An Artificial Perception Model and Its Application to Music Recognition

Andranick Tanguiane
ACROE-LIFIA, 46, av. Félix Viallet, 38000 Grenoble, France

Abstract
The data obtained by signal processing are self-organized in order to separate patterns before their recognition. The performance of the model is similar to perceiving objects in abstract painting without their explicit recognition.

The input information is described in terms of generative patterns and their transformations. Such a representation is less complex than the data themselves. The complexity of data is understood by Kolmogorov, i.e. as the amount of memory required for their storage. Representing the data in the most efficient way, we obtain their description, revealing certain causal relationships.

The approach is applied to voice separation. Chord spectra are described in terms of generative subspectra, providing the decomposition of polyphony into parts and chords into notes.

It is shown that the representation of a chord spectrum corresponding to the causality in the data generation is least complex. The demonstration is based on the factorization of chord spectra into the convolution product of indecomposable spectra, similarly to the factorization of integers into primes.

The model explains logarithmic scaling in pitch perception and the insensitivity of the ear to the phase of the signal as the conditions necessary for voice separation. The model explains also some rules of music theory as simplifying adequate perception of polyphony.

Keywords: Artificial perception, pattern recognition, automatic notation of music, voice separation, rhythm/tempo tracking, signal processing.

1 Introduction
We distinguish two stages in pattern recognition:
(a) object segregation, i.e. grouping data into messages;
(b) object identification, i.e. matching the segregated messages to known patterns.

For example, the first stage corresponds to distinguishing lines, spots, etc. in abstract painting, but their explicit recognition is the task for the second stage.

We deal with the first stage of "not-intelligent" perception. The related model of so-called correlative perception is based on some general principles and properties of human perception rather than on any particular knowledge about the patterns.

The model is applied to voice separation. The contemporary state of the research is reviewed in (Tanguiane 1993), where one can find more related references.

In Section 2, "Principle of Correlativity of Perception", we introduce some basic assumptions about data representation.

In Section 3, "Applications to Voice Separation", we formulate the problem of chord recognition as the problem of recognizing acoustical contours.

In Section 4, "Problems of Justification", we enumerate the questions to be answered in order to substantiate our model.

In Section 5, "Generation of Chord Spectra", we represent a chord spectrum as generated by multiple translation of a tone spectrum.

In Section 6, "Factorization of Chord Spectra", we formulate the problem of chord decomposition as deconvolution of the chord spectrum and show that there is the only non-trivial deconvolution of a musical chord which corresponds to the chord generation.

In Section 7, "Causality and Optimal Data Representation", we show that this non-trivial deconvolution of the chord spectrum is less complex than the spectral data themselves.

In Section 8, "Applications to Psychoacoustics and Music Theory", we discuss the assumptions of our model. It is shown that logarithmic scaling and the insensitivity to the phase of the signal are essential conditions in order to realize the chord recognition. Besides, we show that some statements of music theory can be interpreted from the standpoint of our model as providing the conditions for adequate perception of polyphony.

In Section 9, "Conclusions", we enumerate the main statements of the paper.

68.4  294  ICMC Proceedings 1993
Table 1: Complexity of representation of events in Fig. 2

<table>
<thead>
<tr>
<th>Complexity of rhythmic pattern</th>
<th>Complexity of its transformation</th>
<th>Total complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>16</td>
</tr>
</tbody>
</table>

To explain such an ambiguity in rhythm perception, estimate the complexity of the four representations. Suppose that one byte is needed to code a duration, and two bytes are needed to code a duration with pitch. Also suppose that in order to call the repeat algorithm we need four bytes. Under such conventions the complexity of the representations (in bytes) is given in Table 1. In case of pure rhythm its representation as a single pattern is less complex, whereas in the melodic context its representation as a repeat is more efficient.

Thus the hierarchical principle of data representation is provided with a feedback, guiding the process of data representation in the simplest way.

3 Applications to Voice Separation

A voice is associated with a group of partials which move in parallel with respect to a log-scaled frequency axis. A dynamical trajectory of the group of partials is called a melodic line, or a part in polyphony.

From our standpoint, a melodic line is a high-level pattern, generated by low-level patterns which are associated with notes. A statistical contour generated by a repetition of a group of partials is a high-level pattern of a chord. In the case of voices (recognize chords), one has to represent spectral data in terms of generative elements and their trajectories (contours).

Note that our approach to voice separation is based on recognizing the similarity of voices, in contrast to the approach (Chafe & Jaffe 1986; Mont-Reynaud & Mellinger 1989), based on recognizing their dissimilarity.

