index of predictive association* was calculated for each subject. The index of predictive association, or λ_B, provides a measure of the degree to which each stimulus is associated with a single response category (Goodman & Kruskal, 1954; see Hays, 1963, p. 608).[4] λ_B ranges between 0 and 1, with 0 signifying no association between a given stimulus and response label and 1 signifying a perfect association between stimulus and response. Although no significance test exists for λ_B itself, once calculated, this number can be used as the dependent variable in subsequent analyses.

Measures of predictive association were compared in a 2 × 3 × 3 ANOVA, with the same factors as discussed earlier. The only significant results produced by this analysis involved a main effect for coding dimension, $F(2, 24) = 24.45$, $p < .001$. Multiple comparisons between the mean λ_B for the three conditions (see Table 1), using Bonferroni corrections, revealed that whereas pitch and loudness conditions did not differ from each other, both differed from the duration condition (both *ps* < .01).

Finally, analyses assessed the degree to which practice effects occurred in categorization performance for the three conditions. For each condition, each listener's responses were recoded either as correct (1) or incorrect (0), and the number of correct responses for each trial, summed across listeners, were calculated. In this analysis, practice effects would be manifest as an increasing function across trials. Regression analyses failed to find a significant linear relationship in these data for the pitch ($r = .13$, *ns*), loudness ($r = .27$, *ns*), or duration conditions ($r = -.17$, *ns*), and inspection of these data did not suggest any nonlinear increasing function.

Discussion

In response to our original questions, these results indicate that listeners can reliably distinguish between white, flicker,

[3] Although we were initially quite taken by the idea that musical training might play an important role in the perception of fractal structure, the results of Experiment 1A provide strong evidence that training is unimportant. With hindsight, the fact that musical training is not a prerequisite for sensitivity to this structure fits well with the literature suggesting that even naive subjects (i.e., infants and children) are sensitive to contour information in both speech (Kuhl, 1987) and music (Trehub, Bull, & Thorpe, 1984). Although it seems that perception of some aspects of musical and linguistic structure might require experience, the type of structural information that is being investigated in these studies is available without the need for previous training. As such, we will ignore any effects related to musical training.

[4] Although the index of predictive association determines the degree to which a given stimulus type is associated with a single category label, there is no guarantee that this labeling is veridical, that is, that the white noise sequences were consistently associated with the "white" category label. However, inspection of the data on an individual and group basis (see Table 1) verified that these sequences were given the appropriate label.

and brown noise distributions. However, not all mappings were perceptually relevant for listeners. When encoded by changes in note duration, listeners failed to successfully categorize the noise distributions. In contrast, when mapped into pitch and loudness changes, listeners accurately categorized these stimuli.

Before discussing why successful categorization of noise distributions was so difficult for duration encodings, it is instructive to rule out some obvious (and uninteresting) explanations for this finding. One possibility is that duration differences, in contrast to pitch and loudness changes, were simply not discriminable to our subjects, resulting in the observed inability to categorize these sequences. Unfortunately, appeals to the psychophysical literature on discrimination accuracy for durations are not particularly helpful, given that results vary as a function of the specific experimental methodology used (Fraisse, 1978) and that none of the standard psychophysical methods is particularly applicable to this case. Nevertheless, on some estimates, the smallest perceptible duration difference for tones between 200 and 4,000 ms is a constant differential fraction of less than 0.05 (Getty, 1975; see Fraisse, 1978), meaning that duration differences of 50 ms (the smallest possible duration difference) should have been discriminable for our listeners.

A different possibility is that differences between the pitch–loudness conditions and the duration condition simply reflect different learning curves for these encodings. According to this explanation, differential performance between conditions occurred because listeners are at different points on the learning curve for the duration stimuli than they are for either the pitch or loudness conditions. However, the amount of learning (as assessed by improvement in categorization across trials within a block) was examined and failed to show any strong evidence of increasing performance over trials for any of the three conditions, making it unlikely that there were any strong learning effects.[5]

We suggest that the best explanation for the failure of the duration encodings to accurately convey contour information involves the nature of the duration contours in relation to listeners' abilities to accurately perceive and represent rhythmic–duration patterns. For all three of our coding dimensions, representing noise distributions occurred by translating random numbers into tone attribute changes, with increasing numbers leading to notes that were longer, louder, or higher. Though no data exist on the nature of listeners' encodings of loudness sequences, this translation scheme does appear to match with theories concerning the internal representation of pitch sequences (Deutsch & Feroe, 1981; Jones, 1981). In contrast, investigations of the representation of rhythmic sequences suggest a hierarchical structuring in which low-level events nest within higher level events. As such, musical rhythms generally consist of simple integer ratios like 2:1 or 3:1 (Lerdahl & Jackendoff, 1983; Povel, 1981). It is possible that listeners have difficulty encoding rhythms not using these simple ratios. Essens (1986), for example, suggested that accurate internal representations of a rhythmic pattern are possible only when this pattern consists of simple ratios. Presenting noise distribution information by adding or subtracting 50-ms time constants (or multiples thereof) does not produce a preponderance of these simple ratios, resulting in sequences that listeners are unable to encode accurately.

Unfortunately, there exist other reasons that make accepting this explanation premature. An obvious candidate for the poor performance with duration sequences has to do with the fact that in both pitch and loudness dimensions differences between steps were logarithmically scaled, whereas they were linearly scaled in the duration condition. The failure to recognize duration encodings might have been due to the underlying continuum rather than an inherent difficulty in forming duration contours. Another difficulty involves the fact that the various auditory dimensions covered different overall ranges. Assuming that a doubling in a particular dimension defines an "octave," then the pitch stimuli had a range of 2 octaves, the loudness stimuli ranged over approximately 2.7 octaves (assuming that a 10-dB increase equals a perceived doubling in loudness), and the duration stimuli spanned approximately 2.9 octaves. Although these stimuli are approximately equal (between two and three octaves), the differences in overall ranges raises lingering doubts. Both of these issues are examined in Experiment 1B.

Most important, however, there is a possible mediating factor underlying listeners' categorizations of these sequences. Specifically, there is an issue involving the extent to which the different noise distributions could have produced auditory sequences with various amounts of auditory stream segregation. *Auditory stream segregation* refers to a general process by which connections are formed between events within an auditory sequence (Bregman, 1990). Auditory stream segregation, or *streaming*, involves the effect produced by a sequence of rapidly alternating high and low tones in which the high and low tones split into distinct perceptual groups or streams (Bregman, 1990). As a gross oversimplification, the greater the distance (in some dimension) between two events, the more likely these events will be to stream. The complement of streaming has been called *fusion* or *temporal coherence* and refers to the situation in which an auditory sequence is heard as a single, unitary event (Van Noorden, 1975).

The application of the principles of auditory stream segregation to the fractal noises of this experiment is straightforward. Using the pitch condition as an example, white noise contains a high number of large pitch intervals between successive notes, resulting in sequences that stream. Brown noise contains a high number of small pitch intervals, producing sequences that rarely stream. Finally, flicker noise will be halfway between the two. It is possible that categorization performance in this experiment was based on listeners' detection of the amount of streaming in these passages. Moreover, auditory streaming could help explain the decrement in performance for duration encodings, in that whereas stream segregation is applicable to both pitch and

[5] Unfortunately, ruling out practice effects in categorization with these distributions is difficult. Given sufficient exposure, recognition of these distributions will increase. For example, Schmuckler has noted that after numerous years of exposure to these noises, his categorization performance has improved.

loudness dimensions, variation in duration does not necessarily produce streaming (except possibly for the white noise sequences); the result would be poorer discrimination for this condition.

Unfortunately, it is difficult to wholly dissociate the effect of auditory stream segregation from the fractal structure of the noise distributions. White noise, by definition, will contain power in the upper frequencies, resulting in a lot of point-to-point fluctuation, or large differences in pitch (or loudness or duration) between successive notes. In contrast, brown noise contains little power in the upper frequencies but a great deal of power in the low frequencies. This results in strong general trend, with little point-to-point fluctuation, producing sequences that do not stream.

It is possible, however, to devise a situation in which the importance of streaming for the perception of fractal noises can be assessed. If the entire range of the sequences were manipulated, such that the largest possible difference between successive events was within the boundary for which these sequences could be heard coherently, then streaming would be effectively reduced or eliminated. This sequence would, however, still approximate the fractal structure of the different noise distributions. This strategy was undertaken in the following experiment.

