Chapter 2

Correlativity of Perception

2.1 Introductory Remarks

In the present chapter we discuss a mechanism of perceptual grouping, the *correlativity of perception*, and suggest a way of its modeling. The background and successive steps in reasoning which have prepared the formulation of the principle of correlativity of perception are outlined in section 1.4; here we shall present it in detail.

In Section 2.2, "Principle of Correlativity of Perception", we illustrate the idea of correlative perception with examples of visual and audio perception. We also explain why the same data can be perceived in different contexts in a different way. From the standpoint of our model such a phenomenon is caused by the fact that the complexity of alternative representations depends on the context, implying the contextual dependence of the optimal representation which is associated with the percept. We also explain the effect of apparent motion from stroboscopic images.

In Section 2.3, "Model of Correlative Perception", the mathematical machinery for implementing correlative perception is outlined. We consider an example of motion recognition in a succession of instant images and show that the problem of representing the data in terms of generative elements and their transformations is reduced to the problem of finding correlated messages under various distortions of the data.

In Section 2.4, "Method of Variable Resolution", we describe how a directional search for generative elements and their transformations can be performed. The idea of the method is detecting the correlation between the images with reduced resolution wherein correlated elements are easy to find, and then gradually restoring the resolution while adjusting the images in order to maintain the correlation between the known elements. This way both similar elements and the transformations which provide for their high correlation are found.

In Section 2.5, "Complexity of Transformation as a Distance", we show that the complexity of transformation which links two patterns can be considered as a metrical distance between these patterns. An important property of this distance is that it formalizes the idea of likeness, not being based on the measure of identity. In particular, such a measure is efficient for recognizing similarity, e.g. when patterns have the same shape, being different in size.

In Section 2.6, "Distinctions of the Model", we discuss the peculiarities of our approach. We start with the questions which are to be answered by our model. Then we note that the hierarchization in data self-organization results from optimizing data representations and that the optimization criterion can replace threshold criteria in pattern segregation. We mention that pattern separation in the model is based on the pattern similarity but not on dissimilarity. Finally, we compare our artificial perception approach to pattern recognition with that of artificial intelligence.

In Section 2.7, "Summary of the Chapter", the main items of the chapter are recapitulated.

2.2 Principle of Correlativity of Perception

As already said in Section 1.4, by *correlativity of perception* we understand its capacity to discover similar configurations of stimuli, i.e. structurally arranged groups, and to form configurations of a higher level from them.

Configurations of stimuli themselves are called *low-level patterns*, and configurations of the relationships between low-level patterns are called *high-level patterns.*

This hierarchical scheme of data representation is endowed with a feedback, which guides the process of representing data with least complexity, where *complexity of data* is measured in the sense of Kolmogorov, i.e. as the amount of memory storage required for the algorithm of the data generation (Kolmogorov 1965; Calude 1988).

For example, in Fig. 2.1 one can see a collection of pixels (stimuli) which form symbols A (low-level patterns) which in turn form a contour of symbol B (high-level pattern). Obviously, instead of storing all the pixels, it is more efficient to store their configuration for one symbol A and then to store the contour of B.

Another important property of such a representation is that high-level patterns are stable with respect to changes of low-level patterns. For example, the substitution of Z's for A's in Fig. 2.1 would not influence on the perceptibility of B. The stability of high-level patterns with respect to changes of low-level patterns is described by Palmer (1982; 1983, p. 328).

Moreover, high-level patterns can be recognized without even recognizing

```
AAAAAAAAAAAAAAA
 AAAAAAAAAAAAAAA
 AAAA        AAAA
 AAAA         AAAA
 AAAA         AAAA
 AAAA         AAAA
 AAAA        AAAA
 AAAAAAAAAAAAAAA
 AAAAAAAAAAAAAAAA
 AAAA         AAAA
 AAAA          AAAA
 AAAA          AAAA
 AAAA          AAAA
 AAAA          AAAA
 AAAA         AAAA
 AAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAA
```