To illustrate the procedure of chord recognition, consider two simple chords shown in Fig. 3a which
Figure 3: Spectral representations of chords: a) chords \((c_1; a_1)\) and \((f_1; a_1)\); b) their log-scaled spectra for tones with 5 partials (marked by black or white respectively); c) the discrete clipped audio spectra of the chords to within a semitone (the arrows show the parallel motion of partials)

have the spectra shown in Fig. 3b. Represent the spectra of the chords as shown in Fig. 3c. Thus a chord is identified by a binary string \(s_n(k)\), where \(n\) is the number of chord (in our example \(n = 1, 2\)), \(k\) is the number of frequency band (in our example \(k = 1, \ldots, 34\)), and

\[ s_n(k) = \begin{cases} 1 & \text{if the signal has partials in the} \\
0 & \text{otherwise.} 
\end{cases} \]

We recognize melodic intervals between successive chords by peaks of the correlation function

\[ R_{n, n+1}(i) = \sum_k s_n(k) s_{n+1}(k + i). \]

If this function has a peak at point \(i\) we suppose that the correlating subspectra correspond to similar tones which form interval of \(i\) semitones. In this case the correlating subspectrum is said to be the generative group of partials of the given melodic interval.

Harmonic intervals in the \(i\)th chord are recognised by peaks of the autocorrelation function

\[ R_{n, n}(i) = \sum_k s_n(k) s_n(k + i). \]

Note that the correlation functions may have peaks at the points which don't correspond to real intervals. It is caused by coincidences of the partials belonging to different tones. Fortunately, the structure of groups of partials which correspond to real tones is repeated regularly, forming a stable spectral pattern. On the contrary, a correlating group of partials which doesn't correspond to a real tone has a random spectral structure. Therefore, in order to recognize true intervals, we must examine the generative groups of partials (revealed by correlation analysis) and reject the incidental groups with no stable structure (which occur only once). To recognize stable spectral patterns, we apply correlation analysis to different generative groups of partials.

Our approach is tested with computer experiments on recognizing harmonic and melodic intervals in J.S.Bach chorale (Fig. 4). The chorale has been considered as a sequence of 24 chords whose spectra have been computed and analyzed under various assumptions. The experiments have differed in voice type (harmonic or inharmonic), number of partials per voice (5, 10, or 15), frequency resolution (within 1, 1/2, 1/3, or 1/6 semitone), and optional restriction on the considered intervals up to 12 semitones (in order to avoid the octave autocorrelation). The recognition reliability in these 48 experiments has been always better than 90%.

4 Problems of Justification

Thus a chord spectrum is regarded as generated by a tone pattern translated along the log-scaled frequency axis, according to some interval structure. In this case, the low-level configurations are similar tone spectra, resulting from translations of a generative tone pattern. The high-level configuration is the contour circumscribed by the generative tone pattern, which is associated with the interval structure of the chord.

However, reasonable arguments and good recognition results are not yet a strict substantiation of the model. It may happen that the method is well fitted to the data considered, but will fail in some other experiments. For example, some chord spectra may be representable in terms of generative tone spectra in several ways, implying the ambiguity in their recognition, or the optimal (least complex) representation may differ from the chord decompo-
sition into notes.

Therefore, in order to substantiate our model, we must be sure that:

(a) representing a chord spectrum as translations of a generative tone spectrum, we obtain its description corresponding to its perception;

(b) such a description provides the optimal representation of spectral data;

(c) this optimal representation is unique.

Now we are going to confirm the above items for two-tone intervals, major triads, and minor triads. For this purpose we suggest a special mathematical machinery. First, we show that a chord spectrum is a convolution of a generative tone spectrum by an interval distribution. Then we establish an isomorphism between polynomials over integers and discrete audio spectra with respect to addition and convolution. Unlike polynomial over integers, the unique factorization doesn't hold for polynomials over positive integers (see below) which correspond to discrete power spectra. However, the unique factorization holds for the class of polynomials which corresponds to chord spectra built from harmonic tones. Next we show that power spectra of musical tones are irreducible, as well as interval distributions of two-tone intervals, major triads, and minor triads are irreducible as well. Hence, the only factorization of a chord spectrum equals to its generation.

This approach to spectral decomposition into irreducible factors can be applied in signal processing, speech recognition, and theory of generalized functions where the deconvolution problem arises. The isomorphism between polynomials and discrete spectra enables using algebraic methods in spectral analysis.