Experiment 1B: The Effects of Streaming, Range, and Underlying Scale

The main purpose underlying Experiment 1B was to assess the importance of the different factors that were raised in the previous discussion. First, and most trivially, this experiment controls for the earlier difference in the underlying dimensions between duration and pitch–loudness by scaling these dimensions logarithmically. This experiment also controls for the absolute ranges of the different dimensions by equating them.

A far more interesting extension provided by this experiment is in its attempt to assess the role of streaming in perception of fractal sequences. Here, streaming was manipulated by changing the overall range into which the noise structures were mapped. In one case, the noise distributions could be mapped into a large range, in which streaming would be inevitable, at least for white noise. For comparison, the same noise distribution could also be mapped into a more truncated range, one that does not produce streaming. If streaming is crucial for our ability to discriminate between fractal noises, then there will be better discrimination in the first case (in which noises differ in their streaming) than in the second case (in which none of the noises stream).

Finally, this experiment also examined the influence of the number of discrete levels into which the noise distributions were mapped. One assumption of fractal analyses is that the stimuli used are continuous. Binning the continuous noise distributions into discrete values does violence to the fractal nature of these sequences. Strictly speaking, our sequences only approximate fractal structure because the very act of producing auditory sequences with discrete values changes these noises from continuous to discrete. Manipulating the "coarseness" of the coding of the continuous information by varying the number of bins these noises are mapped into allows us to assess the effect that binning has on our perception of this structure. Less coarse codings (more discrete levels) better approximate the continuous fractal structure of the underlying noise distributions and might therefore be more easily discriminated than more coarse codings.

Method

Subjects

The final sample of subjects consisted of 48 listeners drawn from the University of Toronto, Scarborough, community.[6] These listeners either volunteered their services or received extra credit in an introductory psychology class for participating.

Apparatus

All stimuli were generated on-line using the same equipment as in Experiment 1A, with the only difference being that listeners heard the sequences through a Peavey KB-60 Amplifier located in the room with them.

Stimuli

Overall, generation of these sequences was similar to that of Experiment 1A: 10 samples of 100 random numbers were drawn from white, flicker, and brown noise distributions and were binned into discrete levels, with the break points between bins scaled individually for each sequence. The ranges of the different conditions varied as a function of condition, as did the number of bins into which the noise distributions were mapped.

Because the pitch and loudness conditions of Experiment 1A did not differ, and both differed from the duration condition, this experiment used only pitch and duration encodings. To create stimuli in which the total ranges were equivalent and the underlying continuum was the same, pitch and duration dimensions were given comparable ranges in terms of the number of octaves they spanned, with an octave defined as a doubling within that dimension. For example, pitches between 220 and 440 Hz or 440 and 880 Hz each span a single octave; similarly, durations between 100 and 200 ms or 200 and 400 ms also span single octaves. Within each octave, 12 equally spaced logarithmic steps were derived. Although this is a novel encoding of duration changes, in pitch such an arrangement produces the equal-tempered scale.

Conditions

There were two within-subject conditions in this experiment, and three between-subject conditions. The within-subject conditions corresponded to the auditory dimension (either pitch or duration) into which the noise distribution was encoded. These will be referred to as the pitch and duration encoded conditions.

The three between-subject conditions corresponded to the different mappings of noise distributions into auditory sequences. In the first condition, noise distributions were mapped into a one-octave range containing 13 discrete levels (including the octave

[6] Two additional subjects were run, but their data were not included because of a failure to follow the instructions, as well as failure to complete the experiment.

doubling at the top). For the pitch stimuli, the duration of each tone in the sequence was 200 ms.[7] The pitches used ranged from C4 (middle C) to C5 in 12 equally spaced logarithmic steps. Musically, these pitches are separated by an interval of a semitone. For the duration encoded stimuli, the pitch of each tone was C4. The durations used ranged from 200 to 400 ms in 12 equally spaced logarithmic steps. This condition will be referred to as the *1 octave, 13 bin* condition.

In the second condition, noise distributions were mapped into a three-octave range, again containing 13 discrete levels. For the pitch stimuli, the duration of each tone was 200 ms. The pitches used in this sequence ranged from C3 to C6 in 12 equally spaced logarithmic steps. Musically, these pitches are separated by an interval of a minor third. For the duration stimuli, the pitch of each tone was C4. The durations used ranged from 100 to 800 ms in 12 equally spaced logarithmic steps. This condition will be referred to as the *3 octave, 13 bin* condition.

In the third condition, noise distributions were mapped into a three-octave range containing 37 discrete levels. For the pitch stimuli, the duration of each tone was 200 ms. The pitches used in this sequence ranged from C3 to C6 in 36 equally spaced logarithmic steps. Musically, these pitches are separated by an interval of a semitone. For the duration stimuli, the pitch of each tone was C4. The durations used ranged from 100 to 800 ms in 36 equally spaced logarithmic steps. This condition will be referred to as the *3 octave, 37 bin* condition.

Listeners were assigned to one of the three range–bin conditions. Within each condition, listeners heard both pitch and duration encodings, with the order of these encodings counterbalanced across listeners.

Procedure

The procedure of this experiment was similar to that of Experiment 1A. Listeners saw drawings of the white, flicker, and Brown noise distributions, and the structure of these distributions was explained. The task of the listener was again to determine which noise distribution they had heard. As in Experiment 1A, subjects heard nine practice trials (with feedback about their accuracy) and then began the first block of 30 trials (10 of each noise distribution). After completing the first block, listeners again heard nine practice trials and then performed the second block of trials. The entire experimental session lasted approximately 45 min.

Results

The data analysis for Experiment 1B was similar to that for Experiment 1A. For each listener, the frequency with which they responded *white*, *flicker*, or *brown* was calculated for both pitch and duration encodings. Table 2 lists the average percentage response for the noise distributions for pitch and duration encodings as a function of the range and bin mappings. As in Experiment 1A, listeners categorized the different noise distributions successfully, with this ability again dependent on the particular mode of representation. Generally, pitch encodings led to better categorization performance than did duration encodings. Interestingly, there do not appear to be any systematic differences between categorizations as a function of the range and number of bins of the different maps. Subsequent analyses attempt to quantify these intuitive impressions.

Similar analyses compared performance in the different conditions, using both the chi-square and λ_B measures described earlier. Both variables were calculated for all listeners, with the mean values for these measures also shown in Table 2. Two $3 \times 2 \times 2$ ANOVAs were performed, with a between-subject factor corresponding to the number of bins and range–bins of the stimuli (1 octave, 13 bin; 3 octave, 13 bin; 3 octave, 37 bin), a within-subject factor of encoding dimension (pitch vs. duration), and a between-subject factor of order (pitch first vs. duration first). The dependent variables in these ANOVAs were the chi-squares and the λ_B measures. For the chi-squares, the only significant effect was the main effect of encoding dimension, with the pitch condition significantly greater than the duration condition, $F(1, 42) = 13.98$, $p = .001$. Similarly, for the λ_Bs, the only significant difference was a main effect for encoding dimension, with the pitch condition greater than the duration condition, $F(1, 42) = 10.87$, $p = .002$. In neither case was there a significant interaction between encoding dimension and range–bins, although for the chi-square analysis this interaction approached significance, $F(2, 42) = 2.6$, $p = .09$.

Discussion

The results of this experiment help to clarify the questions raised concerning Experiment 1A. The observed difference between the pitch–loudness encodings and duration encodings of Experiment 1A were not likely due to logarithmic versus linear scaling, given that even with logarithmic scaling for both pitch and duration dimensions categorization was still better for pitch encodings. Similarly, equating of the overall ranges of pitch and duration changes (in terms of the number of octaves spanned) also gave rise to a difference between pitch and duration dimension categorizations.

The most interesting results of Experiment 1B stem from the manipulations of range (1 octave vs. 3 octaves) and the coarseness of the coding (13 bins vs. 37 bins). Stated simply, there were no differences in categorization performance as a function of manipulating either of these dimensions. Although one must always remember the dangers of drawing conclusions from null results, these findings are suggestive. One implication involves the importance of streaming in the perception of these sequences. If streaming were a critical mediating factor, then there should have been a difference in performance as a function of whether the range of the sequences facilitated the occurrence of streaming for some noises (i.e., 1 octave vs. 3 octaves). This result did not occur, however. Although it is tempting to suggest that streaming is unimportant in perceiving these structures, we should be cautious in dismissing such a potent organizing principle. As discussed earlier, the very structure of these noise distributions is such that some noises produce streaming more easily

[7] According to Van Noorden's (1975) seminal work on stream segregation, at a rate of 200 ms per tone a pitch range of one octave (13 semitones) should fall within the *temporal coherence boundary*, which is the "boundary between temporal coherence and fission when the observer is trying to hear temporal coherence" (p. 10). Other support for temporal coherence between notes within an octave (with durations of 200 ms) comes from Miller and Heise (1950) and Schouten (1962, cited in Van Noorden, 1975).

