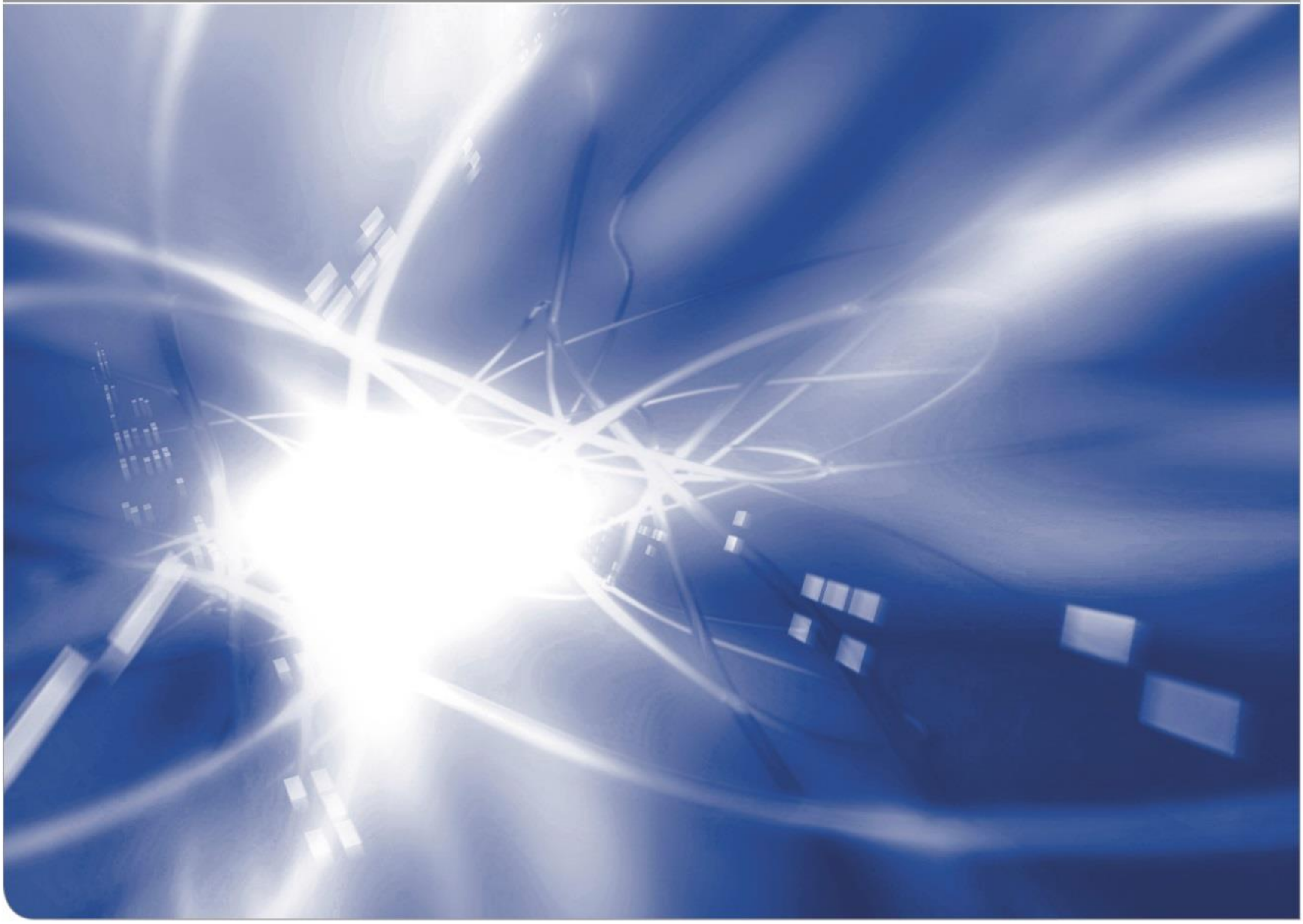
HOW DO WE THINK

Modeling Interactions of Perception and Memory

by Andranik S. Tangian[1]

KIT SCIENTIFIC WORKING PAPERS 166



[1] Institute of Economic Theory and Operations Research

KIT-ECON
Blücherstraße 17
76185 Karlsruhe
Germany
Tel: +49 721 6084 3077
micro.office@econ.kit.edu
andranik.tangian@kit.edu; andranik.tangian@gmail.com

Impressum

Karlsruher Institut für Technologie (KIT)
www.kit.edu

2021

ISSN: 2194-1629

Institute of Economic Theory and Operations Research
Karlsruhe Institute of Technology

HOW DO WE THINK: Modeling Interactions of Perception and Memory[1]

Andranik S. Tangian

Scientific paper Nr. 166

May 2021

E-mail: andranik.tangian@kit.edu
andranik.tangian@gmail.com
Tel: +49 721 6084 3077

Blücherstraße 17 76185 Karlsruhe Deutschland

[1] This paper is the second and improved edition of my 2001 article 'How do we think: Modeling interactions of memory and thinking', *Cognitive Processing*, 2(1), pp. 117–151, which became unavailable on the internet after the journal has been taken over by Springer in 2004 with no copyright for the first four volumes.

Abstract

A model of artificial perception based on self-organizing data into hierarchical structures is generalized to abstract thinking. This approach is illustrated using a two-level perception model, which is justified theoretically and tested empirically. The model can be extended to an arbitrary number of levels, with abstract concepts being understood as patterns of stable relationships between data aggregates of high representation levels.

Keywords: perception, memory, cognition, abstract thinking, data representation, pattern recognition.

Preface to the second edition of the article

The best idea is the simplest one.
Business saying

Advances in artificial intelligence and related disciplines had a great impact on cognition research [Grenander 2012, Grenander and Miller 2007, Kurzweil 2012, Minsky 2006, Mumford and Desolneux 2010]. The developments based on pattern learning, pattern recognition and inference rules have numerous practical applications with outstanding performance and even more promising prospects [European Commission 2020]. At the same time, sophisticated 'intelligent' models pose questions about their relevance to human cognition. For instance, algebraic constructs to logically derive possible conclusions, being an efficient instrument of artificial intelligence, look too artificial for the brain. Since logic is not a model but an invention of mind, putting it at the heart of thinking seems to be too a technie approach. As Niels Bohr once said:

> No, no, you are not thinking, you are just being logical.
>
> — Frisch O.R. (1979) *What Little I Remember*, p. 95.

The article, now re-edited, was based on my two lectures. The first one was read on March 19, 1999, at the *Jahrestagung der Gesselschaft für Biophysik* (Annual Conference of the Biophysical Society), University of Tübingen, on the invitation of Prof. Alexander Borst who looked for models of enigmatic information exchanges between the brain neurons. The second lecture took place on October 12, 1999, at the Department of Psychology of the University of Rome *La Sapienza* on the invitation of Prof. Marta Olivetti-Belardinelli who was interested in some peculiarities of music perception. These lectures articulated some general ideas on cognition scattered throughout my rather technical book *Artificial Perception and Music Recognition* [Tanguiane 1993], which stemmed from two observations. The first one is the effect of a musical band produced by a loudspeaker membrane. Instead of reflecting the physical reality — the single vibrating body — perception *reconstructs* the way the signal was generated by several instruments. An opposite effect is produced by organ registers that activate several differently tuned pipes by pressing one key. Then playing a melody results in a sequence of chords actually perceived as a single voice. Again, instead of reflecting the physical reality — several vibrating pipes — perception reconstructs the way the sound sequence is generated by one-finger playing. In both examples, perception deals with *data representations*, losing sight of physical reality but recognizing the causality in the signal generation.

These examples are not curious paradoxes of perception but illustrations of the brain's regular operation. The brain, receiving signals from the eardrum, which is a single physical body, separates or fuses the sound components at its 'own discretion'. It should be emphasized that no pattern recognition is performed at this early stage. Indeed, if the brain first focused on pattern recognition then the polyphony performed on unknown instruments would be imperceptible — because their sounds could not be identified. On the other hand, the sounds of organ register pipes would not fuse, because their known sounds would be first recognized and thereby separated. All of these argue for the brain's capability to somehow find structures in data with no pattern recognition and without knowing the structures.

The structural perception is studied in Gestalt psychology. In particular, it suggests several laws of perceptual grouping: proximity, similarity, common-fate, past experience and some others, which in the given article are considered derivatives from the simplicity principle. According to [Kolmogorov 1965], the data complexity is measured by the amount of memory needed for the *algorithm* (= rule) of the data generation. Since saving the totality of data requires much memory, compact (= simple) algorithms that *reconstruct* the data are preferable, in line with

> The knowledge of certain principles easily compensates the lack of knowledge of certain facts.
>
> — Helvétius C.A. (1758) *Essays on the Mind and Its Several Faculties.* Essay III, Ch. 1, p. 198.

Obviously, compact data representations save the brain resources and energy. For example, phone number 0123 456 789 is remembered as a progression of numerals, 0123 123 123 — as 0 and 123 repeated three times, etc. The laws of Gestalt psychology are such memory-saving algorithms that result in multi-level data representations.

While constructing new data representations, the algorithms from past experiences, even not simplest, can be efficiently used because they already reside in the memory and require no additional storage. For instance, no new generative rule for the phone number 0123 114 125 is necessary because it is a minor variation of the already known 0123 124 125. The choice between optimally arranged direct percepts and patterns/algorithms from the memory opens a vast field for 'thinking possibilities' up to developing abstract concepts (= high-level patterns/algorithms), which are necessary to perform complex tasks. One can even imagine applications of this artificial perception approach to economic statistics, astrophysical data, etc., expanding the scope of hearing and vision to the areas that are foreign to human senses and human perception.

Thus, we suppose that the brain is guided by its own motivation to minimize memory resources and save energy. For this purpose, it develops ways to pack the information received as compact as possible. Why should this lead to 'world understanding' and revealing the physical causality in observations? Our point is the duality of the naturally optimized material world and its optimal descriptions. The subordination to universal optimization principles determines the nature–information coherence, enabling adequate reasoning solely by means of optimal data processing.

Therefore, cognition is not only learning. The information received is also a trigger that activates cognition and high-level thinking, which run by themselves. Such gaining knowledge 'from nothing' echoes Plato's recollection theory from his dialogues *Meno* (Virtue) and *Phaedo* (Soul), which explains knowledge as remembrance of what the soul has learned in the immaterial world of ideas.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Memory and thinking

In the mid-1980s, I attended a full-length concert of stage performer Youri Gorny. He has demonstrated a phenomenal memory by telling numbers from a telephone directory, memorizing 40-decimal numbers at a glance, or making complex arithmetic operations in mind. In the second part of the show, he has answered questions from the audience. I have expected an interesting conference but his opinions and judgements were neither ingenious nor witty. I got a strange impression of an oldish teenager.

This experience suggested an idea of complementarity of memory and thinking. Probably, an unusual redistribution of brain resources has resulted in an extraordinary memory at the cost of analytical capabilities.

It is known that at the age of about 12 years there is a change in the way of thinking. Children remember everything but do not analyze. Repeating literally is easier than saying in own words. They can learn a poem by heart without understanding the meaning. It is quite different in adults. To remember, they need to understand. They can hardly repeat a sentence in a language they do not know. Instead of replicating word for word, they rather reproduce the sense.

It looks as if by the age of 12 the brain is getting saturated by the incoming information and starts to overcome the shortage of memory. It reduces the amount of data by extracting their most essential features by omitting secondary details and develops schemata to link the new data to what is already stored in the memory.

Instead of unconsciously memorizing data, the brain turns to deriving general rules from particular experiences, shaping patterns of thinking, and forming abstract concepts. A gain in intelligence compensates a decreasing ability to learn foreign languages, master musical instruments, and advance in sensory-motor activities like sport games. In a sense, memory is replaced by mind. An extraordinary memory does not urge this process and can slow down the intellectual development.

In other words, the brain overcomes the shortage of memory by finding new ways of data representation. It requires analytical thinking: Data should be analyzed to be adequately represented. This mechanism is inherent in everybody at every age, differing only in the degree of its development.

Which advantages has the process of replacing memory by mind?

- First of all, it prevents memory from being saturated and overburdened by redundant information. Thus memory storage is saved.

- Next, it supports data classification and information search by analogy and other vague cues. In other words, an efficient data access is provided.