Figure 2.1: High-level pattern of B composed by low-level patterns of A

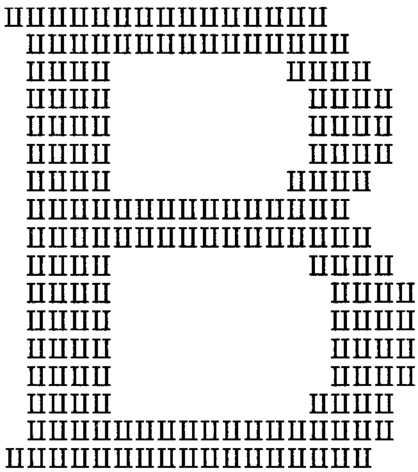

Figure 2.2: Pattern of B composed by unknown symbols

the underlying low-level patterns, Thus if some unknown, unrecognizable symbol were used instead of A, it still would be possible to recognize B by the relationships between these unknown symbols (Fig. 2.2).

This property is also inherent in human perception and learning. Most experiences are obtained from observing interactions of some objects whose peculiarities may be not so much important. For example, perceiving a traffic jam doesn't require any knowledge about the types of cars and their construction. Similarly, some ideas can be learned with no precise knowledge of basic concepts. For example, rules of arithmetics can be learned without strict axiomatization of numbers. Therefore, certain judgements can be made while considering relations between some concepts treated as "black boxes".

As we shall see further, recognizing intervals and chords doesn't require recognizing tones (pitch), which precisely corresponds to human perception. Indeed, most people lack absolute hearing (the capacity to recognize pitch of musical tones), but are capable to perceive intervals, melodies, and types of chords; e.g. distinguish between major and minor chords.

We restrict our attention to the case when important relationships arise between similar objects where the similarity is a cue for their recognition. In other words, we are looking for representations of data in terms of repetitive (correlating) messages, or generative elements, and their transformations.

At the same time, the recognition of similarity depends on some internal or external factors. In our model, such a factor is the total complexity of data representation which is determined by the complexity of low-level generative patterns and the complexity of the high-level patterns of their transformations.

Let us illustrate the influence of complexity criterion on recognizing similarity for further grouping. We shall show that the similarity can be relative, depending on the context, meaning the ambiguity of perception. Therefore, we explain the ambiguity of perception in terms of complexity of alternative data representations.

Consider a sequence of time events whose onsets are shown at the time axis in Fig. 2.3. This sequence can be represented in different ways. Its representation as a single rhythmic pattern under a constant tempo is shown in Fig. 2.4a. The representation corresponding to the repetition of the first three durations performed two times faster is shown in Fig. 2.4b where **R012** designates

- the call for the repetition algorithm **R** with the following parameters:
- begin from time **0**,
- repeat **1** time,
- perform the repetition **2** times faster.

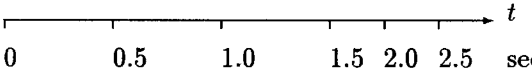

Figure 2.3: A succession of time events

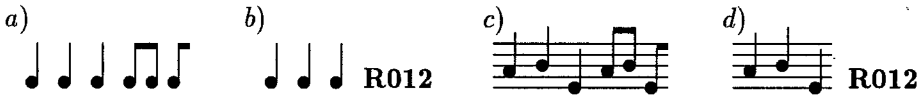

Figure 2.4: Four representations of the same succession of time events

Table 2.1: The Complexity of Representation of Time Events

	Representation *a*)	Representation *b*)	Representation *c*)	Representation *d*)
Complexity of rhythmic pattern	6 bytes	3 bytes	12 bytes	6 bytes
Complexity of its transformation	0 bytes	4 bytes	0 bytes	4 bytes
Total complexity	6 bytes	7 bytes	12 bytes	10 bytes

In this example the data representations has the following levels:

- Stimulus, corresponding to time events;
- Low-level rhythmic patterns which are groups of time events (either a single rhythmic pattern as in Fig. 2.4a, or two rhythmic patterns, the second being equal to the first but performed two times faster as in Fig. 2.4b);
- High-level pattern of the relationships between the rhythmic patterns (either trivial=no pattern in case of a single rhythmic pattern in Fig. 2.4a, or the pattern of repetition with a tempo change coded by the abbreviation **R012** in Fig. 2.4b).

Most likely, the given sequence of events is perceived as a long rhythmic pattern rather than a short one being repeated. This means that the representation in Fig. 2.4a is adequate, while that in Fig. 2.4b being inadequate.