5 Generation of Chord Spectra

We shall consider audio spectra, i.e., we assume the frequency axis to be log-scaled, implying equal distances corresponding to equal musical intervals. We restrict ourselves to discrete power spectra limited in low frequencies and bounded in high frequencies, assuming that both frequency bands and signal levels are expressed by non-negative integers, while the number of bands with a positive level being always finite.

Thus by spectra we understand the expression of the form

$$S = S(x) = \sum_{n=0}^{N} a_n \delta(x - n) = \sum_{n=0}^{N} a_n \delta_n, \quad (1)$$

where $\delta_n$ are Dirac delta-functions and $a_n$ are non-negative integers. The value $a_n$ can be interpreted as the signal power in the $n$th frequency band. The phase of partials is ignored, corresponding to human perception.

The support of spectrum (1) is defined to be the set of its partial frequencies $\Delta_x = \{ n : a_n \neq 0 \}$.

Recall that a musical, or harmonic tone (with a pitch salience) is characterized by a harmonic ratio of the partial frequencies $1 : 2 : \ldots : k$. Then the support $\Delta_x$ of its discrete audio spectrum $S$ has the form

$$\Delta_x = \{ n_k : n_k = \lfloor \log_k (n + 0.5) \rfloor, \ k = 1, \ldots, K \}, \quad (2)$$

where $n_k$ is associated with the fundamental frequency or pitch, the constant $C$ characterizes the accuracy of spectral representation, being equal to the number of frequency bands per octave, and $\lfloor + 0.5 \rfloor$ is the rounding function, since $[x]$ retains the integer part of its argument. A spectrum $S$ whose support $\Delta_x$ has the form (2) is said to be harmonic.

Since we consider a log-scaled frequency axis, pitch shifts correspond to parallel translations of the tone spectrum along to the frequency axis. A translation of a spectrum (1) by $m$ bands to the right corresponds to the convolution

$$\delta_n + S = \sum_{m=0}^{N} a_n \delta_{n+m}. \quad (3)$$

As shown by Togniazi (1993), a chord spectrum can be regarded as generated by a multiple translation of a tone spectrum. By virtue of (3), a chord spectrum $S$ can be represented as follows

$$S = \sum_{m=0}^{N} \delta_n \delta_m \ast T = \sum_{m=0}^{N} b_m \delta_m \ast T = I \ast T, \quad (4)$$

where $T$ is a tone spectrum, and $I = \sum_{m=0}^{N} b_m \delta_m$ is the interval distribution. For example, if the frequency resolution is within one semitone ($C = 12$), then the major triad formed by major third and fifth (4 and 7 semitones, respectively) has the interval distribution $I_{4,7} = \delta_4 + \delta_7$. A spectrum $S$ is said to be simple if all its coefficients are relatively prime and the first coefficient $a_0 \neq 0$. Obviously, (4) can be written down as follows

$$S = a_0 T \ast I, \quad (5)$$

where the interval distribution $I$ and tone spectrum $T$ are simple. A simple interval distribution $I$ corresponds to the intervals between the lowest note and other notes of the chord, while its coefficients $b_m$ determining their relative loudness. The term $a_0 \cdot T$ can be understood as a spectrum of the lowest tone of the chord, having spectral pattern $T$, loudness $a_0$, and pitch $p$. 

ICMC Proceedings 1993

267

68.4
6 Factorization of Chord Spectra

In order to investigate the items enumerated in Introduction, we pose the question: Given a chord spectrum $S$ generated according to (5), does there exist a convolution factorization of $S$ into simple factors other than (5)?

**Lemma 1 (Isomorphism Between Discrete Spectra and Polynomials over Integers)** Define the correspondence between discrete spectra with integer coefficients and polynomials over integers by the equality of their coefficients

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

Then this correspondence is one-to-one, the sum of two spectra corresponds to the sum of the associated polynomials, and the convolution of two spectra corresponds to the product of the associated polynomials.

The proof of this proposition as well as proofs of subsequent ones can be found in (Tanguiane 1993). By analogy with polynomials, a spectrum $S$ is said to be irreducible if it cannot be factored into a convolution product of two spectra, each other than a constant coefficient.

**Lemma 2 (Sufficient Condition for Irreducible Spectra)** Consider a simple spectrum $S$ whose support $\Delta S$ contains two points at least. Suppose that the distance between the last (first) partials in $S$ is less than the distance between any other pair of partials. Then $S$ is irreducible.