- Thirdly, the problem-solving ability is trained due to the practice in distinguishing between important and unimportant, as well as in discovering trends and causal relationships.

1.2 Thinking and compact data representations

Thus a shortage of memory turns to be an advantage. It stimulates analytical thinking and developing abstract concepts (= 'building models'). What is the way of thinking and how does it develop? In the given paper, I attempt to extend the studies in artificial perception [Tanguiane 1993, Tanguiane 1994, Tanguiane 1995, Tangian 1998] to general principles of thinking.

Methodologically, the approach proposed is based on several assumptions.

- *Hierarchical representation of data.* Most information is carried by another type of information. For instance, a radio wave is a carrier of an acoustical signal, the acoustical signal is a carrier of words, and words are carriers of the meaning. Thus, the first principle is data grouping and arranging them into a hierarchy.
- *Least complexity in representing data.* The complexity is understood in the sense of [Kolmogorov 1965] — as the amount of memory storage required for a data representation; see also [Calude 1988]. Data compression is generally based on eliminating repetitiveness. Since deviations from repetitiveness (like modulation of a periodical radio wave), constitute information of the next level (acoustical signal), reducing data redundancy provides a transition to the next information level. Thus the second principle is the simplicity of data representations leading to the data understanding.
- *Duality of optimality in the nature and in its descriptions.* The optimality of the nature (the law of minimal energy in mechanical and electrical systems, etc.) is assumed to correspond to its optimal (minimal) descriptions. Thus, our fundamental conjecture is: *The simplest representations of data reveal the causality in the data generation.*

1.3 Compact representations and causality

Recognizing the causality in data is different from recognizing their physical source. For example, a single physical body (loudspeaker), can produce an impression of many sources (large orchestra). If one could recognize the source, (s)he would perceive a single loudspeaker membrane rather than numerous orchestral instruments.

The cause of the orchestral sound is however neither the loudspeaker nor orchestral instruments but instrumental gestures. Each instrumental gesture causes a number of overtones with a 'common fate', which unites them into an acoustical aggregate, and this aggregate is distinguishable in the mixed sound. The classification of partial tones with respect to their 'common fate' reveals the common cause of their generation.

Data are usually self-organized with a reference to general principles of grouping and simplicity known since the *Gestalt psychology* [Wertheimer 1923]. According to [Gibson 1950, Gibson 1966, Gibson 1979], structures are usually data-driven and require no *a priori* knowledge to be found.

The theoretical non-uniqueness of data representation leaves a room for alternative representations, which are sometimes regarded as paradoxes of perception. A paradoxical perception can be understood as a data structuralization in a way different from the data generation.

Structuring without knowing the structure is like distinguishing objects in abstract painting. An observer can segregate objects but may have difficulties in recognizing (= identifying) them explicitly. On the other hand, one can group certain elements into objects differently than others. The fact that most observers perceive structures in the same way can be explained by the uniqueness of optimal (= simplest) representation.

Identifying objects by labeling is the next step in data understanding. (Throughout the paper, an *object* is an independent data block in vision, hearing and thinking.) It requires learning, determining standards and object matching. Therefore, object recognition falls into two steps:

- 'non-intelligent' perception, that is, data structuring by self-organization with no *a priori* knowledge;
- 'intelligent' labeling, that is, identification of separate structural elements by their matching to standards, which have been determined by learning from past experiences.

The 'non-intelligent' perception backs up the 'intelligence': Matching separate objects is easier than recognizing objects in data streams. In the sequel we illustrate these ideas using a model for recognizing polyphonic voices and rhythm/tempo.

1.4 From perception to abstract thinking

Since thinking deals with data representations anyway, it can process the data aggregates of an arbitrary high level exactly in the same way as the aggregates of primary stimuli. Respectively, structuring data can be performed not only at the level of stimuli but at any level of abstraction.

Creating new abstract concepts optimizes the knowledge representation, making thinking more efficient, less context-dependent and less specific. The knowledge representation is getting free from unimportant details and becomes less memory-consuming.

The hierarchy of thinking, from primary stimuli to abstraction, is homogeneous in the following sense: It is based on the same principles as the first two perception levels — grouping with respect to similarity and saving memory. Moreover, the optimization of data representations is supposed to be the *internal motivation* of thinking.

Thus, the difference between naive perception and abstract thinking is rather quantitative than qualitative:

- Perception operates on hierarchical data representations with a few levels whereas thinking — on that with multiple levels.
- The top levels of these representations correspond to elementary structures of primary stimuli (= percepts) or advanced structures derived from observations (= abstract concepts), respectively.
- Saving memory by optimal data representations is supposed to be the internal motivation of thinking.

1.5 Content of the paper

In Section 2, 'Data representation and correlative perception', we consider two ways of data storage, direct coding and analytical representations. The latter fall into structural ones, which are data-driven and use general algorithms, and knowledge-based ones that require target algorithms stored in the memory.

Structural representations are obtained by hierarchical grouping guided by the simplicity principle. The tight interaction of the grouping and simplicity principles is called *correlative perception*. The data are represented as transformations of a few generative elements. To find such representations, a so-called *method of variable resolution* is proposed.

In Section 3, 'Evidence of correlative perception', the problems of musical chord recognition and voice separation are considered. A polyphonic voice is recognized as a dynamic trajectory drawn by a tone spectrum. The latter is a group of partials with a 'common fate', i.e. that move in parallel with respect to the $\log_2$-scaled frequency axis. A chord is recognized as a static contour drawn by a tone spectrum shifted several times along the frequency axis.

The computer experiments show the same recognition reliability as demonstrated by trained musicians. In addition to empirical tests, a mathematical justification of the model of correlative perception is provided: It is proved that chords can be correctly recognized by optimally representing their spectral data. Finally, we consider three applications to psychoacoustics and music theory: rhythm–tempo interaction, functional definition of interval hearing, and reasons for the prohibition of parallel voice-leading in counterpoint and harmony.

In Section 4, 'Generalization to abstract thinking', it is suggested (a) to extend the model of correlative perception from two to multiple levels and (b) to interface it to a knowledge base. This way the incoming data are optimally represented using both the direct percepts and the patterns from the memory. Furthermore, such an active self-organization of knowledge leads to developing abstract concepts that reflect stable relationships between the data aggregates.

In Section 5, 'Summary', the main statements of the paper are recapitulated.

2 Data representation and correlative perception

2.1 Two types of data storage

At first we make terminological conventions. To be specific, let us consider an example.

Example 1 (Structural representation) *Let us code a sequence of time events coded by 1s and 0s (= beat and offbeat, respectively):*

$$\text{DATA 1} = \{\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ \}\ . \tag{1}$$

Taking into account the data periodicity, we represent them as follows:

$$\text{DATA 1} = \{\ \underbrace{1\ 0}_{\times 8}\}\ . \tag{2}$$

We say that the given data are *memorized* or *directly coded* if all their instances are stored in full like in (1). In contrast to that, the data are *represented analytically* if they are stored using a generative algorithm like in (2). The both types of data storage are compared in Table 1.

Table 1: Two types of data storage

	Direct coding	Analytical representation
Amount of memory storage	Large	Small
Loss of information	No	Possible
Data understanding	No	Yes

Analytical representations can be approximate. For example, (2) can result from approximating the following sequence of time events with accents

$$\text{DATA } 2 = \{\ 2\ 0\ 1\ 0\ 2\ 0\ 1\ 0\ 2\ 0\ 1\ 0\ 2\ 0\ 1\ 0\ \}\ . \qquad (3)$$

In spite of information losses, (2) is nevertheless useful as revealing the data periodicity. If necessary, the comment 'odd beats are strongly accentuated' can be added, or the generative element { 1 0 } can be replaced by a more complex one { 2 0 1 0 }.

Generative algorithms save memory and highlight the data structure, providing a kind of data understanding. Additional details or more complicated generative algorithms burden analytical representations, sometimes making them not shorter as direct coding. In most cases, approximate representations, reflecting most important features of the data, are quite sufficient. For instance, the approximation (2) of rhythm (3) can satisfy a dancer who only cares about the difference between two step and three step.

Below we show how analytical representations are used to recognize chords.

2.2 Two types of analytical representations

The data in (1) are redundant because of repetitions. In some cases, data redundancy is less evident.

Example 2 (Knowledge-based representation)

$$\text{DATA } 3 = \{\ 3\ 1\ 4\ 1\ 5\ 9\ 2\ 6\ 5\ 3\ 5\ 8\ \ldots\} \qquad (4)$$

can be compactly represented using a generative algorithm for computing decimals of π. In the given case, such a representation requires the following:

- *knowledge of an algorithm to compute π,*
- *matching algorithm for recognizing the relation of the data to π,*
- *standard appearance of the data: If (4) started with the 100th decimal of π then the relevance to π would not be easily recognizable (cf. with perceiving an object from an unusual viewpoint), and*

- *standard appearance of the code: For instance, the sequence of notes*

  $$e\flat, c\sharp, e, c\sharp, f, a, d, f\sharp, f, e\flat, f, a\flat, \ldots$$

  would hardly be associated with π, although (4) are their pitch codes.

A representation is *structural* if its generative algorithm uses repetitions and transformations of certain *generative elements*. For example, the periodical sequence (1) is structurally represented as a repetitive rhythm (2) generated by rhythmic pattern $\{1\ 0\}$.

Since recognizing repetitions requires no knowledge, this type of representation corresponds to 'naive' (unconditioned) perception. A structural representation is the first step in visual and audio scene analysis (the repetitiveness of time events 'explains' their generation).

A representation is *knowledge-based* if its generative algorithm uses references to certain known patterns stored in the memory. For example, the non-periodical sequence (4) is representable with a reference to π. This type of representation is associated with intelligence. The knowledge is usually required after the structuring has been done. For instance, the rhythm (1) can be recognized as repetitive and then interpreted as a march, or a painted object can be structured and then identified with a face.

Structural and knowledge-based representation complement each other. On the one hand, structuring simplifies identification: Identifying a separate object is easier than recognizing it in a data flow. For instance, it is easier to watch cartoons than feature films due to their already conceptualized images with fewer details and a simple background. On the other hand, direct recognition contributes to structuring. For instance, having identified some elements of an abstract picture as conditional two eyes, one can continue recognition and then find noise, mouth, hair, etc.

2.3 Hierarchical representations and grouping principle

By *hierarchical grouping* we mean the ability of perception to group stimuli and group the groups of stimuli, that is, perceive patterns of relationships between the groups. The groups of stimuli and relationships between them are called *low-level* and *high-level patterns*, respectively.

Example 3 (Low-level and high-level patterns in vision) *The pixels (stimuli) in Figure 1 are grouped into characters A or symbols* Π *(low-level patterns), which in turn are grouped into the contour of B (high-level pattern).*

Data representations that use hierarchical grouping have three advantages:

Memory saving: Instead of storing all the pixels, it is more efficient to store one pixel configuration for one symbol and then to store the contour of B.

Data understanding through structuring: Since a persistent repetition hardly occurs by chance alone, it is rather a manifestation of an underlying regularity. Therefore, the data structure 'explains' the data generation.

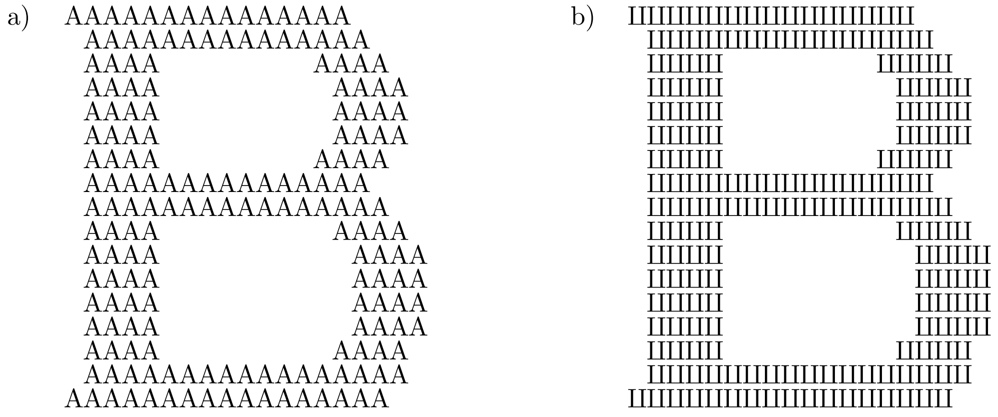


Figure 1: High-level pattern of B composed by (a) low-level patterns of A or (b) II

High-level pattern independence of low-level patterns: High-level patterns are invariant with respect to changes of their carriers — low-level patterns. Even if unknown (unrecognizable) symbols are used instead of A's, it is still possible to recognize B by the relationships between the unknown symbols II.