Table 2
Mean Percentage Categorization Responses for White, Flicker, and Brown Noise Distributions as a Function of Auditory Dimension and Condition

	Pitch encoding			Duration encoding		
Stimulus	White	Flicker	Brown	White	Flicker	Brown
Condition: 1 Octave, 13 Bins						
White	*60.0*	32.5	7.5	*55.6*	32.5	11.9
Flicker	33.7	*49.4*	16.9	29.1	*55.0*	16.9
Brown	10.6	23.9	*62.5*	14.4	31.9	*53.7*
χ^2	17.4*			14.5		
λ_B	0.38			0.28		
Condition: 3 Octaves, 13 Bins						
White	*56.3*	35.0	8.7	*41.9*	41.2	16.9
Flicker	38.8	*50.0*	12.2	28.8	*46.2*	25.0
Brown	11.2	25.0	*63.8*	16.9	25.6	*57.5*
χ^2	15.9*			13.2		
λ_B	0.32			0.29		
Condition: 3 Octaves, 37 Bins						
White	*52.5*	30.6	16.9	*33.7*	44.4	21.9
Flicker	26.2	*51.3*	22.5	29.4	*36.2*	34.4
Brown	13.7	23.1	*63.1*	21.9	30.6	*47.5*
χ^2	17.0*			7.8		
λ_B	0.33			0.18		

Note. Correct categorizations are in italics.
* $p < .05$.

than others. Simply because eliminating streaming did not dramatically affect categorization does not mean that streaming does not play a role when it is potentially available. Nor does it rule out the possibility that it might have a more subtle influence, one that the current situation could not detect. Moreover, it might be that in our one-octave condition the sequences did stream to some extent. We chose a one-octave range primarily because of convenience (and limitations in our equipment). Though 1 octave is below the temporal coherence boundary (Van Noorden, 1975) for tones of 200 ms, it is above the fission boundary. This means that listeners might have been able to hear some streaming in these sequences, with conscious effort. However, the amount of streaming as well as the strength of the effect would be greatly reduced relative to the three-octave condition. Based on our results, as well as the aforementioned caveats, it appears that streaming, though conceivably playing a role in perception of these structures, is unable to account for listeners' categorizations.

A second result of interest arises from the lack of a difference between the 3 octave, 13 bin condition and the 3 octave, 37 bin condition. These conditions differ in the strength of their approximation of the continuous fractal distribution underlying the different sequences, or what we have called the *coarseness* of the coding. Interestingly, it appears that 13 discrete levels (a relatively coarse coding) can adequately convey the distribution information. One possible extension would be to provide even fewer levels to determine the point at which this information can no longer be recovered. Additionally, the efficacy of the coarseness of the coding of these sequences must, in part, be relative to the separation (in fractal dimension) of the different noise distributions being discriminated. Other extensions might vary the separation between noise distributions, along with the number of discrete levels used to represent these distributions, to investigate possible interactions between these factors. Experiment 3 touches peripherally on this issue.

Taken together, Experiments 1A and 1B provide some interesting insights into the perception of fractal structure, as well as the formation of contour more generally. Experiments 2 and 3 both provide more in-depth study of listener's apprehension of this type of information, although these studies move in divergent directions. Experiment 2 extends our investigations into the perception of contour information when coded into different auditory dimensions. Experiment 3 looks more finely at listeners' perceptions of pitch contours and how appreciation of fractal structure varies as a function of fractal dimension.

Experiment 2: Perceiving Simultaneous Pitch and Loudness Contours

Experiment 2 investigated the effects of simultaneous presentation of pitch and loudness contours in identification of fractal information. The basic question examined is whether the availability of multiple, simultaneous sources of information leads to better recognition of a particular noise distribution. That is, if we are searching for a white noise structure in a sequence that has both pitch and loudness variation, is our recognition of white noise better when both pitch and loudness change according to a white noise distribution, as opposed to if just one of the two dimensions (pitch or loudness) is based on a white noise distribution?

We might anticipate that presenting simultaneous pitch and loudness changes in an auditory sequence should lead to more accurate identification of this information. On what basis might we form this expectation? A very simple model for why recognition of fractal information might increase in this situation is that presenting simultaneous target information in two dimensions gives listeners two independent samples on which to base a judgment. Such a model assumes an additive relationship between the different sources of information. If true, we should be able to predict the identification of sequences changing in both pitch and loudness information simultaneously from the identification of sequences in which there is only pitch information, combined with identification of sequences in which there is only loudness information. We develop such a model below.

For convenience, we will develop this model using signal detection terminology. The task of the listeners is to say *yes* if they detect the presence of a specified noise distribution (the target) in either pitch or loudness dimensions. Let $p(+)$ denote the probability of saying *yes* in situations in which both pitch and loudness dimensions vary simultaneously. Within a given dimension (either pitch or loudness), we denote the conditional probability $p(+:+) = p$ as the probability of saying *yes* if the target was, in fact, present (a hit), and $p(+:-) = q$ denotes the probability of saying *yes* if the target was absent (a false alarm). These two probabilities will be given by conditions in which only one dimension carries relevant information. For completeness, denote by $p(-:+) = 1 - p$ and $p(-:-) = 1 - q$ the probabilities of saying *no* if the target information is present or absent, respectively (misses and correct rejections).

To analyze a situation in which both pitch and loudness levels vary simultaneously, we need to compute the probability of a *yes* response to either or both dimensions. In a model where the different dimensions are independent, the individual probabilities compound multiplicatively. In general, the probability of a *yes* response is

$$p(+) = p(1, +)\,p(2, -) + p(1, -)\,p(2, +) + p(1, +)\,p(2, +),$$

where $p(j, +)$ is the probability of saying *yes* to information in the jth ($j = 1$ or 2) dimension and where $p(j, -)$ is the probability of saying *no* to information in the jth dimension, regardless of the nature of the information in either dimension. This equation plus the preceding definitions allow us to compute the probability of identifying a noise distribution in a sequence in which both pitch and loudness vary simultaneously, simply by keeping track of whether a target is present in the first or second dimension and whether the response is a *yes* or a *no*. For the example below, we assume that the listener has been instructed to listen for the presence of white noise (the target).

Case 1: Two Targets Present

White noise is present in pitch and loudness:

$$p(+) = p(+:+)p(-:+) + p(-:+)p(+:+) + p(+:+)p(+:+),$$

$$p(+) = p(1-p) + (1-p)p + p^2 = 2p - p^2.$$

Case 2: One Target Present

White noise is present only in pitch. Because the target can appear in either dimension with equal probability, we can suppose without loss of generality that the target is present in Dimension 1:

$$p(+) = p(+:+)p(-:-) + p(-:+)p(+:-) + p(+:+)p(+:-),$$

$$p(+) = p(1-q) + (1-p)q + pq = p + q - pq.$$

Case 3: 0 Targets Present

White noise is not present in pitch and loudness:

$$p(+) = p(+:-)p(-:-) + p(-:-)p(+:-) + p(+:-)p(+:-),$$

$$p(+) = q(1-q) + (1-q)q + q^2 = 2q - q^2.$$

It is assumed here that the hit and false alarm rates are the same regardless of which dimension carries the target.

The current experiment tests this probability model by examining identification accuracy when fractal information occurred simultaneously in two tone dimensions. Toward this end, we adapted a methodology typically used in work investigating global versus local precedence (Navon, 1977; Pomerantz, 1983; Pomerantz & Sager, 1975), in which subjects report if a target stimulus (i.e., the letter *H*) is present in a multidimensional stimulus.

Method

Subjects

The final sample of subjects consisted of 16 adult listeners recruited from the University of Virginia community.[8] All listeners were paid $5 for participating. All listeners reported normal hearing, and none had absolute pitch.

Apparatus and Stimulus Materials

Stimuli were generated with the same equipment as in Experiment 1A. All sequences were heard over Sennheiser HD414SL headphones plugged directly into the synthesizer. In this study, samples of 100 random numbers from the white and flicker noise distributions were mapped into 14 levels of pitch and 12 levels of loudness, using the same ranges and coding scheme as in Experiment 1A. All tones were 200 ms in duration and played in the same timbre as in Experiments 1A and 1B.

Conditions

This experiment contained a number of different experimental conditions; these are described below and outlined in Figure 2. The primary manipulation involved the simultaneous presentation of sequences changing in both pitch and loudness. For comparison,

[8] One additional subject was removed from this study because of her failure to follow the experimental instructions.