**Lemma 3 (Irreducibility of Harmonic Spectra)** Let $S$ be a simple harmonic spectrum or a simple segment of a harmonic spectrum. If the spectral resolution is sufficiently accurate then $S$ is irreducible.

**Lemma 4 (Irreducibility of Intervals and Triads)** Simple internal distributions, corresponding to two-tone intervals and major or minor triads, are irreducible.

Note that the assumption of non-negativity of spectral coefficients is essential. Otherwise, even the interval of major triad is not irreducible.

**Example 1 (Reducibility of Major Triad)** Let the spectral resolution be one semitone $(C = 12)$, whence the major triad corresponds to 4 frequency bands. By virtue of Lemma 1, the reducibility of the major triad follows from the factorization $4 + z^2 = (2 + \sqrt{2} + z^2)(2 - \sqrt{2} + z^2)$.

Further we shall divide a spectrum

$$S = \sum_{n=1}^{N} a_n \delta_n,$$

whose partial (impulse) tones have frequencies

$$\omega_1 < \omega_2 < \ldots < \omega_N,$$

into lower and higher parts, which are said to be head and tail, as follows:

$$S = \sum_{n=1}^{N-k} a_n \delta_n + \sum_{n=N-k+1}^{N} a_n \delta_n.$$ 

Under these conventions the spectrum's higher part with $Q$ partials

$$S_Q = \sum_{n=N-k+1}^{N} a_n \delta_n$$

is said to be the $Q$-tail of $S$.

We say that two spectra $S$ and $T$ have congruent $Q$-tails if

$$S_Q = T_Q$$

for certain $Q$. This will be denoted

$$S_Q \sim T_Q.$$

**Lemma 5 (Unique Factorization for Spectral Tails)** Let $T, I, U,$ and $J$ be four spectra, each having more than one impulse, such that

$$T \ast I = U \ast J.$$ 

Let $d$ be the distance between the last two impulses of $T$ and $f$ be the distance between the last two impulses of $I$. Suppose that

$$d < f.$$ 

Define the tail $T_Q$ by the given bandwidth $f$, i.e., determine $Q$ so that

$$T_Q = \sum_{n=N-k+1}^{N} a_n \delta_n = \sum_{n=1}^{N-k} a_n \delta_n,$$

and denote the distance between the last and the $Q$th to last partial of $T$ by

$$g = \omega_N - \omega_{N-Q} < f.$$ 

Then either $T_Q \sim T_Q$ and the distance between the last two partials of $I$ is greater than or equal to $g$, or $T_Q \sim T_Q$ and the distance between the last two partials of $U$ is greater than or equal to $g$.

**Lemma 6 (Uniqueness of Interval Decomposition)** Consider an internal distribution $I$ with two

60.4 288

ICMC Proceedings 1993
impulses, and let the distance $f$ between its impulses be smaller than or equal to 12 semitones, i.e.

$$I = a_0 a_1 a_2 \quad (w_1 < w_2);$$
$$f = w_2 - w_1 \leq 12 \text{ semitones}.$$  

Let $T$ be a harmonic spectrum with $N$ successive partials, where $N$ is odd and sufficiently large so that the distance between the last three harmonics of $T$ is less than $f$, i.e.

$$T = \sum_{n=1}^{N} a_n w_n \quad (w_1 \ldots \leq w_N);$$
$$f > w_N - w_{N-2}.$$  

Consider spectrum $S = T \ast 1$. Then its decomposition into non-trivial factors (i.e. having at least two impulses each) is unique to within order and units.

Note that this lemma is valid not only for harmonic tones with all successive partials, but, by virtue of Lemma 3, for segments of harmonic spectra which contain three higher harmonics.

**Theorem 1 (Uniqueness of Interval Decomposition)** Consider a two-impulse interval distribution $I$ and let the distance $f$ between its impulses be smaller than or equal to 12 semitones, i.e.

$$I = a_0 a_2 \quad (w_1 < w_2);$$
$$f = w_2 - w_1 \leq 12 \text{ semitones}.$$  

Let $T$ be a harmonic spectrum with $N$ successive partials, where $N$ is sufficiently large so that the distance between the last few harmonics of $T$ is less than $f$, i.e.

$$T = \sum_{n=1}^{N} a_n w_n \quad (w_1 \ldots \leq w_N);$$
$$f > w_N - w_{N-2}.$$  

Consider spectrum $S = T \ast 1$. Then its decomposition into non-trivial factors (i.e. having at least two impulses each) is unique to within order and units.