Thus, we represent data in terms of repetitive (correlating) messages (e.g. small A's) whose interrelations determine high-level patterns (e.g. B). In other words, we have *generative elements* (A or II) and their transformations (displacements that generate B). Here, low-level patterns are information carriers for high-level patterns.

Grouping is usually based on certain common features of elements or their common behavior ('common fate' principle). In Example 3, the common feature appears as similar pixel configurations (correlation) of characters A. Similar elements are recognizable also in data flows: If the background of Figure 1 were filled with various characters then the contour of B would be still distinct due to the persistent repetitions of A's and II's.

The perception capability to find structures in data without knowing the structures is noted by [Witkin and Tenenbaum 1983]:

> People's ability to perceive structure in images exists apart from the perception of tri-dimensionality and from the recognition of familiar objects. That is, we organize the data even when we have no idea what it is we are organizing. What is remarkable is the degree to which such naively perceived structure survives more or less intact once a semantic context is established: the naive observer often sees essentially the same things an expert does, the difference between naive and informed perception amounting to little more than labeling the perception primitives. It is almost as if the visual system has some basis for guessing *what* is important without knowing *why*. . .
>
> . . . The aim of perceptual organization is the discovery and description of spatio-temporal coherence and regularity. Because regular structural relationships are extremely unlikely to arise by the chance configuration of independent elements, such structure, when observed, almost certainly denotes some underlying unified cause or process. A description that decomposes the image into constituents that capture regularity or coherence therefore provides

descriptive chunks that act as ‘semantic precursors,’ in the sense that they deserve or demand explanations.

— Witkin A.P. and Tenenbaum J.M. (1983) On the role of structure in vision, pp. 482–483

2.4 Optimal representations and simplicity principle

The grouping principle provides certain advantages in analytically representing and understanding data. However, grouping can be made in different ways. To overcome the grouping ambiguity, we apply the simplicity principle: From all possible groupings, the least complex one is selected. The *complexity* is understood in the sense of Kolmogorov, that is, as the amount of memory storage required for the algorithm of the data generation [Kolmogorov 1965, Calude 1988]. The next example illustrates the interaction between the grouping and simplicity principles.

Example 4 (Overcoming the logical loop of definitions of rhythm and tempo) *In existing literature, rhythm is defined with respect to a certain tempo, and tempo assumes a certain rhythm. Under these definitions, any sequence of time events can be regarded either as (a) a sequence of (unequal) durations under a constant tempo or (b) a sequence of equal durations under varying tempo or (c) unequal durations under varying tempo in an infinite number of ways.*

However, most listeners perceive rhythm and tempo in the same way, which can be explained in terms of data representation complexity. Each hierarchical representation of time events has some generative low-level pattern — a repetitive rhythmic figure, and the high-level pattern is some tempo curve, which reflects the augmentations and diminutions in time of the rhythmic pattern. The complexity of this representation is split between the rhythmic pattern and the tempo curve. The optimal representation with the least total complexity is selected, which corresponds to the rhythm/tempo perception. In most cases such a representation us unique.

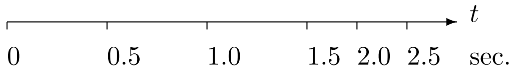
As shown in Table 2, repetitive rhythmic patterns are analogous to repetitive symbols A or П in Figure 1. The tempo curve is the contour of ‘time density’, which is determined by augmentations or diminutions in time of repetitive rhythmic patterns. It is analogous to the contour of B generated by displacements of a repetitive symbol.

Table 2: Analogy between visual and time data

	Visual data	Time data
Stimuli	Pixels	Time events
Low-level pattern	A or П	Rhythmic pattern
High-level pattern	B	Tempo curve

The context dependency of perception can also be explained in terms of data representation complexity.

Example 5 (Context dependency of rhythm and tempo perception) *Figure 2 displays a sequence of six time events. Most listeners perceive it as a single rhythmic pattern*



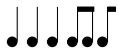
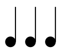
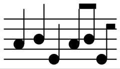
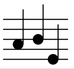
Representation of time events	*A*	*B*	*C*	*D*
		 **R012**		 **R012**
Complexity of rhythmic pattern	6	3	12	6
Complexity of its transformation	0	4	0	4
Total complexity	6	7	12	10

Figure 2: Complexity of representations of six time events, in number of bytes

(Representation A) rather than as the first three events repeated twice faster (Representation B). Here, **R012** *denotes the repetition algorithm call* **R** *with three parameters: return to time* **0***, play* **1** *time, perform* **2** *times faster). However, if the events have pitches as in Representation C then the repetitive melodic contour enhances the sensation of a rhythmic repetition (cf. with the recognizability of a fugue theme in diminution). This means that Representation D is preferred to Representation C.*

To explain the perception of the same rhythm as non-repetitive or repetitive, we estimate the complexity of these representations. We assume that each duration is coded by one byte, a duration with pitch — by two bytes, and the repetition algorithm call — by four bytes. For the rhythm alone, Representation A is less complex than B (6 bytes against 7), resulting in the perception of a single rhythmic pattern under a constant tempo. In the melodic context, Representation D is simpler than C (10 bytes against 12), resulting in the perception of repetition under a double tempo.

Thus, low-level rhythmic patterns are reference time units for tempo tracking, and the tempo curve is a high-level pattern of their transformations — augmentations and diminutions in time. The interdependence of rhythm and tempo is overcome thanks to the intervention of the criterion of least complexity (= simplicity principle), which enables to interpret time events in terms of both rhythm and tempo in their interaction.

2.5 Correlativity of perception

By *correlativity of perception* we mean the ability to hierarchically group percepts (data) in the simplest way. This is attained by endowing the grouping principle with a feedback that optimally controls the process of data representation to the end of minimizing its complexity. Among all possible structures, the correlative perception finds the optimal one, which is usually unique, resulting thereby in an *unambiguous* structural perception.

For instance, the 'audio scene' (cf. with [Bregman 1990]) from Example 5 is analyzed with no special knowledge. The decision about the time structure with or without repetition needs no learning or special constructs like 'conceptual frames' in the sense of [Minsky 1975] or 'meaningful settings' in the sense of [Palmer 1975]. In Example 5, the context dependency of perception is conditioned by no cues other than general criteria of similarity and simplicity.

To find similarities, one has to perform correlation analysis under various data transformations. The search for correlated data blocks under data transformations can be realized

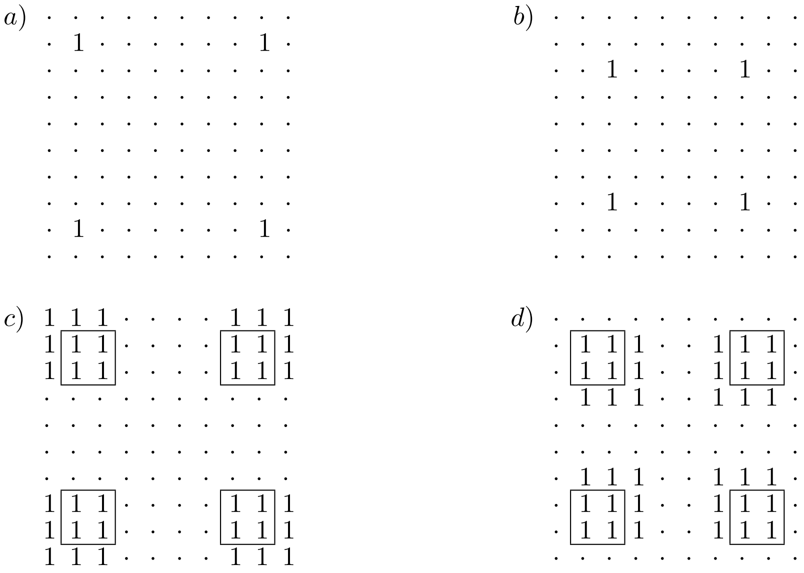


Figure 3: Illustration to the method of variable resolution

by the *method of variable resolution*, which we explain using the following example.

Example 6 (Method of variable resolution in vision) *Figures 3a–b display two apparently similar 2D configurations of 1s on the background of zeros denoted by dots. To establish their similarity, we have to find a* simple *transformation that makes them equal. For this purpose, configurations* a) *and* b) *are superimposed and the number of matching 1s is counted. Since the configurations are unequal, the matching 1s are too few (= no correlation). Then the 'resolution' of both images is reduced by replacing 1s by clusters of 1s as shown in Figures 3c–d. Now the framed areas in Figures 3c–d coincide (= correlate). Once the correlation is captured, the original resolution is gradually restored while keeping control over the areas correlated: If they lose some of the matching 1s (= correlation decreases) then these 1s are shifted to retain the correlation high. These local adjustments determine the image transformation required. In the given case, the transformation is simple, which confirms the similarity of configurations* a) *and* b)*.*

In the following example, the method of variable resolution is applied to one-dimensional time data in a more formal way.

Example 7 (Variable tempo tracking) *We consider the quasi-periodic sequence of time events*

$$s(n) = \underbrace{1000000000}_{10}\underbrace{100000000}_{9}\underbrace{10000000000}_{11}10 \tag{5}$$

whose autocorrelation

$$R_s(p) = \sum_n s(n-p)s(n)$$

returns the number of matching 1s under self-superimposing string s with shift (period) p. Let us apply the method of variable resolution step by step as traced in Table 3.

Table 3: Autocorrelation $R_s(p)$ of time events

$$
\begin{aligned}
s &= \underbrace{1\ 000000000}_{10}\underbrace{1\ 00000000}_{9}\underbrace{1\ 0\ 000000000}_{11}\ 1\ 0\\
s_1 &= \underbrace{\boxed{1}100000000}_{10}\underbrace{\boxed{1}10000000}_{9}1\underbrace{\boxed{1}000000000}_{11}\boxed{1}1\\
s_2 &= \underbrace{1000000000}_{10}\ \ \underbrace{100000000}_{9}\ \underbrace{\overset{\rightarrow}{1}\ 0^*\ 000000000}_{11}1\ \ 0\\
s_3 &= \underbrace{\boxed{1}000000000}_{10}\underbrace{\boxed{1}00000000\ 0}_{10}\underbrace{\boxed{1}000000000}_{10}\boxed{1}0
\end{aligned}
$$

p	$R_s(p)$	$R_{s_1}(p)$	$R_{s_2}(p)$	$R_{s_3}(p)$
...	0	0	0	0
8	0	1	0	0
9	1	3	1	0
10	1	4	1	3
11	1	3	1	0
12	0	1	0	0
...	0	0	0	0

- *Since s has no period, $R_s(p)$ shows no significant autocorrelation.*

- *Reducing the resolution of s by duplicating 1s, we obtain string s_1 whose autocorrelation $R_{s_1}(p)$ has a peak at $p = 10$ with the matching 1s shown by frames.*

- *Restoring the original resolution, we obtain the initial string $s_2 = s$ where the position of the lost correlation is indicated by 0^*. To restore the high autocorrelation we transform s_2 into s_3 by moving $\overset{\rightarrow}{1}$ to the position of 0^*.*

- *The peak of $R_{s_3}(p)$ at $p = 10$ suggests that s is either quasi-periodic or periodic under the variable tempo*

$$
T(n) = \begin{cases} 10/10, & 1 \leq n \leq 10 \quad \textit{(initial tempo)}\\ 10/9, & 11 \leq n \leq 19 \quad \textit{(acceleration)}\\ 10/11, & 20 \leq n \leq 30 \quad \textit{(deceleration)} \end{cases} .
$$

If tempo curve T looks too complex then s should be considered a rhythmic pattern under a constant tempo, otherwise s is considered a regular pulse train under a variable tempo, which in notation of Example 5 looks as follows:

$$s = \{1000000000\}\ \mathbf{R\ 0\ 2\ 10/9\ 10/11}\ ,$$

where $\{1000000000\}$ *is the generative beat with off-beats and* **R 0 2 10/9 10/11** *means* **R***epeat from* **0** *moment,* **2** *times, with tempos* **10/9, 10/11***, respectively. Of course, the final decision about the least complex representation depends on coding conventions and complexity measures.*

All of these resemble the operation of perceptrons [Minsky and Papert 1988] and pyramidal data structures [Hummel 1987]. Resolution reduction (= filtering) is common while recognizing similarities [Palmer 1983, Witkin 1983, Bouman and Liu 1991], but we apply it as an intermediate step to find correlated elements, which are then slightly modified to the end of building simple data representations.