		Pitch Dimension	
		White Noise	Flicker Noise
Loudness Dimension	White Noise	Redundant: 1. Congruent 2. Incongruent Response: "present"	Non-redundant Response: "present"
	Flicker Noise	Non-redundant Response: "present"	Redundant: 1. Congruent 2. Incongruent Response: "absent"

Unidimensional Conditions:

Pitch Dimension	White Noise Response: "present"	Flicker Noise Response: "absent"

Loudness Dimension	White Noise Response: "present"	Flicker Noise Response: "absent"

Note: The task of the listener is to listen for the presence of White noise in either Pitch or Loudness dimensions.

Figure 2. Schematic representation of the bidimensional and unidimensional conditions of Experiment 2. (Also shown are the appropriate responses for the conditions of Experiment 2 when listening for the white noise distribution.)

sequences were included in which variation in only one dimension occurred, while the second dimension was held constant. There were, therefore, two general classes of sequences: bidimensional stimuli, in which both pitch and loudness varied simultaneously, and unidimensional sequences, in which either pitch or loudness varied, while the other dimension stayed constant. For both bi- and unidimensional sequences, pitches and loudnesses could vary in accordance with either a white noise or a flicker noise distribution.

Bidimensional sequences. Crossing the two dimensions available for change (pitch and loudness) with the two noise distributions each on which could be based (white or flicker noise) produced four different bidimensional sequences of two general types. The first type were "redundant" sequences, in which both pitch and loudness changes were based on the same noise distribution (i.e., both determined by white noise or both by flicker noise), whereas the second type were "nonredundant" sequences, in which the pitch dimension was determined by one noise distribution and the loudness dimension was determined by the second noise distribution (i.e., white noise pitch changes and flicker noise loudness changes, or vice versa; see Figure 2). A final manipulation for the bidimensional sequences occurred within the redundant sequences, which could be either "congruent" or "incongruent." For the congruent redundant sequences, pitch and loudness changes were based not only on the same general noise distribution (i.e., white noise) but also on the same set of 100 random numbers. Thus, changes in pitch and loudness were almost perfectly correlated.[9] For the incongruent redundant sequences, pitch and loudness varied in accordance with the same general noise distribution, but the actual set of random numbers determining these changes differed for these dimensions. In sum, then, there were six bidimensional conditions: two congruent redundant conditions, two incongruent redundant conditions, and two nonredundant conditions (see Figure 2).

Unidimensional sequences. For the unidimensional sequences, the stimuli contained one dimension (either pitch or loudness) that changed in accordance with either white or flicker noise, while the second dimension remained constant. For those sequences in which the loudness changed, the pitch of each note was held constant at C4; when the pitch changed, the loudness level was approximately 84 dBC. These sequences are essentially a replication of the pitch and loudness conditions of Experiment 1A. Both pitch and loudness unidimensional stimuli were crossed with each noise distribution, producing four unidimensional conditions (see Figure 2).

This study contained one final experimental manipulation. Although each sequence was generated on the basis of samples of 100 random numbers, the actual length of the sequences heard by listeners was manipulated. Specifically, five different sequence lengths were used: 15 notes in length, 30 notes, 45 notes, 60 notes, or 75 notes.

In sum, then, there were 10 experimental conditions: six bidimensional conditions and four unidimensional conditions (see Figure 2). Each of these conditions contained five different lengths, producing 50 experimental trials. One set of 50 trials was considered a single experimental block of trials. Four exemplars of each trial type were created, producing four blocks of 50 trials each.

Procedure

Prior to the start of the experiment, listeners saw drawings of the white and flicker noise distributions (see Figure 1), with the structure of the sequences explained as in Experiments 1A and 1B. Listeners were told that these distributions generated auditory sequences varying in pitch, in loudness, or in both pitch and loudness simultaneously. The listener's task was to listen for one of the two noise distributions in each sequence (i.e., to listen for the white noise distribution) and to respond *present* if that noise distribution occurred in either pitch or loudness and *absent* if it was not present. We will refer to the noise distribution that subjects were instructed to listen for as the *target* and will distinguish between sequences on the basis of whether this target information was present. *Present* and *absent* responses were indicated by pressing different keys on the computer keyboard. Figure 2 also shows the different conditions of this study, along with the appropriate responses when listening for the presence of the white noise distribution. Half the subjects listened for the white noise distribution, and the other half listened for flicker noise. After responding, listeners received feedback as to whether their answer was correct. Feedback occurred throughout the experimental session.

Listeners heard one practice block of trials and four blocks of experimental trials. The practice block was one of the later experimental blocks. All listeners heard the trials within each block in different random orders. The entire experimental session lasted approximately 1.5 hr.

Results

For each listener, the percentage of correct responses (answering *present* when the target noise distribution was present and *absent* when the target noise distribution was not present) was calculated for the different conditions. These data were then analyzed with a three-way ANOVA, with factors of noise type (listening for white vs. flicker noise), sequence length (15, 30, 45, 60, or 75 notes), and condition—the 10 bi- and unidimensional sequence conditions. Sequence length and condition were within-subject factors, whereas noise type was a between-subject variable. This analysis revealed a main effect of sequence length, $F(4, 56) = 8.89, p < .001$, and condition, $F(4, 56) = 11.60, p < .001$, but no main effect for noise type, $F(1, 14) = 1.82$, *ns*. None of the two-way interactions was significant, but the three-way interaction among length, condition, and noise type was significant, $F(36, 504) = 2.44, p < .001$. The percentages correct for the 10 conditions (averaged across noise type and length) are shown in the top half of Table 3. Given the lack of a main effect for noise type, all further analyses collapsed across this variable.

Before discussing any subsequent analyses, it is important to note one interesting aspect of these results. Inspection of the top half of Table 3 suggests that accuracy differed depending on whether the target information was actually present in the sequence. For sequences in which the target information was present, accuracy increased relative to sequences in which the target information was not present. This effect is particularly pronounced for the bidimensional sequences, which had response levels of 89.7% (both congruent and incongruent sequences) when the target was present, versus 59.4% and 56.7% (congruent and incongruent sequences, respectively) when the target was not present. Because of this difference, all subsequent analyses distinguished between sequences in which the target information was present and sequences in which the target information was absent.

Analyses were performed in which the different redundant, nonredundant, and unidimensional conditions were examined, averaging across the sequence lengths. In these analyses, the congruent redundant sequences were compared with the incongruent redundant sequences, the unidimensional pitch sequences were compared with the unidimensional stimuli loudness sequences, and the two nonredundant

[9] Pitch and loudness changes were not perfectly correlated because there were 14 levels of pitch and only 12 levels of loudness.

Table 3
Mean Percentage Correct for the Uni- and Bidimensional Conditions of Experiment 2

Condition	Target information	Percentage correct	*F* for difference between the means
Bidimensional redundant			
Congruent	Present	89.7	$F(1, 15) = 0.0$, *ns*
Incongruent	Present	89.7	
Bidimensional nonredundant			
Pitch = White, Loudness = Flicker	Present	79.1	$F(1, 15) = 3.83$, *ns*
Pitch = Flicker, Loudness = White	Present	74.1	
Unidimensional			
Pitch	Present	79.1	$F(1, 15) = 1.16$, *ns*
Loudness	Present	75.0	
Bidimensional redundant			
Congruent	Absent	59.4	$F(1, 15) = 0.35$, *ns*
Incongruent	Absent	56.7	
Unidimensional			
Pitch	Absent	66.9	$F(1, 15) = 2.58$, *ns*
Loudness	Absent	74.1	
Bidimensional redundant	Present	89.7	
Bidimensional nonredundant	Present	76.6	
Unidimensional	Present	77.0	
Bidimensional redundant	Absent	58.0	
Unidimensional	Absent	70.5	

sequences were compared. The results of these five comparisons are shown in the last column of the top of Table 3; none of the pairs of means differed significantly. Accordingly, these pairs were averaged, producing five conditions: redundant sequences in which the target information was present, redundant sequences in which the target information was absent, unidimensional sequences in which the target information was either present or absent, and nonredundant sequences in which the target information was present.

Comparisons then looked for effects of having the target information present (or absent) simultaneously in pitch and loudness dimensions. The means for the five conditions were compared with one-way ANOVAs with Bonferroni corrections. The most important test involved comparison of the bidimensional redundant condition with the bidimensional nonredundant condition and the unidimensional condition when the target was present. In this case, detection accuracy for the bidimensional redundant condition was significantly greater than detection in both the bidimensional nonredundant condition, $F(1, 15) = 28.64$, $p < .01$, and the unidimensional condition, $F(1, 15) = 17.83$, $p < .01$. No difference in performance occurred between the bidimensional nonredundant condition and the unidimensional condition, $F(1, 15) = 0.01$, *ns*.