However, if the same rhythmic sequence is placed into the melodic context shown in Fig. 2.4c, then the sensation of repetition becomes quite clear. Now the representation in Fig. 2.4d, where the idea of repetition is displayed, is rather natural and even can be considered as more adequate to perception than that in Fig. 2.4c.

To explain such an ambiguity in rhythm perception, that the same rhythmic passage can be perceived either as a long rhythmic pattern or as a repetition of a shorter one, estimate the complexity of the given representations. Suppose that one byte is needed to code a duration, and two bytes are needed to code a duration with pitch. Also suppose that to call the repetition algorithm, we need four bytes. Under such conventions the complexity of the given representations is estimated in Table 2.1. One can see that the representation of pure rhythm (Fig. 2.4a–b) as a long pattern is less complex (6 bytes against 7), whereas the representation of the same rhythm in melodic context (Fig. 2.4c–d) is more efficient as a repeat (10 bytes against 12).

A similar effect of recognizing twice shorter or twice longer durations as a repetition of the same pattern with another time scale arises while listening to a fugue whose theme is played in augmentation or diminution. For example, such a device is used at the beginning of the fourth fugue from J.S.Bach's *The Art of Fugue* shown in Fig. 2.5 where T indicates the entry of the inversion of the theme, A indicates its entry in augmentation, and D indicates its entry in diminution.

Thus the perception correlativity appears in two instances: Firstly, the low-level patterns are chosen similarly to each other (that is correlativity at the same level); and, secondly, they are chosen with respect to their interaction in high-level patterns (that is correlativity between levels). This hierarchical scheme of data representation is provided with a feedback, the criterion of least

Figure 2.5: Theme in augmentation and diminution from Bach's *The Art of Fugue*

complexity. It guides the process of data representation in the least complex way, while the complexity being shared between the generative patterns and their transformations (see Table 2.1).

In a sense, the high-level pattern shows how the similarity of the low-level patterns should be understood in the given context, yet without such a high-level pattern the similarity of low-level patterns may be dubious, as in the example illustrated by Fig. 2.4a–b. This implies that the measure of similarity between low-level patterns is influenced by the way how they are confronted in the high-level configurations.

Threfore, we can speak of *contextual similarity*, or *functional similarity* with respect to some unifying high-level pattern. Within certain context two patterns may be perceived not so much dissimilar as if taken separately, since in the given context they are charged with a common function with respect to the high-level pattern.

Such a contextual similarity does arise in the representation in Fig. 2.4d, and does not when there are no additional cues as in case of melodic context in Fig. 2.4a–b. The effect of contextual similarity with respect to melodic cues arises also while recognizing a fugue's theme in diminution or in augmentation.

Revert to the example illustrated by Fig. 2.4a–b where three eights are not recognized as a replication of thee quarter notes. We have shown that in a melodic context these two groups of durations can be identified as similar. The same effect can be obtained with no melodic cues but by placing these two patterns in an appropriate rhythmic context.

For example, insert accelerating triads between the two patterns shown in Fig. 2.6a as it is shown in Fig. 2.6b. Owing to a gradual acceleration, the pattern marked by the bracket is unambiguously perceived as a repetition of the pattern formed by the first three durations. Here, the representation of the time events as a repeat of the first three events under a tempo acceleration is simpler than using complex fractional durations with which one can write down the same progression of events (see Fig. 2.6c). This means that the last three durations are recognized as similar to the first three durations, as required.

The stroboscopic effect, i.e. the illusion of apparent motion from successive images with slightly different locations of an object which is used in moving pictures and television, can be also explained within our model. Some psychologists explain the effect of apparent motion as the "solution to the problem of what is occurring in the world that might yield this unusual sequence of stimulation" (Rock 1983, p. 14 and Chapter 7). From our point of view, finding the meaning is not necessary, the motion illusion results from the least complex description of the scene in terms of generative elements and their transformations, i.e. objects and their trajectories.