One can always represent the given data as generated by a few simple patterns. For example, every sequence of time events can be represented as generated by the regular beat $\{1\ 0\}$ with a tempo leap at every beat. However, transformations of low-level patterns, in the given case described by the tempo curve (= second-level pattern), may be too complex. Since the goal is reducing the total complexity, high complexity of second-level patterns makes pointless simplicity of first-level patterns, so that a certain compromise of sharing the complexity among the levels must be found.

The correlative perception modeling requires a lot of, however, elementary calculations that are easily implemented in neural networks with parallel processing. Since neural networks are considered models of the brain [Rossing 1990, p. 164], the operations described can be relevant to human cognition.

Let us recapitulate the main statements of the section.

1. The correlative perception is a tight interaction of two principles of data self-organization — grouping and simplicity.

2. These principles are used to represent data in terms of generative elements (low-level patterns) and their transformations (high-level patterns).

3. The simplicity principle is modeled using Kolmogorov's criterion of least complex data representation (least memory for the generative algorithm).

4. The method of variable resolution performs a directed search for generative elements and structural data representations. Under poor resolution, it finds similar elements and suggests their local adjustments (displacements, deformations, etc.) that enable building compact data representations.

5. The final choice of data representation is made with respect to the criterion of total complexity, which in turn depends on coding conventions and complexity measures.

3 Evidence of correlative perception

In this section, the problem of voice separation and chord recognition is considered in terms of correlative perception. It should be noted that the number of voices perceived may differ from the number of actual sound sources. On the one hand, a single source like a loudspeaker or a piano deck produces an effect of several voices. On the other hand, several organ pipes can produce an effect of a single voice. Therefore, chord recognition is not recognition in the physical sense but a question of acoustical data representation.

3.1 Chord recognition as representation of auditory data

In our consideration, a *voice* is an acoustical *trajectory* drawn by a generative spectrum whose partials move in parallel (have a ‘common fate’) with respect to the $\log_2$-scaled frequency axis. The generative spectrum is a low-level pattern associated with a tone, and its trajectory is a high-level pattern of *melodic line.* Similarly, a *chord* is an acoustical *contour* drawn by shifts of a generative spectrum (associated with a tone) along the $\log_2$-scaled frequency axis. Thus, if movements of a tone spectrum are in dynamics then we have an acoustical trajectory — a voice; if their displacements are in statics then we have an acoustical contour — a chord. The analogy between visual and auditory data is shown in Table 4.

These definitions of voice and chord prompt an approach to voice tracking and chord recognition:

- Represent sound spectra as generated by synchronous moves of groups of partials (= tones with similar spectral structures) both in dynamics and statics.

Finding such groups of partials requires no knowledge of tones and their properties. It suffices to detect *similarities* in sound spectra — unlike (or complementary to) the approach by [Mont-Reynaud and Mellinger 1989, Mont-Reynaud and Gresset 1990], which is based on finding known spectral patterns, recognizing tone dissimilarities and their minor asynchrony in chords.

3.2 Correlative perception and chord recognition

Example 8 (Recognition of chords and voice separation) *Let us consider two simple chords (in fact, two intervals) shown in Figure 4 assuming that each voice has five partial tones of equal power. The spectra of the chords are depicted in the middle of the figure, where the horizontal axis shows time t (in fact, the chord number), the vertical* $\log_2$*-scaled axis shows frequencies in Herz and corresponding pitches, and the black and white bars denote the partials belonging to the lower and upper voices, respectively. On the right hand, the same spectra are represented in the binary form for semitone (1/12 octave) frequency bands ranging from* e_1 *to* $c\sharp_3$*. The arrows indicate the parallel motion (‘common fate’) of partial tones of each voice. Thus, each chord is identified with the binary string*

$$s_t(n) = \begin{cases} 1 & \textit{if the signal level in the n-th frequency band} \neq 0, \\ 0 & \textit{otherwise} \ , \end{cases}$$

where

Table 4: Analogy between visual and auditory data

	In statics		In dynamics	
	Visual data	Auditory data	Visual data	Auditory data
Stimuli	Pixels	Partial tone	Pixels	Partial tone
Low-level pattern	A	Note	Object	Note
High-level pattern	B-contour	Chord	Trajectory	Voice

$t = 1, 2$ *is the chord number, and*

$n = 1, \ldots, 34$ *is the frequency band index.*

The voices' melodic intervals *— between the tones in successive chords— are recognized by peaks of the correlation function*

$$R_{t,t+1}(i) = \sum_n s_t(n) s_{t+1}(n+i) \ .$$

If $R_{t,t+1}(i)$ has a peak at i, we suppose that the correlated sub-spectra (= similar groups of partials) correspond to the tones that differ in the interval of i semitones. In this case, the sub-spectrum correlated is said to be the generative group of partials *of the given interval. Table 5 displays the most salient melodic intervals. The generative group of partials of each melodic interval is described by the corresponding frequency band indices. Although pitch plays no role in recognizing intervals, the generative groups of partials are conditionally identified with the pitch of the lowest partial.*

Similarly, the harmonic intervals *in Chord t are recognized by peaks of the autocorrelation function*

$$R_{t,t}(i) = \sum_n s_t(n) s_t(n+i) \ .$$

Table 6 and Table 7 display the most salient harmonic intervals inherent in the first and in the second chord, respectively. The most salient harmonic interval in both tables is the interval of prime ($i = 0$) because the autocorrelation of an unshifted spectrum is always maximal.

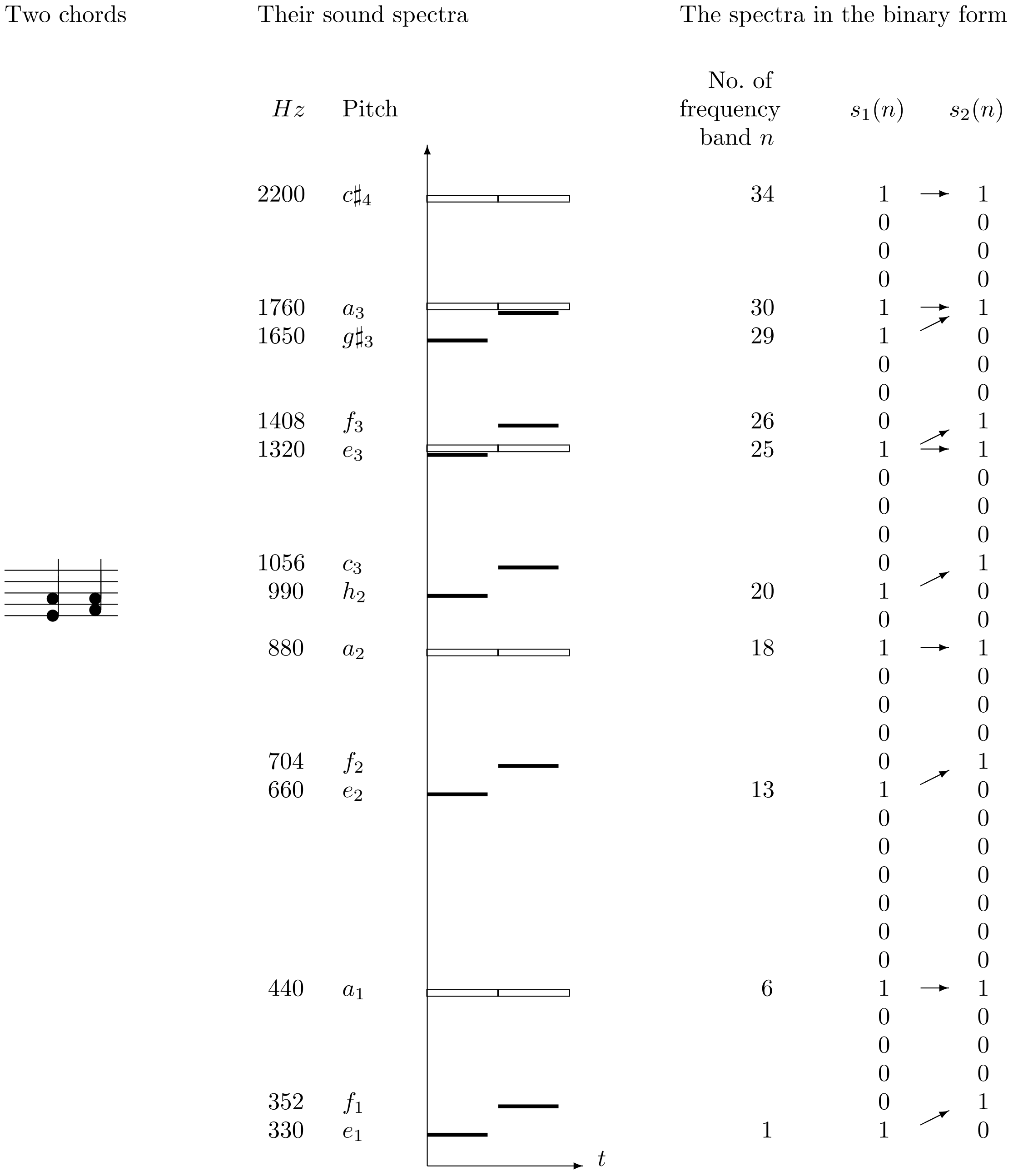
As show Tables 5–7, a high correlation indicates the true intervals, both melodic and harmonic.

Since chord spectra consist of numerous partials, certain groups of partials can correlate by chance being associated with no tones. Such occasional correlates usually have irregular structures that differ from the repetitive structure of generative tones. To distinguish true generative tones, the correlated groups of partials are tested on multiple correlations by multi-correlation analysis.

Theorem 1 (Necessary condition of a generative tone spectrum) *Let a binary chord spectrum $S = s(n)$ be generated by a binary tone spectrum T shifted by intervals $j_1 < \ldots < j_k$. Then it holds*

1. *The multi-autocorrelation functions of degree $k, k-1, \ldots$ have the following peaks:*

$$\begin{array}{rcll} R(i_1, \ldots, i_k) & = & \sum_n s(n)s(n-i_1)\cdots s(n-i_k) & at \ \ (j_1, \ldots, j_k) \\ R(i_1, \ldots, i_{k-1}) & = & \sum_n s(n)s(n-i_1)\cdots s(n-i_{k-1}) & at \ \ (j_1, \ldots, j_{k-1}) \\ \multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\ R(i_1) & = & \sum_n s(n)s(n-i_1) & at \ \ j_1 \ . \end{array} \qquad (6)$$



Hz	Pitch	No. of frequency band n	$s_1(n)$		$s_2(n)$
2200	$c\sharp_4$	34	1	→	1
			0		0
			0		0
			0		0
1760	a_3	30	1	→	1
1650	$g\sharp_3$	29	1	↗	0
			0		0
			0		0
1408	f_3	26	0		1
1320	e_3	25	1	↗ →	1
			0		0
			0		0
			0		0
1056	c_3		0		1
990	h_2	20	1	↗	0
			0		0
880	a_2	18	1	→	1
			0		0
			0		0
			0		0
704	f_2		0		1
660	e_2	13	1	↗	0
			0		0
			0		0
			0		0
			0		0
			0		0
			0		0
440	a_1	6	1	→	1
			0		0
			0		0
			0		0
352	f_1		0		1
330	e_1	1	1	↗	0

Figure 4: Spectral representations of two chords

Table 5: Most salient melodic intervals between two chords

Correlation $R_{1,2}(i)$	Interval i	Generative group of partials	Interval notation
6	-4	6 18 25 29 30 34	$(a_1; f_1)$
5	0	6 18 25 30 34	$(a_1; a_1)$
5	1	1 13 20 25 29	$(e_1; f_1)$
5	5	1 13 20 25 29	$(e_1; a_1)$
2	-7	13 25	$(e_2; a_1)$

Table 6: Most salient harmonic intervals in the first chord

Autocorrelation $R_{1,1}(i)$	Interval i	Generative group of partials	Interval notation
9	0	1 6 13 18 20 25 29 30 34	$(e_1; e_1)$
5	5	1 13 20 25 29	$(e_1; a_1)$
3	7	6 13 18	$(a_1; e_2)$
2	4	25 30	$(e_3; g\sharp_3)$
1	1	29	$(g\sharp_3; a_3)$

Table 7: Most salient harmonic intervals in the second chord

Autocorrelation $R_{2,2}(i)$	Interval i	Generative group of partials	Interval notation
9	0	2 6 14 18 21 26 30 34	$(f_1; f_1)$
5	4	2 14 21 26 30	$(f_1; a_1)$
2	5	21 25	$(c_3; f_3)$
2	7	14 18	$(f_2; c_3)$
1	1	25	$(e_3; f_3)$

2. *The multi-correlated groups of partials $T_{j_1,\dots,j_k}, T_{j_1,\dots,j_{k-1}}, \dots, T_{j_1}$ that cause the peaks of multi-autocorrelation functions (6) include generative tone spectrum T; moreover*

$$T \subset T_{j_1,\dots,j_k} \subset T_{j_1,\dots,j_{k-1}} \subset \dots \subset T_{j_1,j_2} \subset T_{j_1} \quad . \tag{7}$$

The inclusions (7) prompt how to find a repetitive sub-spectrum recursively. To find a multi-correlated sub-spectrum of degree p, it suffices to check the multi-correlated sub-spectra of degree $p-1$ for the multi-correlation of degree p and reject the ones that fail the test.