The pattern of results becomes more complicated for conditions in which the target information was not available. Before correcting for multiple comparisons, the bidimensional redundant condition differed significantly from the unidimensional condition; after correction, this difference approached significance, $F(1, 15) = 7.94$, $p < .08$. What makes this result so striking is that the difference between these means is in a direction opposite from what was expected—the bidimensional redundant condition produced substantially poorer performance than the unidimensional condition. Two additional analyses compared means across response type. First, the two bidimensional redundant conditions (the target information was and was not present) were compared; these means differed significantly, $F(1, 15) = 57.37$, $p < .001$. Second, performance in the unidimensional conditions (across the presence or absence of target information) was examined; here, no difference occurred, $F(1, 15) = 1.66$, *ns*.

Finally, the effects of the different sequence lengths were examined. A two-way ANOVA, with the factors of sequence length and condition, revealed significant main effects for both variables, $F(4, 60) = 8.24$, $p < .001$, and $F(4, 60) = 17.08$, $p < .001$, respectively. The two-way interaction did not reach significance, $F(16, 240) = 0.82$, *ns*. Figure 3 shows the average correct percentage for the different conditions as a function of sequence length. Although the length of the sequence influenced accuracy, this effect was constant across all conditions.

Discussion

The results of this study suggest, in keeping with our earlier predictions, that accuracy for identification of fractal structure changes when this information is simultaneously available in both pitch and loudness dimensions of a tone sequence. Specifically, when target information was present in both dimensions accuracy increased relative to when this target information was available in only a single dimension. Interestingly, when the target information was not present in either dimension, accuracy decreased substantially. Although somewhat nonintuitive, such a result falls naturally out of the probability model that we described earlier.

How well does our probability model predict the results of this study? To test our model, we first determine the hit and

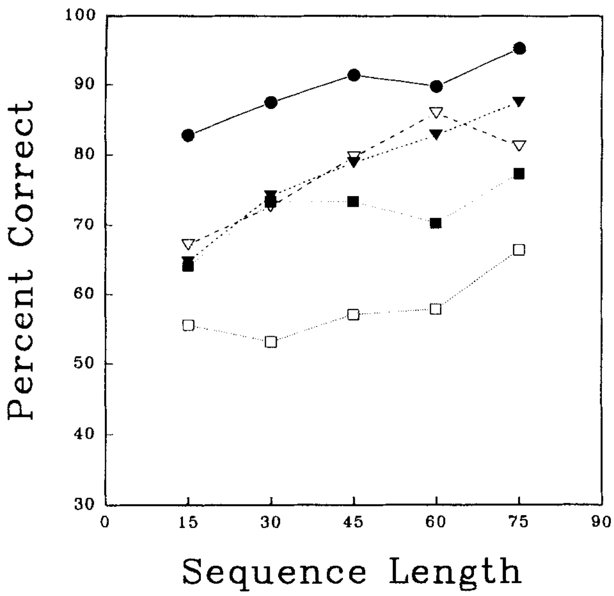

Figure 3. Average percentage correct as a function of sequence length for the bidimensional and unidimensional conditions of Experiment 2.

false alarm rates, or $p(+:+)$ and $p(+:-)$, using the results of the unidimensional condition when the target was present and absent, respectively. From Table 3, we see that $p(+:+) = p = .77$ and that $p(+:-) = q = .30$. Substituting these values in the equations described earlier, we calculate that $p(+$, 2 targets) (the probability of a *present* response given that two targets were present) is .95, $p(+$, 1 target) = .84, and $p(+$, 0 targets) = .51. From Table 3, we have $p(+$, 2 targets) = .90, $p(+$, 1 target) = .77, and $p(+$, 0 targets) = .42. Slight adjustments in p and q allow us to fit our data more closely. For $p = .70$ and $q = .25$, we calculate $p(+$, 2 targets) = .91, $p(+$, 1 target) = .78, and $p(+$, 0 targets) = .44. In Figure 4, we show the predicted hit and false alarm rates as a function of target number together with the observed rates. The agreement between the additive model and our data indicates that loudness and pitch are processed independently and that identification accuracy is understandable simply in terms of the detection of two independent samples.

Throughout this experiment, we have focused on the ways in which listeners might be made more aware of the underlying structure of the auditory sequences presented to them. It is instructive to note, however, that the current experiment is related to work on the integrality and separability of dimensions in perceptual organization (Garner, 1974, 1981; Garner & Morton, 1969; Pomerantz, 1981). Although a full account of this topic is beyond the scope of this article, the issues involved in the integrality versus separability of stimulus dimensions can be loosely characterized as follows. When perceiving multidimensional stimuli (such as pitch and loudness in auditory stimuli or the size and orientation of visual objects), two dimensions are perceptually independent if perception of one dimension is not influenced by the perception of the other dimension (Ashby & Townsend, 1986; Garner & Morton, 1969). Perceptual interaction occurs when perception of one dimension is somehow contingent on perception of a second dimension. In one particular test of perceptual interaction (speeded sorting), integral dimensions interact by facilitating identification of a stimulus when the dimensions of this stimulus are redundant relative to when a single dimension defines a category (Garner, 1976). Such facilatory effects have been referred to as *redundancy gains* and are a hallmark indicator of perceptual interaction.

Previous researchers have found evidence for integrality of auditory dimensions, using both speech sounds (Day & Wood, 1972; Wood, 1974; Wood & Day, 1975), as well as more basic tone attributes of pitch and loudness (Grau & Kemler Nelson, 1988). The most comprehensive research on interactions among auditory dimensions is recent work by Melara and Marks (1990a, 1990b, 1990c). This work provides a wealth of interesting results and generates extensive empirical support for the idea that tone attributes, such as pitch, loudness, and timbre, are processed interactively by listeners. Our current study is methodologically related to studies examining perceptual interaction, and our results indicate the presence of redundancy gains in that the cooccurrence of a common fractal structure in two auditory dimensions facilitated detection of this information. This suggests, contrary to our earlier probabilistic model, that pitch and loudness contours are in fact being processed interactively by listeners. As such, we are in the curious position that the same set of data simultaneously supports both an independent and interactive model of performance.

Given this situation, we are faced with a problem. How do we reconcile the divergent implications of our results? In answer to this question, we should stress that our experiment is not an adequate test of whether pitch and loudness contours are perceptually interactive or independent. Garner (1976, 1981) has proposed that dimensional interaction be defined by a set of converging operations involving a variety of related tasks. In the current case, we have only a single task

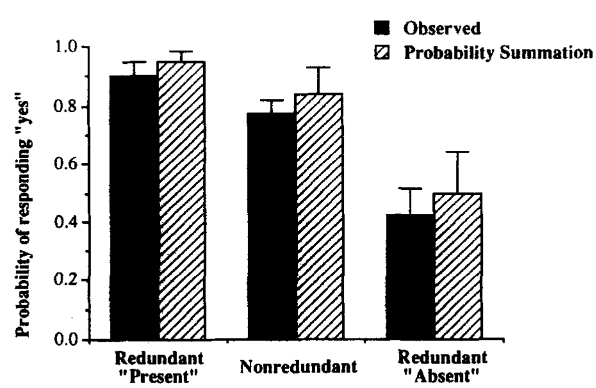

Figure 4. Comparison of the observed probability of responding *yes* with the probability summation model for Experiment 2.

(speeded sorting), and our method is at best analogous to this task, not identical to it. Moreover, even within the task of speeded sorting, we did not include the different combinations of stimulus sets (i.e., baseline, positively correlated, negatively correlated stimuli, filtering stimuli) that make the evaluation of perceptual interaction possible. Given that our initial interest was not examining the possible interaction of pitch and loudness dimensions, this oversight is understandable; we are interested primarily in listener's apprehension of the underlying fractal structure and only peripherally in the relationship between pitch and loudness dimensions of tone sequences.

Methodological differences suggest that the current study is not a fair test of perceptual interaction between pitch and loudness contours. As such, it seems most conservative to accept the implications of our probability model, which suggests independent processing of pitch and loudness contours. Unfortunately, though, this raises a subsequent problem. Namely, how do we reconcile our findings of independence between pitch and loudness dimensions with other research suggesting perceptual interaction between pitch and loudness (Grau & Kemler Nelson, 1988; Melara & Marks, 1990b)?