For example, consider a sequence of movie frames with a flying ball in an

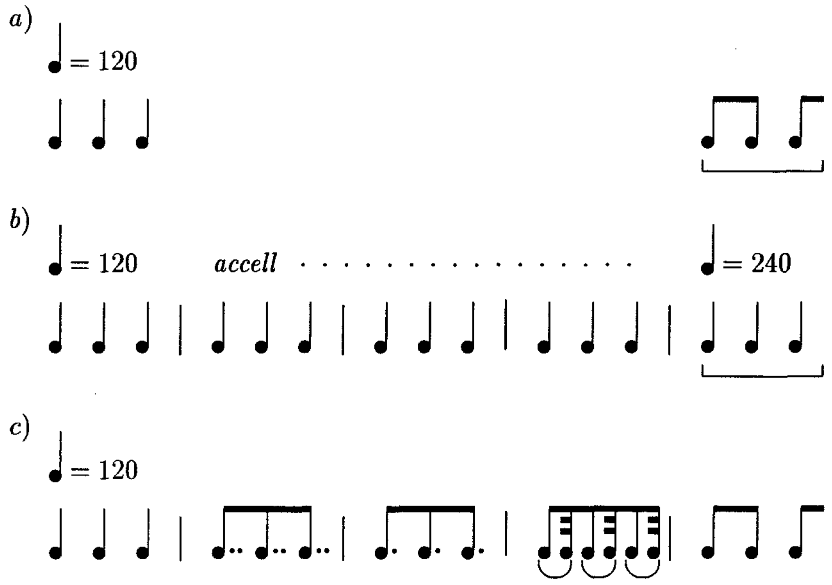

Figure 2.6: Contextual similarity of rhythmic patterns

invariable background shown in Fig. 2.7. From our standpoint, the apparent motion arises because of reducing the total visual information to its representation in terms of generative elements and their transformations (objects ant trajectories). Such an explanation doesn't use any assumption for "active perception", i.e. that "perceptual processing is guided by the effort or search to interpret proximal stimulus ... and the motivation for it must be the result of evolutionary adaptation" (Rock 1983, p. 16).

Note that the above effects of functional similarity, apparent motion, and ambiguity of perception, i.e. its dependence on the context, are explained with no reference to the meaning of percepts. We do not use any special constructions like conceptual frames (Minsky 1975) or meaningful settings (Palmer 1975). In other words, we provide for an explanation of percept dependence on the context which is *internal* with respect to perception, being based on no external cues (certainly, our explanation doesn't exclude others).

This pseudo-semantic self-organization of data is the most important feature of the correlative perception.

2.3 Model of Correlative Perception

Since we are looking for a representation of data in terms of variations of some generative elements, in other words, similar submessages, it is natural to apply methods of correlation analysis of data. However, since we are trying to recognize the transformations of generative elements, we have to use the correlation analysis under various deformations of data.

For example, to recognize a moving object in a cinema sequence as a more or less stable submessage, we can apply correlation analysis to distorted instant images. If these instant images are appropriately shifted, turned, stretched, etc., then the instant states of the object will correlate. In order to illustrate this idea, consider the problem of describing a dynamical scene at a computer display in terms of objects and their trajectories.

Imagine a flying ball in an invariable background. Identify the total visual information with a series of instant images, or with a cinematic sequence, wherein each frame differs from others only in the location of the ball (Fig. 2.7). In dynamics, a moving object is associated with a group of pixels which "move" according to a common law of motion (which have a "common fate"). This common law of motion is perceived as the object trajectory.

In order to separate the pattern of the ball from the pattern of the background, it is necessary to compare successive frames and to discover similar groups of pixels which are shifted with respect to each other. The pixels associated with the background are united by their immobility. Obviously, the groups of pixels associated with the ball are correlated in the statistical sense. If the ball's motion is complex and includes, for example, rotation, motion towards or away from the plane of the computer screen, etc., then the correlation model must provide for possible deformations of images, corresponding to the rules of motion in perspective.

Concerning the image transformations, it must be taken into account that every two images can be brought into correlation by some transformation. Hence there arises a risk of recognizing random configurations as intermediate states of the same object. Fortunately, such transformations are usually very complex. For this reason the model is provided with the criterion of least complexity, which rejects data descriptions which are too complex.