Let us summarize the main points of the section.

1. Binary chord spectra can be decomposed into tone spectra by peaks of the multi-autocorrelation function of the chord spectra. The multi-correlated tone spectra can be found by recursion on the degree of multi-correlation.

2. Since all musical tones (= with pitch salience) have harmonic spectra (= with the partial frequency ratio $1 : 2 : 3 : \dots$), the multi-correlation analysis finds them regardless of their timbre.

3. The approach described can be applied to track polyphonic voices. For this purpose, multi-correlation analysis of successive short-time spectra should be performed.

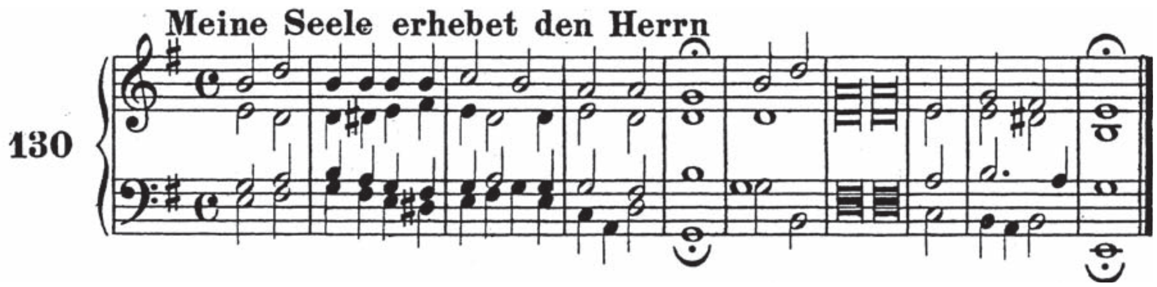


Figure 5: The 130th chorale from J.S.Bach's *371 Four-part chorales*

3.3 Computer experiments

The model of chord recognition based on multi-autocorrelation analysis was tested in a series of computer experiments. The recognition results for a five-tone chord $(c; e; g; h, d_1)$ are collected in Table 8. Correct recognition is denoted by OK, and the misidentifications are given in full with harmoniously alien pitches in bold. Recognition experiments differed in the spectral resolution of 1, 1/2, 1/3 or 1/4 semitone, which is interpreted as 'sharpness of hearing'. The experiments used synthetic voices with 5, 10, 16 or 32 harmonics of equal power (the more harmonics, the noisier the voices), which determined the 'task difficulty'. Table 8 shows that the recognition accuracy improves as the 'hearing sharpens' and deteriorates as the 'task difficulty' increases.

The same audio tests with humans showed that the model's recognition limits are approaching that of trained musicians. Table 8 also evidences that the model has a human-like ability to recognize most easily the major/minor function, then the harmony to within octave transpositions of notes, and finally the correct chord's interval structure.

The simple correlation approach (with no multi-correlation) was applied to recognize harmonic and melodic intervals in the 130th Bach's chorale in Figure 5. The choral was considered a sequence of 24 chords whose audio spectra were synthesized and analyzed under various conditions. The experiments differed in voice type (harmonic, or inharmonic), number of partials per voice (5, 10 or 16), spectral resolution (1, 1/2, 1/3 or 1/4 semitone), and considering all intervals or only the ones < 12 semitones — to avoid octave correlations. For the spectral resolution $\leq 1/2$ semitone ($\leq 1/4$ tone), the recognition reliability was about 98% (94 correctly recognized notes out of 96) or better. The 100%-correct recognition was attained using the multi-correlation analysis of the 3rd degree (3 spectra analyzed at a time). For details see [Tanguiane 1993, Tanguiane 1994, Tanguiane 1995].

We conclude the following:

Table 8: Recognition of chord $C_{7M/9} = (c; e; g; h; d_1)$

Accuracy in semitones	Number of partials per voice (task difficulty)			
(sharpness of hearing)	5	10	16	32
1	OK	$(c; e; g)$	$(c; g; \boldsymbol{a\flat}; \boldsymbol{b}; c_1; \boldsymbol{e\flat_1})$	$(c; g; \boldsymbol{a\flat}; \boldsymbol{b}; c_1; \boldsymbol{e\flat_1})$
1/2	OK	OK	OK	$(c; c_1; d_1; e_1; g_1; \boldsymbol{a\flat_1}; \boldsymbol{a_1}; h_1; c_2)$
1/3	OK	OK	OK	$(c; e; g; h; c_1; d_1; e_1; g_1)$
1/4	OK	OK	OK	OK

1. The model's recognition capabilities are similar to that of humans: Correct recognition is easier for voices with fewer partials and under more accurate spectral resolution; the chord type (major or minor) is recognized first, next the harmony to within octave tone transpositions, and then all the intervals.

2. Already the simple (pairwise) correlation approach demonstrates a reliable recognition of Bach's four-part polyphony: For the spectral resolution of 1/2 semitone, the recognition reliability is about 98%.

3.4 Mathematical foundations

Now we justify the following statements theoretically.

1. If a chord spectrum is generated by shifts of a harmonic tone spectrum (with the partial frequency ratio $1 : 2 : 3 : \ldots$) along the $\log_2$-scaled frequency axis then the model can correctly recognize both the generative tone and the chord's interval structure.

2. The representation of this chord's spectrum as generated by shifts of a harmonic tone spectrum is optimal (= is the least complex); moreover, this optimal representation is unique.

The formal proofs of these statements for two-tone intervals and most common chords like major and minor triads are given in [Tanguiane 1993, Tanguiane 1994, Tanguiane 1995]. Here, we only outline the basic facts, starting with a few assumptions on hearing and correlative perception.

Axiom 1 (Logarithmic pitch) *The frequency axis is* $\log_2$*-scaled.*

Axiom 2 (Insensitivity to the phase of the signal) *We consider exclusively power spectra.*

Axiom 3 (Grouping by structural similarity) *Chord spectrum partials are grouped into sets with equal partial frequency ratios.*

Axiom 4 (Simplicity principle) *Spectral data are represented in the least complex way in the sense of Kolmogorov, that is, using the most compact generative algorithm.*

Our reasonings use the concept of *discrete power spectrum*

$$S = \sum_n c_n \delta_n \ ,$$

where

n are indices of frequency bands (e.g., corresponding to semitones),

δ_n are point impulses given by Dirac delta functions

$$\delta_n = \delta_n(x) = \begin{cases} +\infty & \text{if} \quad x = n \\ 0 & \text{if} \quad x \neq n \end{cases} \qquad \text{such that} \qquad \int_{-\infty}^{+\infty} \delta_n(x)\, dx = 1 \ ,$$

c_n are the impulse powers given by non-negative integers.

To shift a spectrum S by k frequency bands, the *convolution product* (a kind of multiplication of spectra) is applied

$$S * \delta_k = [S * \delta_k](x) = \sum_n c_n \delta_{n+k} \ .$$

To generate a chord spectrum C by a tone spectrum T, several shifts of the tone spectrum T by intervals $i_1, \dots, i_k$ (measured in frequency bands) are applied:

$$C = T * (\delta_{i_1} + \cdots + \delta_{i_k}) = T * I \ , \tag{8}$$

where the spectrum $I = \delta_{i_1} + \cdots + \delta_{i_k}$ is called the chord's *interval structure*.
To decompose a chord into notes, we have to find the chord spectrum's deconvolution (= convolution factorization) as in (8). The next lemma reduces the problem to factorization of polynomials.

Lemma 1 (Isomorphism between spectra and polynomials) *Discrete power spectra with respect to addition and convolution are isomorphic to polynomials over non-negative integers:*

$$S = \sum_{n=0}^{N} a_n \delta_n \quad \longleftrightarrow \quad p(x) = \sum_{n=0}^{N} a_n x^n \ .$$

By Lemma 1, an *indecomposable spectrum*, that is, not generated by shifts of its proper sub-spectrum, corresponds to an irreducible polynomial (= that cannot be factored). Unlike polynomials over all integers, polynomials over non-negative integers do not constitute an algebraically complete system. In particular, their factorization into irreducible polynomials is not necessarily unique. At first we illustrate this phenomenon using an algebraically incomplete set of integers.

Example 9 (No unique factorization into primes) *If number 2 is removed from natural numbers then 4 and 8 are not factorable, implying two factorizations of 64 into these 'primes':*

$$64 = 8 \times 8 = 4 \times 4 \times 4 \ .$$

Such a non-unique factorization is inherent in polynomials over non-negative integers.

Example 10 (No unique factorization for polynomials over non-negative integers) *In the system of polynomials over non-negative integers, the factorization into irreducible polynomials is not necessarily unique:*

$$\begin{aligned} 1 + x + x^2 + x^3 + x^4 + x^5 &= (1 + x^2 + x^4)(1 + x) \\ &= (1 + x + x^2) \ \underbrace{(1 + x^3)}_{(1-x+x^2)(1+x)} \ . \end{aligned}$$

The following lemma formulates an important property of musical tones with a pitch salience. It says that *harmonic spectra*, i.e. with partial frequency ratios $1:2:3:\ldots$, are indecomposable acoustical units.

Lemma 2 (Indecomposability of harmonic spectra) *Under sufficiently fine spectral resolution, spectra of harmonic tones are indecomposable.*

Lemma 2 does not hold without Axiom 1 that postulates the $\log_2$-scaled pitch.

Example 11 (Decomposability of harmonic spectra for a linear pitch axis) *Let us consider a linearly scaled frequency axis with 12 frequency bands for the first octave (hence, 24 bands for the second octave). By the isomorphism between spectra and polynomials established in Lemma 1, the harmonic spectrum with four partials (with pitches c_1, c_2, g_2, c_3) is decomposable:*

$$\begin{aligned} \delta_0 + \delta_{12} + \delta_{24} + \delta_{36} &\leftrightarrow 1 + x^{12} + x^{24} + x^{36} \\ &= (1 + x^{12})(1 + x^{24}) \\ &\leftrightarrow (\delta_0 + \delta_{12}) * (\delta_0 + \delta_{24}) \ . \end{aligned}$$

Lemma 3 (Indecomposability of two-tone intervals and major or minor chords) *Under sufficiently fine spectral resolution, the interval structures of two-tone intervals, as well as of major and minor triads are indecomposable.*

Lemma 3 does not hold without Axiom 2 that postulates the ear's insensitivity to the phase of the signal.

Example 12 (Decomposability of the structure of minor third under sensitivity to signal phase) *Let the frequency axis be $\log_2$-scaled with the frequency bands corresponding to semitones. The simplest sensitivity to the signal phase — the phase inversion (180^0) — means that negative coefficients in spectral representations are allowed. Under these assumptions, the interval structure of minor third $I = \delta_0 + \delta_3$ is decomposable:*

$$\begin{aligned} \delta_0 + \delta_3 &\leftrightarrow 1 + x^3 \\ &= (1 + x)(1 - x + x^2) \\ &\leftrightarrow (\delta_0 + \delta_1) * (\delta_0 - \delta_1 + \delta_2) \ . \end{aligned}$$

The *complexity* of spectrum S is defined to be the number of its partial tones. The complexity of its decomposition $S = T * I$ is the sum of the complexities of T and I.