There are some important differences between previous work investigating interaction among auditory dimensions and the current study, however. Previous work (Grau & Kemler Nelson, 1988; Melara & Marks, 1990b) observed interactions among pitch–loudness dimensions during processing of a single note. In contrast, our study examined the processing of pitch and loudness contour information that is available only over time and as a result of integrating pitch–loudness differences between successive tone events. This is a dramatically different process than occurs when processing a single note. This difference in the inherent nature of the judgment is quite important and could underlie the distinction between our findings of additivity and other work suggesting interaction. A related difference is that traditional work on dimensional interaction uses stimuli that are well defined, whereas our stimuli are more ambiguous. For example, in Melara and Marks's (1990b) Experiment 1a, listeners classified single tones that could have one of two values of pitch and loudness: 900 or 950 Hz for pitch and 60 or 70 dB for loudness. As such, the categories as well as the exemplars of these categories were well defined; loud or soft and high or low pitch. In contrast, our categories were defined by the slope of the power spectrum, and we had multiple exemplars for each power spectrum. Thus, it is likely that our categories of stimuli, white versus flicker noise, were much less well defined for our listeners. Given both of these differences, it is not surprising that we did not necessarily find evidence for interaction between pitch and loudness dimensions. Simply put, although pitch and loudness might interact when processing single tones, there is no reason that these dimensions should interact in the formation of more complex auditory objects.

Finally, brief mention should be made of two other results of this study. First, this experiment found that having congruent changes between pitch and loudness did not produce better detection. This result is striking, given that congruent sequences provided stimuli that were isomorphic in terms of pitch and loudness variation. It is possible, of course, that listeners were simply unaware of the relationship between these dimensions. However, this result fits nicely with the idea that pitch and loudness contours combine additively; facilatory effects produced by dimensional congruity tend to be an indicator of perceptual interaction between dimensions (Melara & Marks, 1990b, 1990c). Second, this experiment found that sequence length had a relatively consistent and straightforward impact on identification of these sequences, with increasing accuracy associated with increasing sequence length. Probably the most surprising aspect of this manipulation, though, was that identification was quite good even for very short sequences. For the bidimensional redundant condition in which the target was present, accuracy was over 80% at sequence lengths as short as 15 notes. Such a result suggests that sufficient information exists for detecting gross differences in fractal structure in relatively small samples of the noise distribution.

Experiment 3: Discrimination of Noise Distributions

The purpose underlying Experiment 3 was to look in more detail at listeners' ability to discriminate between noise distributions. In this case, however, rather than look at noise discrimination as a function of the number of dimensions within which this information is embedded, the current study looked at changes in discrimination as a function of varying fractal dimension. Specifically, we investigated listeners' abilities to distinguish between noise distributions across a range of fractal dimensions. A secondary manipulation in the current study was to further examine the influence of the total range or coarseness of the coding on discrimination.

Why might sensitivity vary as a function of fractal dimension? Earlier, we noted that many natural sources, both auditory and visual, can be conveniently described in terms of their fractal dimension. Voss and Clarke's (1978) research demonstrated that a number of diverse auditory sources could be related to a single fractional Brownian noise—$1/f$ or flicker noise. Additionally, fractional Brownian noises have been recognized as providing useful descriptions for visual scenes. For example, Keller et al. (1987) found that the silhouettes of tree lines and mountain tops could be characterized as self-affine noises varying in their fractal dimension. Other examples include work by Sayles and Thomas (1978a, 1978b), who measured the roughness of a number of objects ranging from steel balls to runways and found that all could be described by their fractal dimension. This work found a variety of fractal dimensions for these sources, primarily in the range between −2.0 and −3.0.

The observation that both auditory and visual sources are fractal suggests that people might be sensitive to fractal information. Of course, in a general sense, investigation of this idea has been a recurrent theme of these experiments. More specifically, though, it is possible that perceptual sensitivity might be somehow "tuned" to the regularities of fractal structure in the environment. As such, sensitivity to fractal structure might vary according to fractal dimension, with greater sensitivity occurring for structures whose fractal dimension matches naturally occurring fractal sources.

For example, in the work by Knill et al. (1990) described earlier, maximal discrimination of fractal images occurred for displays with a fractal dimension of 2.5—the same range of fractal dimensions of naturally occurring landscapes. Knill et al. (1990) interpreted this finding in terms of the efficiency of perception, with our perceptual systems being maximally tuned for structures that are likely to occur in the environment.

When considering visual sources, one might expect increased sensitivity to stimuli with a fractal dimension between −2.0 and −3.0. For auditory stimuli, we might anticipate increased sensitivity around the region of −1.0, based on Voss and Clarke's (1978) results. It is possible, moreover, that there will be crossover between visual and auditory modalities, resulting in a discrimination function having roughly an inverted U shape, with either 1 or 2 peaks, in the regions of −1.0 and −2.0 to −3.0. Alternatively, it might be that listeners are not maximally sensitive to the fractal structure occurring naturally within our environment. If so, then we should see relatively constant discrimination across a range of fractal structure.

Method

Subjects

Five listeners from the University of Toronto, Scarborough, were paid for participating. Each listener ran in multiple experimental sessions over the course of 1–2 months.

Apparatus and Stimulus Materials

Stimuli were generated with the same equipment as in Experiments 1A, 1B, and 2. All listeners heard the auditory passages over a Peavey KB-60 amplifier.

A series of fractional Brownian noise distributions were generated (see Figure 5) in which the slope of the power spectra for these noises varied systematically from 0.0 to −3.9 in steps of 0.3. These noise distributions (coded as sequences of random numbers) were used to generate auditory sequences varying in pitch using the same method as in the previous three experiments. Although each sequence consisted of 100 notes, listeners heard only the first 50 for each sequence. This decrease in the length of each sequence is justified, given the results of Experiment 2, which found that identification of noise distributions could be quite accurate even in sequences as short as 15 notes. Thirty different sequences were generated for each noise distribution.

Conditions

There were two conditions in this experiment, corresponding to the total range into which the random numbers could be mapped. In the first case, there were 24 discrete levels of pitch, corresponding to two octaves of a chromatic scale starting from C4 (middle C, 260 Hz) and ending on B5. This condition will be referred to as the *narrow* condition. In the second case there were 48 discrete levels of pitch, corresponding to four octaves of a chromatic scale, beginning with C3 and ending on B6. This condition will be referred to as the *wide* condition. For all stimuli, the duration of each tone was 200 ms.

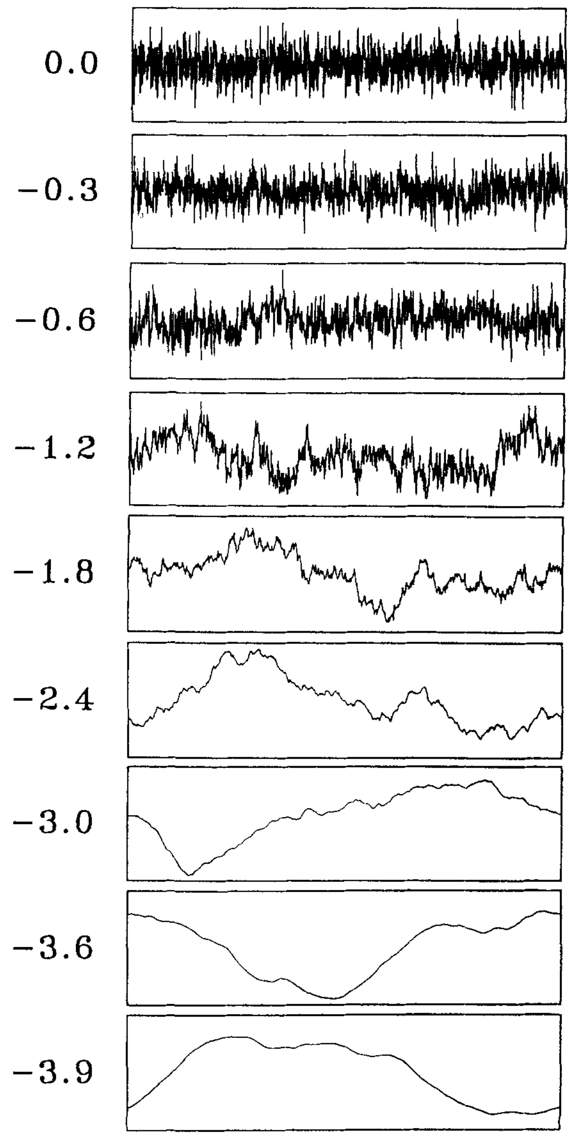

Figure 5. Samples of noise distributions used to generate stimuli for Experiment 3. (Slopes of the power spectra range from 0.0 to −3.9.)