Thus, to describe a dynamical scene in terms of objects and their trajectories, it is necessary to analyze successive images and to discover pixel groups which are correlated under not very complex transformations. Next, among all such representations there must be found the one which is least complex. This scheme of calculations constitutes the *model of correlative perception.*

In statics, as contrasted with dynamics, the trajectory is replaced by the contour which is drawn by some generative element (as in Fig. 2.1). In general, everything said about trajectories is valid for contours with the only exception that correlation analysis is applied not to successive images but to the same image; in which case it is called autocorrelation analysis.

Note that the notion of contour is intended in a broad sense: A contour need not be continuous, closed, or even unidimensional. In this sense, the

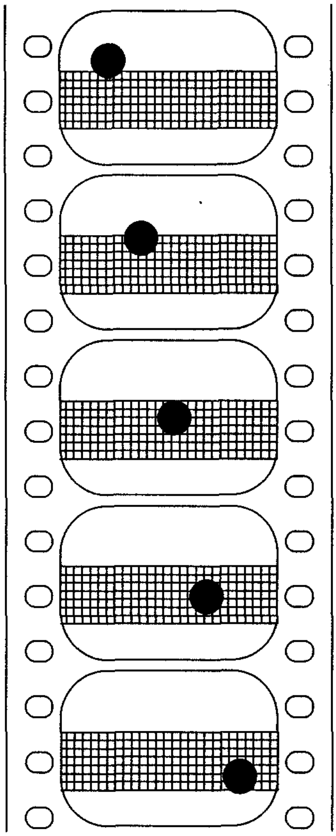

Figure 2.7: Flying ball in a cinema sequence

translations of A in Fig. 2.1 generate the contour of B, which is neither continuous, nor unidimensional. If instead of B we considered C, the outcome would be open (not closed).

2.4 Method of Variable Resolution

An implementation of the model of correlative perception requires considerable computing resources. The correlation analysis itself is rather slow and moreover it has to be performed under various transformations of data. Therefore, it is impossible to realize the correlation analysis, trying all possible distortions.

To end of realizing a directional search for the deformations of images (messages) which provides their high correlation (or that of certain submessages), a *method of variable resolution* is proposed.

The idea of the method is as follows:

- First, the resolution of the images is reduced in order to make the effect of small transformations negligible;
- Next, the correlated elements must be found;
- After the correlated elements have been discovered, the resolution is restored gradually while *locally* adjusting (distorting) the images in order to maintain the correlation between the known elements. These distortions correspond to the image transformations which provide the required high correlation.

The operation of the method can be traced in the following example. Consider two similar configurations of pixels in Fig. 2.8a–b coded by 1s on a background of dots (zeros). These pixels may be thought of as vertices of two squares of different sizes. Since the squares are unequal, the correlation function has no salient peaks. Indeed, under displacements and rotations, no pair of 1s can be superimposed between Fig. 2.8a and Fig. 2.8b. However, after a reduction in resolution, as in Fig. 2.8c and Fig. 2.8d, the two images are correlated. Indeed, some of the 1s can be superimposed; they are shown by frames. After the correlated elements have been discovered, the correlation can be improved by locally distorting neighborhoods of the correlated elements. Having determined the distortions which provide the highest correlation of the two images, one obtains the required deformation which transforms one original image into another, i.e. under which the two images in Fig. 2.8a–b are most correlated. If necessary, reducing and restoring the resolution can be realized gradually, in several steps.

One can see that the principal advantage of the method is that the search for *global distortions* is reduced to *local adjustments*. Note that reducing the