Theorem 2 (Chord recognition by optimal spectral representation) *Let the spectrum C of a two-tone interval or major triad or minor triad be generated by shifts of a harmonic spectrum T with sufficiently numerous partials, and let the spectral resolution be sufficiently fine. Then the least complex representation of spectrum C is unique and corresponds to the way of its generation: $C = T * I$, where I is the interval structure.*

Theorem 2 does not hold if the generative tone is inharmonic (= has no pitch salience since the partial frequency ratio $\neq 1:2:3:\ldots$).

Example 13 (No unique decomposition of a chord generated by an inharmonic spectrum) *Let the frequency axis be* $\log_2$*-scaled with the frequency bands corresponding to semitones. Then the polynomial from Example 10 corresponds to the spectrum of minor triad generated by an inharmonic tone spectrum, which has alternative deconvolutions into indecomposable spectra:*

$$\begin{aligned}
\delta_0+\delta_1+\delta_2+\delta_3+\delta_4+\delta_5 &= \underbrace{(\delta_0+\delta_1+\delta_2)}_{\text{Inharmonic spectrum}} * \underbrace{(\delta_0+\delta_3)}_{\substack{\text{Interval structure}\\ \text{of minor triad}}} \\
&= \underbrace{(\delta_0+\delta_2+\delta_4)}_{\text{Inharmonic spectrum}} * \underbrace{(\delta_0+\delta_1)}_{\substack{\text{Interval structure}\\ \text{of minor second}}} .
\end{aligned}$$

Thus, we conclude

1. Under natural assumptions, the model of correlative perception correctly recognizes intervals and most common chords constituted by harmonic tones (= with a pitch salience).

2. The seeming imperfection of hearing — the non-linear frequency scale and the insensitivity to the phase of the signal — turns out to be an advantage. Just this 'imperfection' enables functional hearing with sound decomposition and tracking parallel voices.

3. Harmonic tones are indecomposable units of auditory communication, explaining their role in speech and music. The interval structures of major and minor chords are also irreducible, explaining their role in music as basic harmonies.

3.5 Applications to psychoacoustics and music theory

Definition of rhythm and tempo In music theory, rhythm and tempo are defined with respect to each other, which results in a vicious circle. Indeed, rhythm does not exist without tempo, and tempo cannot be recognized with no reference to rhythmic patterns. Tempo is therefore imagined as time modulations of the periodicity generated by a rhythm, similarly to frequency modulation (FM) in broadcasting.

The model of correlative perception implies an operational definition of rhythm and tempo as patterns of different levels in a two-level *optimal representation* of time events. As noted at the end of Section 2.4, the vicious circle is overcome due to the intervention of optimality criterion. For further details see [Tangian 1998].

Definition and function of interval hearing The model of correlative perception suggests a general definition of musical interval: It its the shift distance d between two tone spectra with *similar* structures. This definition makes no reference to pitch because no reference points are needed. Therefore, this definition is applicable to similar inharmonic sounds (with no pitch salience) like bell-like sounds, or similar band-pass (pink) noises.

A visual analogy is the distance between similar and dissimilar objects shown in Figure 6. For similar objects, the distance can be defined with no reference points

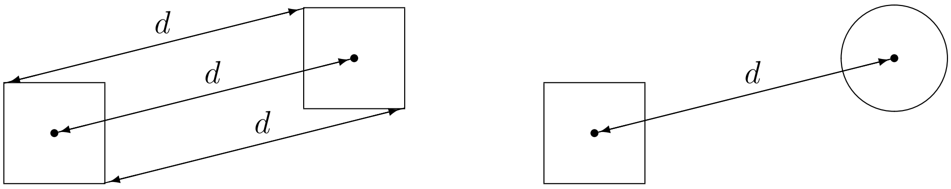


Figure 6: The distance between similar and dissimilar objects

since the shift distance is the same for all the object's points. For dissimilar objects, like a square and a circle, the distance cannot be defined other than between some special points like geometric centers, closest points, etc. In the same way, the distance between tones with similar spectral structures (with or without pitch salience) can be estimated with no reference to pitch.

It should be emphasized that most humans cannot identify pitch but can recognize intervals. The interval hearing is a particular instance of correlative perception in the auditory domain. It is a manifestation of the ability to recognize similar groups of partials (= with a 'common fate'), which enables following simultaneous acoustical processes in a much broader context than music.

Prohibition of parallel voice leading In harmony and counterpoint, parallel fifths and octaves are strictly prohibited. From the viewpoint of correlative perception, the parallel voices, with all their partials moving in parallel, fuse into one voice with a new timbre. Such a 'disappearance' of some independent voices, entry of a new one with a foreign sound, and then recovery of the voices lost violate the polyphonic homogeneity. On the other hand, parallel voice leading is used as a special effect in some musical instruments like pipe organs or synthesizers as well as in orchestral arrangements.

This is illustrated by two excerpts from *Bolero* (1928) by M.Ravel (1875–1937). In Figure 7, the horns' theme is doubled by celesta in one and two octaves and by flutes at the twelfth and the seventeenth, resulting in a strictly parallel motion of five harmonics. The parallel voice leading is emphasized by the key signatures: The theme is in C whereas the flute parts that enhance the 3rd and the 5th harmonics are in G and E, respectively. The overall effect is a harmonically 'thin' voice with a synthetic timbre, resembling the sound of electric organ. Figure 8 illustrates an alternative way of using orchestral instruments. The melodic parts are not strictly parallel, producing an effect of a harmonically 'thick' voice texture.

Thus, the prohibition of parallel voice leading in harmony is explained by the effect of violating the polyphonic homogeneity. Parallel voices, losing their independence, fuse into one *functional* voice, within which the harmony rules are irrelevant.

For other applications to music theory, see [Tanguiane 1993, Chapter 7].

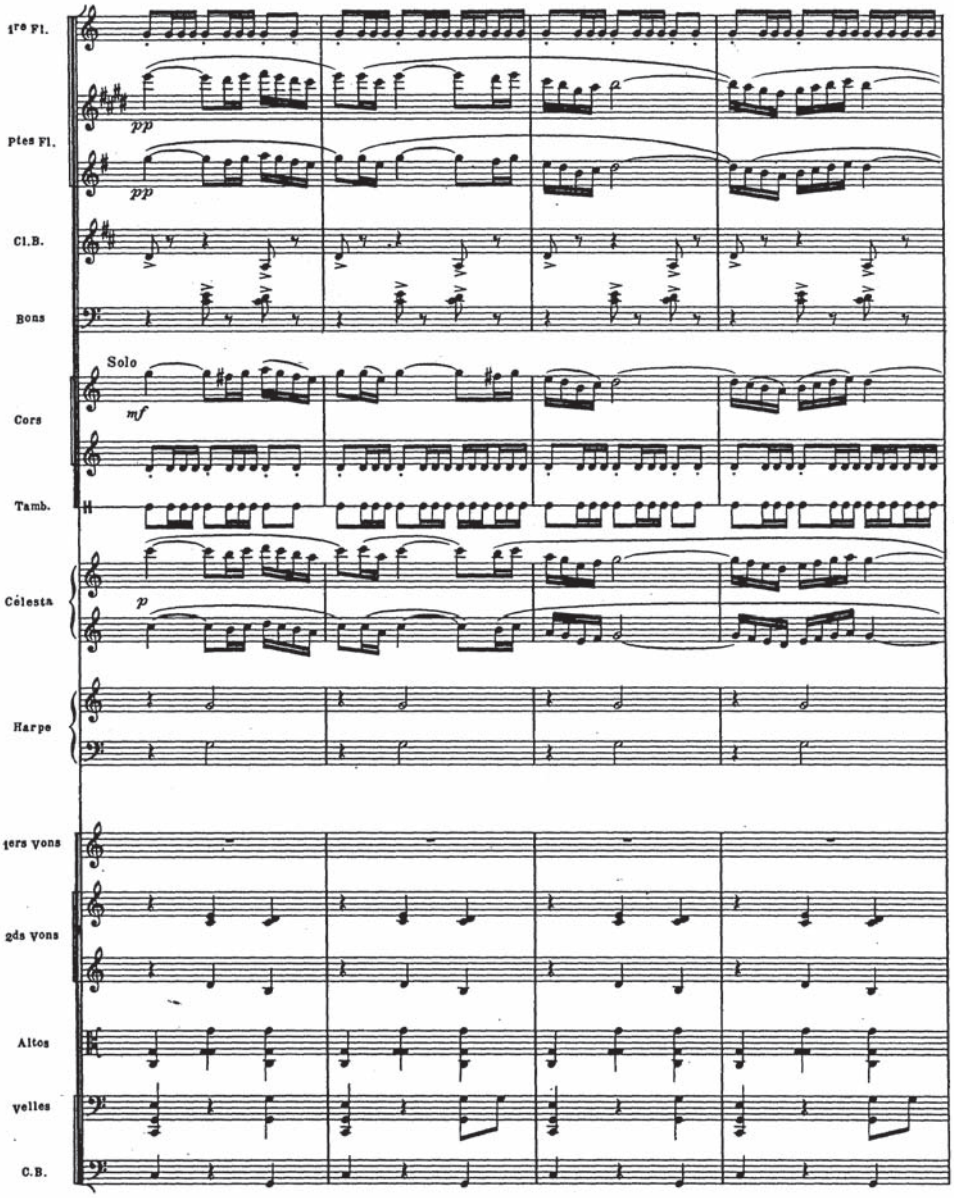


Figure 7: Strictly parallel voice leading in Ravel's *Bolero* (rehearsal number 5)

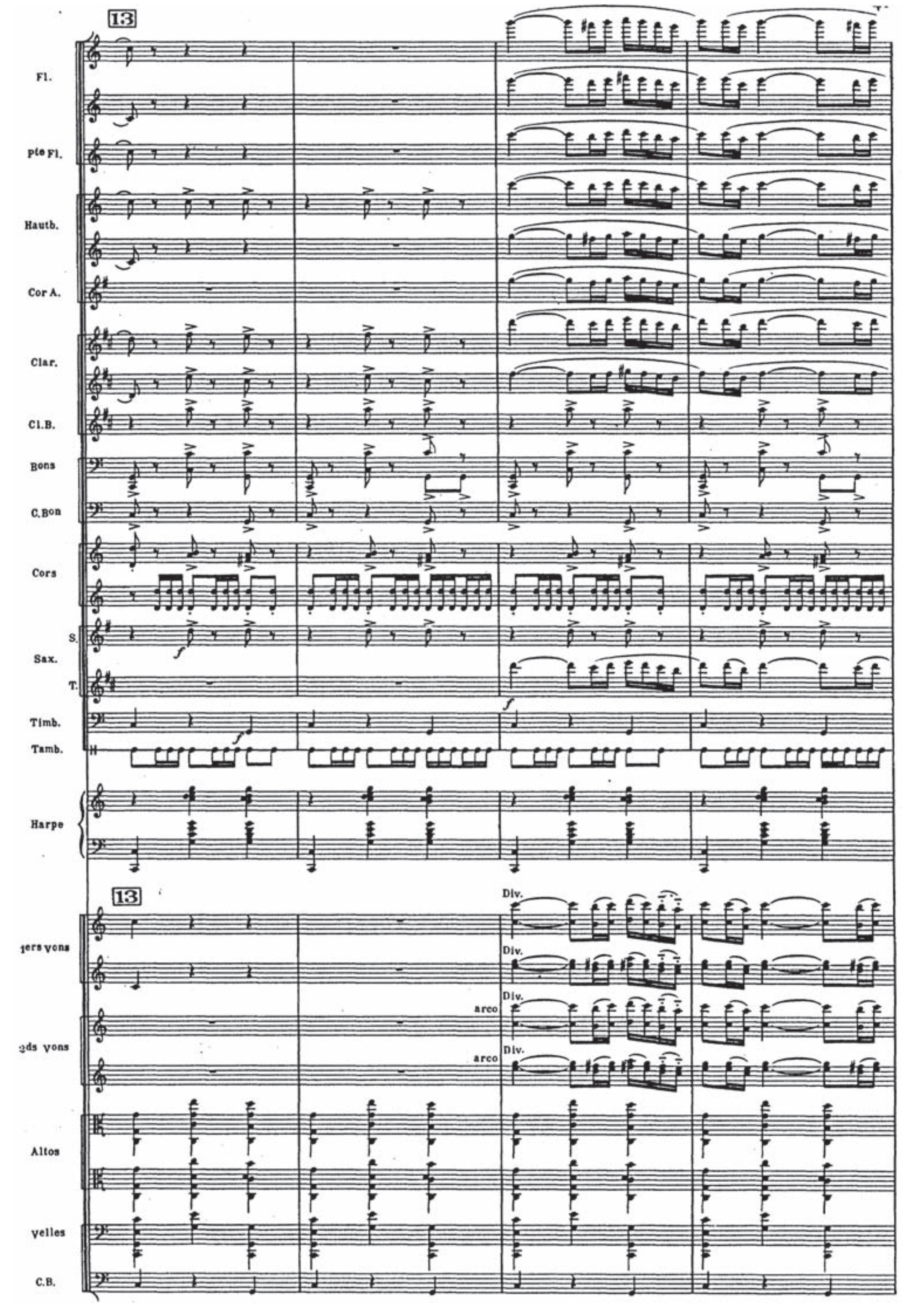


Figure 8: Harmonic voice leading in Ravel's *Bolero* (rehearsal number 13)