Procedure

During each experimental session, listeners heard a series of sequences based on two different noise distributions. One of the noise distributions was arbitrarily labeled Category A, whereas the other noise distribution was labeled Category B. The listener's task was to correctly categorize each of the sequences into either the A or B category. A single session consisted of 200 experimental trials, 100 from each category. As there were only 30 different exemplars of each category, individual sequences were repeated within sessions. For each trial, the sequence was chosen randomly from the total set of exemplars of that noise distribution.

At the beginning of each session, listeners heard three samples of each of the categories. Subsequently, listeners began the block of 200 trials. After hearing each trial, the computer prompted the listener to respond as to whether the sequence was an exemplar of

Category A or Category B. Feedback as to whether the listener's categorization was correct was given throughout the entire experimental session. A single session lasted approximately 1 hr.

Training phase. Prior to starting the experimental sessions, listeners received training in distinguishing between the different noise distributions. At the beginning of this training phase, the structure of the different noise distributions was explained in a similar fashion as in Experiments 1A, 1B, and 2. Listeners then ran a number of blocks of trials in which the separation between Categories A and B, in terms of fractal dimension, was quite large. For example, initially listeners discriminated between noise distributions with a slope of 0.0 and −3.0. The difference, in terms of fractal dimension, between the two noise distributions was gradually decreased (over multiple experimental sessions) until listeners were discriminating between noise distributions with slopes of 0.0 and −1.0. At this point the listener began the experimental phase.

Experimental phase. During the experimental phase, listeners discriminated between two noise distributions differing in the slope of their respective power spectra. Based on pilot work, it was decided that a reasonable separation between noise distributions, in terms of fractal dimension, was a difference of 0.6. The range from 0.0 to −3.9 was sampled at intervals of 0.3. Thus, in one session listeners would discriminate sequences in which the slopes of the power spectra were 0.0 and −0.6. Subsequent to this session, listeners would then discriminate between slopes of −0.3 and −0.9 and so on up to slopes of −3.3 and −3.9. This produced 12 different pairs of noise distribution comparisons.

Three of the listeners discriminated noise distributions starting with the least coherent noises and ending with the most coherent noises (i.e., 0.0:-0.6; −0.3:-0.9; . . . , −3.3:-3.9), whereas the remaining two listeners ran the experimental sessions in the reverse order (i.e., −3.3:-3.9; −3.0:-3.6; . . . , 0.0:−0.6). All listeners participated in the narrow condition first and the wide condition second.

Results

The percentages of correct and incorrect responses for Categories A and B were converted into d', using the percentage correct for Category A as the hit rate, and the percentage incorrect for Category B as the false alarm rate. Initial analyses determined that there were no obvious differences in performance as a function of the order these sequences were heard; thus, this variable was ignored in subsequent analyses.

The d's were analyzed with a two-way ANOVA with within-subjects factors of condition (narrow versus wide) and slope (0.0:-0.3, . . . , −3.3:-3.9). This analysis revealed a main effect of slope, $F(11, 44) = 8.01, p < .001$, no effect of condition, $F(1, 4) = 1.62$, *ns*, and no interaction between the two, $F(11, 44) = 0.61$, *ns*. Figure 6 shows the average d' values across listeners for the narrow and wide conditions, as well as the average of the two conditions. These values are plotted in terms of the intermediate slope value between the two noise distributions being discriminated (i.e., discrimination performance for slopes of 0.0 and −0.6 are plotted at −0.3). Also shown in Figure 6 is the range of fractal dimensions for the previously discussed inventory of visual landscapes and auditory sources.

Discussion

The discrimination function shown in Figure 6 demonstrates that sensitivity to fractal structure varies as a function

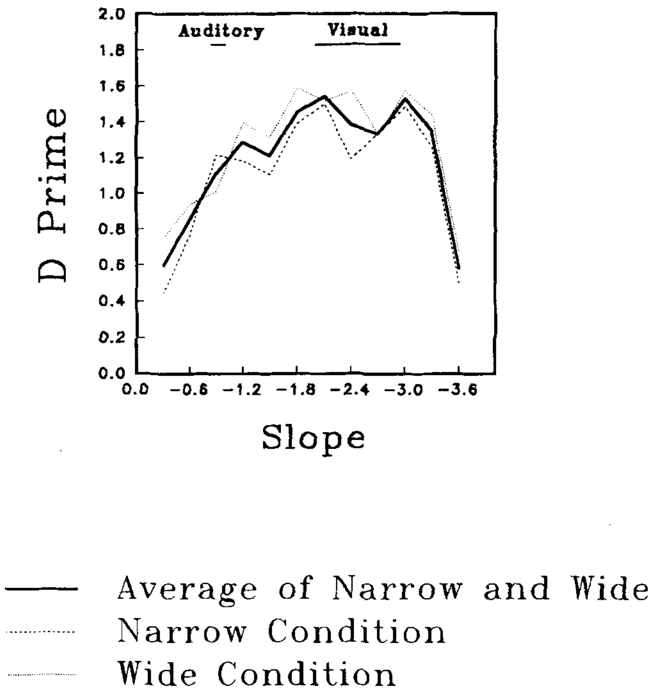

Figure 6. Mean d' for the narrow, wide, and averaged narrow and wide conditions of Experiment 3. (Also shown are the ranges of visual and auditory sources.)

of fractal dimension, with the discrimination function approximating an inverted U-shaped curve. Interestingly, however, we find only limited support for the idea that sensitivity to fractal structure is related to the statistical regularities of fractal structure in our environment. Although there was a discrimination peak for slopes between −2.0 and −3.0 (the range of visual landscapes), there was no heightened sensitivity for slopes around −1.0 (the slope value underlying auditory sources). It is unclear why, if perceptual sensitivity is related to statistical regularities of the environment, that auditory discrimination should peak in the range of visual stimuli, with no heightened sensitivity in the range of auditory sources.

Alternatively, it might be that the shape of the discrimination curve is not a result of learned environmental regularities but is simply a fortuitous by-product of the process of discriminating between the noise distributions. In an experiment similar to the current study, Gilden, Schmuckler, and Clayton (in press) had observers discriminate between pairs of line drawings of fractal noises differing in the slope of their power spectra. Discrimination functions for these displays had essentially the same shape as seen in the current study: The curves had an inverted U shape, peaking in the region between −2.0 and −3.0. In subsequent analyses, however, this discrimination was modeled by algorithms that were essentially blind to the integrated hierarchical nature of fractal structure. Instead, these models operated on the basis of signal–noise distinctions or by decomposing the fractal structure into smooth and rough components of the sequences and comparing the variances of these components. The success of this model argues strongly against the idea that our perceptual systems are tuned to the statistical distribution of environmental forms (Knill et al., 1990). Currently, we are extending these results, looking at discrimination of different fractal and nonfractal noises, in an attempt to provide further data bearing on perceivers' abilities to detect the information inherent in fractal structure. Viewed in this light, the current experiment is best considered as a preliminary report of the discrimination of fractal structure.

General Discussion

The results of these studies can be looked at in a number of ways. First, these experiments represent an investigation into listeners' abilities to use "pure" contour information, or the recognition of contours that are mathematically featureless and relatively independent of their mode of representation. The ability to identify contour is critical for auditory perception, being implicated as a crucial component in perceiving the overall structure of the auditory environment, or what has been termed auditory scene analysis (Bregman, 1990). From this view, these studies examine our abilities to perceive relational information when presented by different auditory dimensions (pitch, loudness, and duration) either in isolation (Experiments 1A and 1B) or simultaneously (Experiment 2). All of these studies demonstrated listeners' sensitivity to the global structure of the featureless noise distributions used to generate the auditory passages. Experiments 1A and 1B showed that listeners could perceive the abstract information underlying contours, although not all modes of representing this information were perceptually relevant. Experiment 2 demonstrated that the presentation of contours simultaneously in pitch and loudness dimensions of a sequence facilitated identification of these noise distributions in an additive fashion. Overall, these results fit well into the expanding literature on perception of complex auditory information.

In our attempt to study contours that were independent of sources of information other than relational, we used random number distributions having fractal structure. Because of this, a secondary framework for understanding these studies involves people's abilities to perceive the integrated hierarchical information characterizing fractal sources. It has been noted in recent years that our environment is replete with fractal structure; evidence for the psychological relevance of fractal structure has been much less forthcoming, however. Essentially, this project is a step toward making good on this implied promissory note by investigating the sensitivity of listeners to fractal structure when mapped into changes in auditory sequences. Moreover, this work has attempted, albeit informally, to compare sensitivity to fractal structure to other, more traditional characterizations of auditory sequences, namely auditory stream segregation.