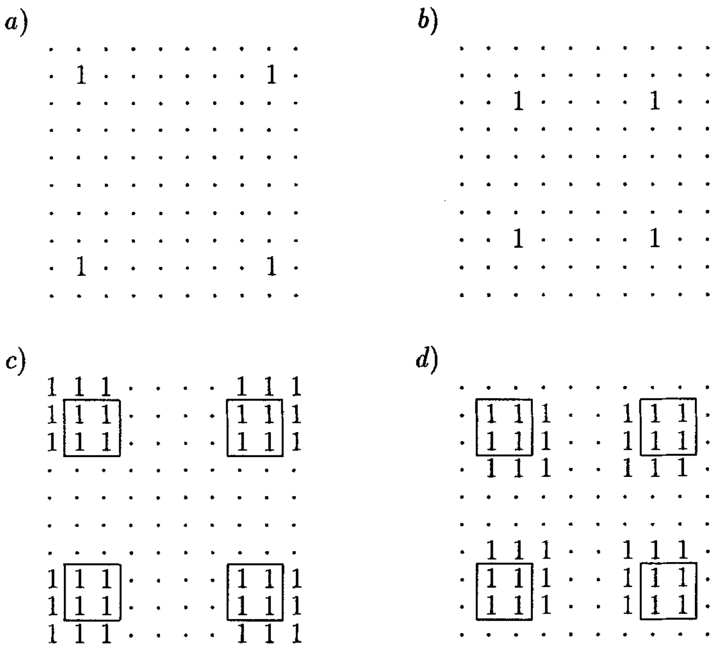

Figure 2.8: Illustration to the method of variable resolution
(dots denote zeros)

resolution (filtering) is usual in recognizing similarity; e.g. see (Palmer 1983; Witkin 1983; Bouman & Liu 1991). On the other hand, note the similarity of the method of variable resolution to the operational principle used in perceptrons (Minsky & Papert 1988) and pyramidal data structures (Hummel 1987).

However, we go further, providing the filtering and correlation schemes with a kind of a double feedback which guides the directional search for the image deformations required. It is realized by applying the correlation criterion which reveals the deformations which provide the highest correlation of successive images. At the same time, the complexity criterion sorts out those deformations which are too complex.

Note that both correlation analysis and method of variable resolution are realizable on neuron nets with parallel processing. Therefore, the method can be quite efficient on a specially designed hardware.

Since neuron nets are usually considered as models of the brain (Rossing 1990, p. 164), the observation mentioned is also compatible with our hypothesis about the existence of the related mechanism in human cognition.

2.5 Complexity of Transformation as a Distance

There are several definitions of distance between patterns. The distance which can be defined with respect to characteristics of the mapping which transforms one pattern into another is of particular interest for our purposes. The *translational distance* and *transformational distance* are discussed by Palmer (1983, p. 289); *distortion measure* is used by Rangarajan & Shah (1991); different metrics are enumerated by Witkin & Tonenbaum (1983b, p. 517).

According to our complexity approach, it is natural to characterize the distance between two patterns in terms of complexity of the deformation which transforms one pattern into another. Such a distance has the same properties as the metrical distance:

- **(Positivity)** The complexity of deformation is equal to 0 if and only if two patterns are identical, otherwise it is positive.

  Indeed, if two patterns are equal, the deformation required is trivial, implying its complexity to be equal to zero. If two patterns are dissimilar, the deformation required is significant, implying its complexity to be significant too.

- **(Symmetry)** The complexity of deformation (pointwise defined) equals to the complexity of the inverse deformation.

Indeed, if we consider the deformation of patterns as a pointwise correspondence, the complexity of the inverse transformation requires the same memory storage as the direct one.

- **(Triangle inequality)** The complexity of a superposition of two deformations (the sum of two complexities) is less than or equal to the complexity of the resulting direct deformation.

The complexity of deformation is well adapted to estimating the dissimilarity of patterns. For example, two patterns may be identical but remote in the Euclidean space. Since the complexity of deformation is small, their similarity is recognizable. Any distance based on absolute scales will fail in recognizing such an identity.

Moreover, learning can change the "distance" between patterns, providing some transformations with standard descriptions (e.g. rotations). Even if complex themselves, the reference complexity of these standard transformations can be small since their use requires calling related algorithms with a few parameters (cf. the call **R012** in Fig. 2.4).

In a sense, the complexity of deformation is complementary to the correlation coefficient, being a measure of dissimilarity instead of the measure of identity. However, using these two concepts is not equivalent. Indeed, two objects may be similar but not identical. In such a case the correlation analysis fails in recognizing similarity (because there may be no identity at all, as in Fig. 2.8a–b), whereas the complexity of the deformation, being small, indicates at the similarity (as recognized in Fig. 2.8d).

Note that the two complementary concepts, the correlation between two objects and the complexity of transformation of one into another, correspond to the two levels of our model of data representation: At the first level we have correlating generative elements, and the second level is reserved for the description of their deformations.

2.6 Distinctions of the Model

The need for an axiomatic theory of perception was realized while developing models of perceptual organization (Rock 1983, pp. 328–335; Palmer 1983; Leyton 1986).