4 Generalization to abstract thinking

4.1 Multi-level hierarchies

Computers and the brain function differently. To process specific cases, a computer requires a more or less general algorithm. A computer will not develop a universal problem solver basing on a few examples. Without an algorithm, it won't even start running. One can say that a computer needs an algorithm to process examples, whereas the brain uses examples to derive algorithms. Indeed, people learn from particular instances better than from ready-made rules. To master arithmetic, a pupil exercises with numbers instead of learning its foundations, and the theoretical issues are only taught after illuminating examples. In practice, humans often develop general principles themselves and apply them to new situations. Such a creative systematization of own experiences goes in line with Plato's theory of knowledge as recollection of what the soul has seen in the world of ideas [Plato 380 BC, Plato 360 BC].

All we have done so far is an attempt to model this process. We have built two-level data representations (in fact, generative algorithms) just by recognizing repetitive data blocks and reducing data redundancy. Since these representations reflect the data generation causality and recognize data structures without knowing the structures, we obtain knowledge 'from nothing' or, better to say, thanks to the universal optimality principle, which operates in parallel in the physical and information worlds.

This approach is generalizable to hierarchies with more than two levels. The following example illustrates the idea of information hierarchy where each level serves as a carrier for the next-level information — similarly to tones that are interval structure carriers or rhythmic patterns that are tempo curve carriers.

Example 14 (Information hierarchy in an AM-wave) *Let us see how speech information is transmitted by an AM (= amplitude modulation) radio wave with a constant frequency, say, of 200 kHz. The amplitude of this carrier wave is modulated by the audio wave — second-level wave — with frequencies up to 20 kHz. The audio wave is the carrier for short-time sound spectra that constitute the third information level. The fourth information level consists of variable spectral envelopes that determine speech formants recognizable by the brain.*

4.2 Repetitive carriers in multi-level hierarchies

As illustrated by Example 14, repetitive elements carry the next-level information generating some message. It is a very general principle that is also used in much less technical environments.

Example 15 (New meaning carried by repetitions) *Let us imagine a person who waits for a vital message. At first (s)he states:*

— No message.

Then the concern increases:

— No message. No message! No message!!

Desperate, (s)he repeats monotonously:

— No message ... No message ... No message ...

Here, the fact of repetition and the way they are made carry an important information beyond the verbal content.

Another meaning of repetitions is suggested by the Chinese proverb 'You said — I believed, you repeated — I doubted, you began to insist, and I realized that you were lying'.

Repetitions and their variations play an important role in non-verbal communication (music, dance, circus, painting, architecture, etc.).

Example 16 (Structural perception of classical and contemporary music) *Classical music is based on various forms of repetitions, which organize time events into structures. Due to melodic and rhythmic repetitions, notes are grouped into motives, motives into phrases, and so on, up to the level of musical form. Rhythmic music is essentially multi-level and, on the other hand, 'self-sufficient' to be perceived.*

Contemporary music, sometimes without or little repetitiveness, is different in this respect. Since perceiving high semantic levels is often possible due to redundant carriers, no repetitiveness implies no evident information hierarchy. Therefore, the direct message of the music thus organized is restricted to the level of single events.

For instance, any form of repetition is avoided in serial music, making its structure so little redundant and complex that it cannot be perceived. Alternatively, stochastic music makes a similar impression but with almost no rules at all. Its inventer, Iannis Xenakis (1922–2001) declared in 1954:

> *Linear polyphony destroys itself by its very complexity; what one hears is in reality nothing but a mass of notes in various registers. The enormous complexity prevents the audience from following the intertwining of the lines and has as its macroscopic effect an irrational and fortuitous dispersion of sounds over the whole extent of the sonic spectrum. There is consequently a contradiction between the polyphonic linear system and the heard result, which is surface or mass. This contradiction inherent in polyphony will disappear when the independence of sounds is total. In fact, when linear combinations and their polyphonic superpositions no longer operate, what will count will be the statistical mean of isolated states and of transformations of sonic components at a given moment. The macroscopic effect can then be controlled by the mean of the movements of elements which we select. The result is the introduction of the notion of probability, which implies, in this particular case, combinatory calculus. Here, in a few words, is the possible escape from the "linear category" in musical thought.*
>
> *— I. Xenakis (1971)* Formalized Music*, p. 8. [Xenakis 1971]*

It is noteworthy that the complexity of music is assessed regarding the perceptibility of its structure, which in our terminology corresponds to the complexity of generative algorithms.

If structural cues are insufficient then perception matches incoming data with patterns from past experiences. If no appropriate pattern is found then the data can hardly be perceived adequately.

Example 17 (Knowledge-based perception of contemporary music) *Certain contemporary compositions use structures that have nothing to do with music. For instance, there are musical pieces encoding chess games or river water levels [Dodge and Bahn 1986, Wuorinen 1979]. Since these structures are neither repetitive nor musically meaningful they are not perceptible at all. One gets a strange impression of something unconventional, and the contrast with the past experience is the only criterion to evaluate the piece.*

In vision, repetitions play the same organizing role as in hearing.

Example 18 (Dynamic scene analysis) *Figure 9 shows a sequence of cinema frames with a motionless object and a moving one — a tennis net and a ball, respectively. Recognizing similar elements in successive frames enables interpreting them as consecutive states of the same object. The optimal representation of this image sequence (= the shortest program to generate the scene) is object-trajectory-oriented (cf. with object-oriented programming):*

- *The repeated objects (ball, tennis net) are the generative elements, the carriers.*
- *The trajectories are high-level patterns of their transformations. For the motionless tennis net, the trajectory is zero.*

No knowledge is required to distinguish between the two objects and separate them. The knowledge is only required to identify them as a tennis net and a ball.

4.3 Abstract concepts as labels for stable high-level patterns

Many cognitive studies discuss building abstract notions from practical experiences; for a review see [Giunchiglia and Walsh 1992]. So far we have considered data representations for single experiences but the patterns occurring repeatedly in various situations can be processed in the same way — memorized as 'typical' and labeled. We can imagine that a model of multiple-level data representation operates on multiple experiences as follows:

- The high-level patterns extracted from numerous cases are memorized and compared.
- The similar patterns extracted are considered as carriers for new patterns.
- A more advanced pattern hierarchy is built over the pattern aggregates that are stored in the memory.

The meaning of a pattern can be understood as a system of its relationships and interactions with other patterns. For instance, the notion 'table' includes its configuration (board, meaning the relationship with other boards, four legs, meaning the relationship with legs of other furniture); height (twice higher than a chair seat, meaning the relationship with 'twice' and chair), etc. Functionally, 'table' interacts with 'board', 'leg', 'chair', number 2 ('twice higher'), number 4 ('four legs'), etc.

Proceeding in the same way, we come to a multi-level model, where every patterns is formed by stable relationships between patterns of the lower level. For example, 'two'

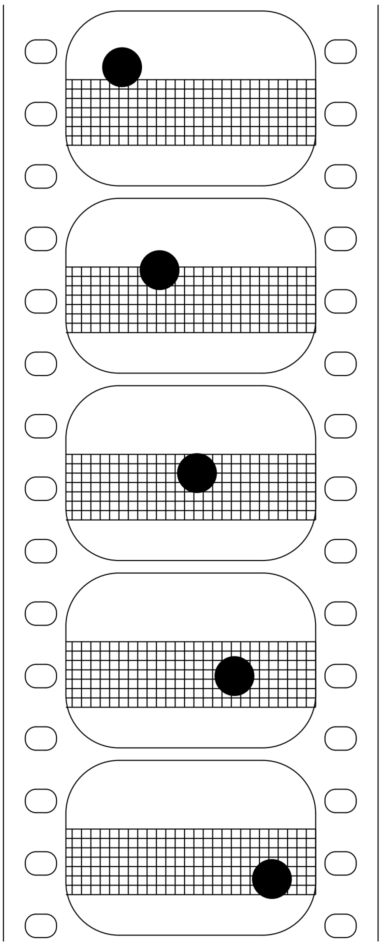


Figure 9: Flying ball in cinema frames

fixes a certain repetitive relationship emerging in every couple of objects like two eyes, two ears, two hands, etc. In this sense, the repetitive common feature 'to consist of two elements' is the carrier of the abstract pattern of 'number two' that is not associated with particular objects.

After 'two', 'three', and so on, have been realized as patterns of thinking, one also realizes that all these integers have common features like similar definitions or possibilities to be added with each other. These repetitive properties (= 'common fate') are carriers of some stable relationships, which constitute the basis for a higher abstraction level — the notion of a number system. In turn, this leads to the idea of 'infinity' and other absolutely new concepts.

Generally speaking, labeling a new notion is not necessary. A label is aimed at calling the associated pattern/algorithm by a short code. Indeed, using a data block is easier if it has a recognizable name; by the same reason a text file 'Untitled-1' is saved as 'Report_2021'. Then the label bears a certain meaning. At the same time, labeling depends not only on the context where it emerges but also on the language. As a word, a label is associated with several other words of the same language, having meaning nuances. For instance, the French and German equivalents to 'money' are respectively 'argent' (= silver) and 'Geld' (∼ 'Gold' = gold). Obviously, the English 'money' is most neutral, whereas the German

‘Geld’ is most seductive and tempting. These nuances can be critical while translating from one language to another. A literal translation is sometimes misleading because of differing etymologies, morphologies, common usage, etc. Therefore, a good translation takes into consideration the correspondence between the lists of associated words and meanings in both languages.

Thus, a model of abstraction is imagined as a pattern hierarchy, which optimally organizes itself and is provided with labels. In order to compare patterns from numerous experiences, such a model is interfaced to a data base. It finds stable relations, generates their patterns and defines new ‘mental frames’ by optimizing the data storage. In other words, cognition is understood as jointly analyzing the input and the stored data with the aim of their joint aggregation and discovering the correlation between the aggregates.

This type of cognition is inherent in developing a new mathematical theory. Unlike elementary mathematics, where the main skill is applying given definitions and facts, the main skill in higher mathematics is introducing new concepts and inventing new definitions. This type of creativity takes into account the existing concepts, their interactions, and using the new concepts as entireties. That is, a mathematician recognizes a certain configuration of phenomena that emerges in a number of cases. Then (s)he conceptualizes this configuration by providing its definition and assigning a specific label to it. As the relations between the new concepts and their implications have been established, the creative process goes further. This is very similar to determining high-level patterns carried by repetitive low-level patterns in the correlative perception.

4.4 Representation of current data by memory patterns

Interfacing the model of correlative perception to a data base implies storing selected patterns/algorithms in the long-term memory. Then the incoming data are represented in terms of either new patterns/algorithms or that from the memory or both in certain combinations. These representations, being not necessarily optimal if created from scratch, can be optimal in the extended model as requiring little additional memory.

In other words, the knowledge patterns stored in the memory compete with the current patterns of ‘naive’ perception. Obviously, the memory patterns have impact not only on the optimal representations of incoming data but also on the derivation of new high-level patterns. This corresponds to different reactions of different people conditioned by education, past experience or expectations.