In this sense, the current studies are somewhat analogous to work by Cutting and Garvin (1987), in which the importance of fractal structure in similarity ratings of images was compared with the role of more classical characterizations of these images. Similar to Cutting and Garvin's (1987) results, we find that fractal structure plays a role in discrimination, although it is not an overwhelming source of information. Experiments 1A and 1B, for example, demonstrate that the perception of auditory fractal contours varying in pitch, loudness, and duration cannot be wholly accounted for on the basis of a more simple explanation relying on auditory streaming. Though we are reluctant to discount the importance of streaming in these studies (our reasons were enumerated earlier), these results suggest that there was clearly more to these sequences than streaming. Experiment 3, however, points out the limits of explanations relying in terms of fractal structure. This experiment demonstrated that discrimination of noises varying in their fractal structure could not be wholly accounted for by the differences in fractal dimension (which predicts constant discrimination with constant difference in fractal dimension) or by heightened sensitivity to statistical regularities of the environment (which predicts increased discrimination for noises with fractal dimension of −1.0 and between −2.0 to −3.0). It appears, then, that fractal information might simply be another potential source of structure, working in concert (or sometimes in opposition) with other aspects of auditory structure.

The fact that both visual scenes and auditory stimuli have fractal descriptions raises a variety of cross-modal questions, some of which have already been discussed. One of the more fascinating results of the current project involved finding that not all mappings of the contour information could be perceived by listeners. Similar questions arise concerning the effectiveness of different encodings in other sensory modalities. Gilden and Schmuckler (1989) examined contour perception when contour information was presented in different visual dimensions. In a comparable study to Experiments 1A and 1B, Gilden and Schmuckler (1989) chose five different isomorphic representations of white, flicker, and brown noise distributions and had observers categorize these representations. The different representations used consisted of bands of rectangles differing in height or width, bands of stripes differing in monochromatic brightness, the vertical motion of a line, and the brightness of a patch over time. The results of this study revealed a hierarchy of category discriminability among the isomorphs in contour formation. Categorization performance was enhanced when contour information occurred in displays making use of spatial layout and was maximal when the varying dimension of the representation used scales familiar from everyday experience (e.g., distance metrics for height and width are more familiar than brightness metrics). In addition, as already discussed, Gilden et al. (in press) examined sensitivity to line drawings of scaling noises as a function of their power law exponent and found that sensitivity to contour information varied over the range of fractal dimensions tested, with the discrimination curve also approximating an inverted U-shaped function.

The current studies represent an initial investigation into contour formation and the perception of fractal structure. Unfortunately, the results of these studies raise more questions than they answer. Issues concerning the perceptual relevance of different encodings of auditory structure have, for example, been only lightly touched. Another question needing further work involves the effect of the coarseness of coding from continuous to discrete representation. In a few different manipulations (Experiments 1B and 3), we consistently failed to find an effect of this variable, an unusual and somewhat nonintuitive result. There are similar issues concerning the additivity versus interaction of simultaneous pitch and loudness contours. Although our study did not provide a fair assessment of the possible interaction of these sources, the results did strongly suggest independence between these dimensions. Finally, we have a host of issues concerning the apprehension of fractal structure per se, its psychological relevance as a factor in perceptual processing, and its relationship to other factors traditionally thought to play a role in auditory perception. Investigating these issues promises to be an illuminating task, providing insight into contour perception and auditory processing generally.

References

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Cutting, J. E., & Garvin, J. J. (1987). Fractal curves and complexity. *Perception & Psychophysics, 42,* 365–370.

Day, R. S., & Wood, C. C. (1972). Interactions between linguistic and nonlinguistic processing. *Journal of the Acoustical Society of America, 51,* 79 (Abstract).

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88,* 503–522.

DeWitt, L. A., & Crowder, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception, 3,* 259–274.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85,* 341–354.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology, 1,* 30–49.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49,* 524–531.

Dowling, W. J., & Hollombe, A. W. (1977). The perception of melodies distorted by splitting into several octaves. *Journal of the Acoustical Society of America, 21,* 60–64.

Essens, P. J. (1986). Hierarchical organization of temporal patterns. *Perception & Psychophysics, 40,* 69–73.

Fraisse, P. (1978). Time and rhythm perception. In *Handbook of perception* (Vol. 8, pp. 203–254). New York: Academic Press.

Gardner, M. (1978, April). Mathematical games: White and brown music, fractal curves and one-over-f fluctuations. *Scientific American,* 16–31.

Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology, 8,* 98–123.

Garner, W. R. (1981). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 287–308.

Garner, W. R., & Morton, J. (1969). Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin, 72,* 233–259.

Getty, D. J. (1975). Discrimination of short temporal intervals. *Perception & Psychophysics, 18,* 1–8.

Gilden, D. L., Bertenthal, B. I., & Othman, S. (1990). Image statistics and the perception of apparent motion. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 693–705.

Gilden, D. L., & Schmuckler, M. A. (1989). *The perception of fractal structure.* Paper presented at the 1989 meeting of the Association for Research in Vision and Opthalmology, Sarasota, FL.

Gilden, D. L., Schmuckler, M. A., & Clayton, K. (in press). The perception of natural contour. *Psychological Review.*

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross-classifications. *Journal of the American Statistical Association, 49,* 732–764.

Grau, J. W., & Kemler Nelson, D. G. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General, 117,* 347–370.

Hays, W. L. (1963). *Statistics.* New York: Holt, Rinehart & Winston.

Jones, M. R. (1981). A tutorial on some issues and methods in serial pattern research. *Perception & Psychophysics, 30,* 492–504.

Keller, J. M., Crownover, R. M., & Chen, R. U. (1987). Characteristics of natural scenes related to the fractal dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 9,* 621–627.

Knill, D. C., Field, D., & Kersten, D. (1990). Human discrimination of fractal images. *Journal of the Optical Society of America, 7,* 1113–1123.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception: Volume 2. From perception to cognition* (pp. 275–382). New York: Academic Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Mandelbrot, B. B. (1977). *Fractals: Form, chance and dimension.* San Francisco: Freeman.

Marshak, F. M. (1986). *Stochastic modes in the visual perception of natural surfaces.* Unpublished manuscript.

Melara, R. D., & Marks, L. E. (1990a). HARD and SOFT interacting dimensions: Differential effects of dual context on classification. *Perception & Psychophysics, 47,* 305–325.

Melara, R. D., & Marks, L. E. (1990b). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics, 48,* 169–178.

Melara, R. D., & Marks, L. E. (1990c). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 398–414.

Miller, G. A., & Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America, 22,* 637–638.

Navon, D. (1977). Forest before the trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383.

Pentland, A. P. (1984). Fractal-based description of natural scenes. *IEEE Pattern Analysis and Machine Intelligence, 6,* 661–674.

Pentland, A. P. (1985). *On perceiving 3-D shape and texture.* Paper presented at the Symposium on Computational Models in Human Vision, Rochester, NY.

Pomerantz, J. R. (1981). Perceptual organization in information processing. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization.* Hillsdale, NJ: Erlbaum.

Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General, 112,* 516–540.

Pomerantz, J. R., & Sager, L. C. (1975). Asymmetric integrality with dimensions of visual pattern. *Perception & Psychophysics, 18,* 460–466.

Povel, D. J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 3–18.

Sayles, R. S., & Thomas, T. R. (1978a). Surface topography as a nonstationary random process. *Nature, 271,* 431–434.

Sayles, R. S., & Thomas, T. R. (1978b). Topography of random surfaces. *Nature, 273,* 573.

Schmidt, R. C., Beek, P. J., Treffner, P. J, & Turvey, M. T. (1991). Dynamical substructure of coordinated rhythmic movements. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 635–651.

Schouten, J. F. (1962). On the perception of sound and speed; Subject time analysis. *4th International Congress on Acoustics, Copenhagen Congress Report II,* 201–203.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perceptions of melodies: The role of melodic contours. *Child Development, 55,* 821–830.

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.

Voss, R. F., & Clarke, J. (1978). 1/f noise in music: Music from 1/f noise. *Journal of the Acoustical Society of America, 63,* 258–263.

White, B. W. (1960). Recognition of distorted melodies. *American Journal of Psychology, 73,* 100–107.

Wood, C. C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception & Psychophysics, 15,* 501–508.

Wood, C. C., & Day, R. S. (1975). Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception & Psychophysics, 17,* 346–350.

Received May 20, 1991
Revision received September 11, 1992
Accepted September 14, 1992 ■