In a sense, we develop an axiomatic approach (not a theory yet) to perception modeling. Postulating the self-organization capacity of perception aimed at data reduction, we attempt to derive its much less evident properties like: The capacity to segregate patterns, to arrange patterns into hierarchies, to recognize the causality in pattern generation, etc. While investigating hierarchical representations of data, we pose the following questions:

- Why a hierarchy?
- Which hierarchy? and
- How does the hierarchy correspond to the reality?

From the standpoint of our approach to minimizing the complexity of representations, the answers to these questions are, respectively:

- The hierarchization makes data representations compact.
- Consequently, a better hierarchy is the one which makes the data representation least complex.
- Under certain assumptions such a hierarchy reveals perception patterns and causal relationships in the data, making the first step towards their semantical description.

In our model, the self-organization capacity of perception is characterized in terms of optimal data representation but not in terms of recognition capabilities. We argue that the pattern recognition is determined by the criterion of least complexity, whereas most of recognition systems are based on threshold criteria which are adjusted at the learning phase.

Next, we elaborate the approach to separating patterns with respect to their *similarity*. Usually, the recognition of patterns is based on their classification with respect to their dissimilarity which is fixed by threshold criteria. Since we don't recognize dissimilarity, we also don't need thresholds for pattern separation.

Comparing the two approaches to pattern recognition, ours and the one based on threshold criteria, we see that the latter has two principal disadvantages:

- It requires a learning stage in order to determine thresholds,
- thresholds, being determined, make the recognition hardly adaptable to new circumstances (or additional learning is needed).

Unlike threshold criteria, the criterion of least complexity is self-adjustable to current circumstances. It guides the self-organization of data, not requiring any learning stage. This may be important in ambiguous cases where threshold criteria can fail. For instance, it is difficult to incorporate a threshold criterion in example illustrated by Fig. 2.4 in order to judge whether a sequence of durations should be interpreted as a single or repetitive pattern.

An important distinction of our approach is the way of recognizing the similarity. Usually, the measure of similarity is estimated by the degree of identity; this is the main idea of correlation analysis. On the contrary, we measure the

Table 2.2: Artificial intelligence and artificial perception in pattern recognition

	Artificial perception	Artificial intelligence
Function	Object segregation	Object identification
Performance	Data representation	Knowledge representation
Principle	Self-organization	Classification
Means	Data processing	Learning
Cues	Similarity	Dissimilarity
Criterion	Optimal representation	Threshold adjustment
Measure of similarity	Complexity of transformation	Correlation

difference between the objects by the degree of dissimilarity measured by the complexity of the deformation which is necessary to make objects identical. Such an approach is useful in cases when there is a similarity but there is no identity.

The enumerated features distinguish our artificial perception approach to pattern recognition from that of artificial intelligence which is traditionally based on learning, representation of knowledge, and classification of patterns. The complementarity of artificial intelligence and artificial perception in pattern recognition, as understood in our model, is shown in table 2.2.

2.7 Summary of the Chapter

Thus we enumerate the main distinctions of the proposed approach to perception modeling.

1. We consider a perception mechanism, the correlativity of perception, which is the interaction of two principles of data self-organization, "common fate" principle and simplicity principle. In our model these two grouping mechanisms control each other, resulting in a new quality of grouping.

2. The "common fate" principle is modeled in terms of generative elements and their transformations, corresponding to stable configurations of stimuli and relationships between them. Thus the structure is recognized with respect to replications but not with respect to dissimilarity.

3. The simplicity principle is formalized by the Kolmogorov criterion of optimal data representation (least memory storage required). By constructing optimal representations of data we attempt to segregate patterns and to reveal the causality in the data generation.

4. The search for stable configurations of stimuli is realized as finding similar messages in data arrays by means of correlation and autocorrelation analysis applied to the data arrays under their various distortions.

5. In order to realize the directional search for similar configurations of stimuli, the method of variable resolution is proposed. At first it reveals the similarity in general, and then localizes the search for fine matching.

6. The distance between two patterns is defined to be the measure of complexity of the transformation of one pattern into another. This measure can be modified while using some standard transformations (shifts, rotations, etc.) which are not to be described each time but simply coded with their parameters.