Thereby, the methodology of artificial perception, in particular correlative perception, is complementary to that of artificial intelligence. Artificial perception operates on incoming data, finds their structure and represents them in a compact form. Artificial intelligence stores flexibly accessible knowledge patterns with their labels and interactions. If artificial perception models are interfaced to artificial intelligence models they can optimize data storages. Since the correlative perception is data-driven and self-organizing, it can find new regularities in the knowledge stored in the memory and develop some new concepts. That is, artificial perception models can transform a passive data storage into an active knowledge processor and ‘stimulate’ it to derive a new knowledge by the ‘internal’ motivation of saving memory.

Let us recapitulate the main points.

- The two-level model of correlative perception is generalizable to a multiple-level one. High-level patterns are relationships between repetitive patterns of the lower level, i.e. they are meta-structures and/or generalized observations.

- Labeling and storing the patterns/algorithms created by such a multiple-level model enables their use in optimally representing the incoming data, combining thereby 'naive' and 'intelligent' perceptions.

- A data base with a self-optimizing functionality and an ability to create high-level patterns from relationships between the known patterns gives a model of active abstract thinking.

- The model's motivation for the new knowledge generation is internal. It is the knowledge optimization aimed at saving memory.

4.5 Perceptual world and particularities of thinking

As already mentioned, perception is much influenced by personal experiences because the incoming data can be represented in alternative ways, depending on the patterns saved in the memory. For instance, contemporary music is perceived differently by sophisticated and inexperienced listeners.

Another source of alternative representations is the way the input data are coded. This is determined by the input sensors. For instance, music perception and music thinking depend on the properties of hearing. In our study, we assume the logarithmic pitch and the insensitivity to the phase of the signal. The linear pitch and the sensitivity to the signal phase would imply quite different hearing properties and recognition abilities.

Even more important are the senses. The human vision, hearing, touch, smell and taste determine a particular perceptual world. For instance, those who can only see think differently than those who can only hear. For instance, visual data have two or three spatial dimensions in three colors (red, green, blue), which cover in total about one octave (the range of perceptible wavelengths is 400–750nm). The fewness of color sensors and their individual tuning makes the color perception rather subjective (implying every photograph's own preference for color films). The time factor in vision is not critical, and one gets a lot of information from a static image like a photo or a picture.

Auditory perception is very different. Compared to vision, it has a weak spatial sensitivity, but the perceptible frequency range is about ten octaves (20 – 20 000 Hz). The number of frequency sensors is not three but about 1000, making the pitch perception (= equivalent of color in vision) equal for all listeners. Finally, the time dimension in hearing is extremely important because an instant impulse carries too little information.

One can hardly imagine the impact of alternative senses on the way of thinking. For instance, dolphins can generate directed ultrasonic impulses and 'see' the reflected signals, transforming them into acoustical images. Such an active acoustical vision results in a specific perceptual world, which is completely foreign for humans. This can be one of reasons why dolphins are much less communicative than might be expected given their high intelligence.

4.6 Algorithmic thinking

Thus, thinking is a special way of memory operation. It derives rules from particular instances and makes generalizations, which replace large amounts of raw data. The memory stores some basic facts, and the analytical thinking provides their flexible use. For instance, pupils memorize the multiplication table up to 10 × 10 and then use rules to multiply arbitrary numbers.

Both types of inference — memory-based or rule-based — is inherent in the domains other than mathematics. For example, the British legal procedure uses precedents (stored in archives), whereas the French tradition prescribes directly applying the law (which is a system of analytical rules). As the archives grow immensely, the 'analytical' jurisprudence becomes more practical.

Rules are kinds of algorithms, which are also stored in the memory. Therefore, the memory contains both the raw data and the 'algorithms', which represent some other raw data in a compressed (= analytical) form and help process the incoming information. The example of phenomenal memory that is mentioned in the Introduction suggests that the part of memory, which is reserved for raw data, can be more developed than that reserved for 'analytical algorithms' and vice versa.

At the same time, memory is not opposed to intelligence. Memorizing is necessary for thinking, and thinking is necessary for memorizing. For instance, learning a poem by heart is much easier if one knows the language and understands the content. Understanding guides memorizing, and retelling a poem has a shade of *recreation* suggested by the first line, rhythm, rhymes, metaphorical images and other cues.

As noted by [Herriot 1932, p. 172], the culture is what is left when everything else is forgotten. In our consideration, this is not a paradox. Even if the facts (= raw data) are forgotten, the way of thinking (= algorithms) remains. Indeed, the way of thinking is more important than particular facts.

5 Summary

Optimality as a link between physical and information worlds Cognition is understood as memory organization. The raw data are replaced by analytical representations, that is, by algorithms of their generation. It is supposed that the physical optimality is mirrored by optimal data representations. Hence, optimal (least complex) data representations reflect the physical causality in observations.

Correlativity of perception Data representations are either structural or knowledge-based. The former use general principles of data processing (are data-driven), the latter — target algorithms. Structural representations use hierarchical grouping and simplicity criteria. Their tight interaction is the principle of correlativity of perception. The correlative perception is confirmed experimentally and theoretically. It also manifests itself in some psychoacoustic phenomena and rules of music theory.

Model of thinking A multiple-level model of correlative perception interfaced to a knowledge data base is imagined as a model of thinking. The active self-organization of knowledge aimed at saving memory is supposed to generate abstract concepts.

References

[Bouman and Liu 1991] Bouman Ch, Liu B (1991) Multiple resolution segmentation of textural images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 13(2), 99–113.

[Bregman 1990] Bregman AS (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge MA: MIT Press.

[Calude 1988] Calude C (1988) *Theories of Computational Complexity.* Amsterdam: North-Holland.

[Dodge and Bahn 1986] Dodge Ch, Bahn C (1986) Musical Fractales. *Byte,* 11(6), 185–196.

[European Commission 2020] European Commission (2020) *White Paper: On Artificial Intelligence — A European Approach to Excellence and Trust.* Brussels 19.2.2020 COM (2020) 65 final.

[Frisch 1979] Frisch OR (1979) *What Little I Remember.* Cambridge: Cambridge University Press.

[Gibson 1950] Gibson JJ (1950) *The Perception of the Visual World.* Boston: Houghton Mifflin.

[Gibson 1966] Gibson JJ (1966) *The Senses Considered as Perceptual Systems.* Boston: Houghton Mifflin.

[Gibson 1979] Gibson JJ (1979) *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin.

[Giunchiglia and Walsh 1992] Giunchiglia F and Walsh T (1992) A theory of abstraction. *Artificial Intelligence,* 57, 323–389.

[Grenander 1993] Grenander U (1993) *General Pattern Theory: A Mathematical Study of Regular Structures.* Oxford: Clarendon Press.

[Grenander 1996] Grenander U (1996) *Elements of Pattern Theory.* Baltimore: John Hopkins University Press.

[Grenander 2012] Grenander U (2012) *A Calculus of Ideas: A Mathematical Study of Human Thought.* Hackensack NJ: World Scientific Publishing.

[Grenander and Miller 2007] Grenander U, Miller M (2007) *Pattern Theory: From Representation to Inference.* New York: Oxford University Press.

[Helvetius 1758] Helvétius CA (1758) *De l'Esprit, or Essays on the Mind and Its Several Faculties.* Engl transl Mudford W (1807). London: Jones. `https://books.google.de/books?id=TrdcAAAAcAAJ&printsec=frontcover&dq=GOOGLE+BOOKS+HELVETIUS+ENGLISH+ESPRIT+1807&hl=fr&sa=X&ved=2ahUKEwjo8eaS1f3vAhUF_rsIHZ7bDuEQ6AEwAHoECAMQAg#v=onepage&q=GOOGLE%20BOOKS%20HELVETIUS%20ENGLISH%20ESPRIT%201807&f=false`

[Herriot 1932] Herriot E (1932) *Normale.* Paris: La nouvelle société d' édition, collection 'Nos grandes écoles'.

[Hummel 1987] Hummel R (1987) The scale-space formulation of pyramid data structures. In: Uhr L (ed) *Parallel Computer Vision.* Boston: Academic Press.

[Kolmogorov 1965] Kolmogorov AN (1965) Three approaches to defining the notion 'quantity of information'. *Problemy Peredatchi Informatsii,* 1(1), 3–11. Reprinted in: Kolmogorov AN (1987) *Theory of Information and Theory of Algorithms.* Moscow: Nauka, 213–223. (Russian)

[Kurzweil 2012] Kurzweil R (2012) *How to Create a Mind: The Secret of Human Thought Revealed.* New York: Viking Penguin.

[Minsky 1975] Minsky M (1975) A framework for representing knowledge. In: Winston PH (ed) *The Psychology of Computer Vision.* New York: McGraw-Hill.

[Minsky 2006] Minsky M (2006) *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind.* New York: Simon and Schuster.

[Minsky and Papert 1988] Minsky M, Papert S (1988) *Perceptrons,* 2nd ed. Cambridge MA: MIT Press.

[Mont-Reynaud and Gresset 1990] Mont-Reynaud B, Gresset E (1990) PRISM: Pattern recognition in sound and music. *Proceedings of the International Computer Music Conference'1990.* Glasgow, 153–155.

[Mont-Reynaud and Mellinger 1989] Mont-Reynaud B, Mellinger D (1989) Source separation by frequency co-modulation. *Proceedings of the First International Conference on Music Perception and Cognition, Kyoto, Japan, 17–19 October, 1989*, 99–102.

[Mumford 1996] Mumford D (1996) Pattern theory: A unified perspective. In: Knill D, Richards W (eds) *Perception as Bayesian inference.* Cambridge MA: Cambridge University Press, 25–62.

[Mumford and Desolneux 2010] Mumford D, Desolneux A (2010) *Pattern Theory, the Stochastic Analysis of Real World Signals.* AKPeters/CRS Press.

[Palmer 1975] Palmer SE (1975) Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In: Norman DA et al (eds) *Exploration in Cognition.* Hillsdale NJ: Erlbaum.

[Palmer 1983] Palmer SE (1983) The psychology of perceptual organization: A transformational approach. In: Beck J, Hope B, Rosenfeld A (eds) *Human and Machine Vision.* New York: Academic Press, 269–339.

[Plato 380 BC] Plato (380 BC) Meno. Engl transl Jowett B. `http://classics.mit.edu/Plato/meno.html`

[Plato 360 BC] Plato (360 BC) Phaedo. Engl transl Jowett B. `http://classics.mit.edu/Plato/phaedo.html`

[Rossing 1990] Rossing TD (1990) *The Science of Sound*, 2nd ed. Reading MA: Addison-Wesley.

[Tanguiane 1993] Tanguiane AS (1993) *Artificial Perception and Music Recognition.* Berlin: Springer (Lecture Notes in Artificial Intelligence No. 746).

[Tanguiane 1994] Tanguiane AS (1994) A principle of correlativity of perception and its applications to music recognition. *Music Perception,* 11(4), 465–502.

[Tanguiane 1995] Tanguiane AS (1995) Towards axiomatization of music perception. *Journal of New Music Research*, 24 (3), 247–281.

[Tangian 1998] Tangian AS (1998) An operational definition of rhythm and tempo. *General Psychology (GenPsy).* Special issue 'Temporal Dynamics and Cognitive Processes', 55–89.

[Wertheimer 1923] Wertheimer M (1923) Untersuchungen zur Lehre von der Gestalt, II. *Psychologische Forschung*, 4, 301–350. Condensed transl in: Ellis WD *A Source Book of Gestalt Psychology, Selection 5.* New York: Humanities Press, 1950. Also in: Beardslee DC, Wertheimer M (eds) *Readings in Perception, Selection 8.* Princeton NJ: Van Nostrand Reinhold, 1958.

[Witkin 1983] Witkin AP (1983) Scale-space filtering. *Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany,* 1019–1024.

[Witkin and Tenenbaum 1983] Witkin AP, Tenenbaum JM (1983) On the role of structure in vision. In: Beck J, Hope B, Rosenfeld A (eds) *Human and Machine Vision.* New York: Academic Press, 481–543.

[Wuorinen 1979] Wuorinen Ch (1979) *Simple Composition.* New York: Longman.

[Xenakis 1971] Xenakis I (1971) *Formalized Music.* Bloomington: Indiana University Press.

KIT Scientific Working Papers
ISSN 2194-1629

www.kit.edu