**Lemma 7 (Uniqueness of Chord Decomposition)** Consider an interval distribution $I$ with three-impulses. Let the distance between its extreme impulses be smaller than or equal to the octave and let the distances between its adjacent impulses be different, i.e.

$$I = a_1 a_2 \quad (w_1 < w_2);$$
$$f = w_2 - w_1 \leq 12 \text{ semitones},$$
$$w_2 - w_1 \neq w_2 - w_3.$$  

Let $T$ be a harmonic spectrum with $N$ successive partials, where $N$ is not divisible by 2 and 3, and is sufficiently large so that the distance between the last few harmonics of $T$ is less than $f$, i.e.

$$T = \sum_{n=1}^{N} a_n w_n \quad (w_1 \ldots \leq w_N);$$
$$f > w_N - w_{N-3}.$$  

Consider chord spectrum $S = T \ast 1$. Then its decomposition into non-trivial factors (i.e. having at least two impulses each) is unique to within order and units.

**Theorem 2 (Uniqueness of Chord Decomposition)** Consider a three-impulse interval distribution $I$. Let the distance between its extreme impulses be smaller than or equal to the octave and let the distances between its adjacent impulses be different, i.e.

$$I = a_1 a_2 a_3 \quad (w_1 < w_2 < w_3);$$
$$f = w_3 - w_1 \leq 12 \text{ semitones},$$
$$w_3 - w_1 \neq w_3 - w_2.$$  

Let $T$ be a harmonic spectrum with $N$ successive partials, where $N$ is sufficiently large so that the distance between the last seven harmonics of $T$ is less than $f$, i.e.

$$T = \sum_{n=1}^{N} a_n w_n \quad (w_1 \ldots \leq w_N);$$
$$f > w_N - w_{N-3}.$$  

Consider chord spectrum $S = T \ast 1$. Then its decomposition into non-trivial factors (i.e. having at least two impulses each) is unique to within order and units.

Note that Theorem 2 is valid also for segments of harmonic spectra with seven successive harmonics. Thus spectra of two-tone intervals and major or minor triads are decomposable in the unique way. Consequently, the only decomposition of a chord spectrum, which is built from harmonic tones with a sufficient number of harmonics, reveals its generation.

Theorem 2 doesn't state the unique decomposition of chords if the number of partials of generative tones is small. This however can be done by directly testing each particular case from a finite number of cases.

To finish this section, we shall show that the harmonicity of tones is an important condition of the Unique Decomposition Theorem. For arbitrary power spectra the unique decomposition doesn't hold. By virtue of Lemma 1 this is seen from the following example proposed by Alain Chateau and in personal communication.
Example 2 (No Unique Factorization of Polynomials over Positive Integers) Consider the following polynomials:

\[ p(x) = x^2 + 2x + 2 \]
\[ q(x) = x^2 - 2x + 2 \]
\[ r(x) = x(x^2 + 2x + 2) + 1 \]
\[ = (x^2 + x + 1)(x + 1), \]

where polynomials \( p(x), q(x), x^2 + x + 1, \) and \( x + 1 \) are irreducible over integers. And, consequently, over non-negative integers as well. By virtue of Lemmas 1 and 4, polynomial

\[ p(x)q(x) = x^4 + 4 \]
is irreducible over non-negative integers. Therefore, polynomial with non-negative integer coefficients \( p(x)p(x)q(x) \) can be factored into irreducible polynomials over non-negative integers as follows

\[ p(x)p(x)q(x) = p(x)p(x)(x^2 + 4) = (x^2 + 4)(x^2 + x + 1)(x + 1). \] (7)

At the same time, polynomial \( p(x)p(x)\) can be factored into polynomials over non-negative integers in a different way:

\[ p(x)p(x) = (x^2 + 2x + 2)(x^2 + 4) = (x^2 + 2x + 2)(x^2 + x + 1)(x + 1). \] (8)

where the first factor, \( x^2 + 2x + 2 \), is irreducible, being different from all irreducible factors of factorization (7). Hence, there exist two different factorizations of polynomial \( p(x)p(x) \) into irreducible polynomials over non-negative integers, one given by (7) and another given by (8) with a further factorization of the second term \( x^2 + 4 \) with \( x^2 + x + 1 \), if such a further factorization exists.

7 Causality and Optimal Data Representation

One can ask a question: Why do we perceive chords as chords but not as single sounds? From the standpoint of our consideration we can reformulate this question as follows: What are the reasons in favor of decomposing spectra instead of considering them as they are?

In order to compare different representations we refer to the criterion of least complex data representation. We shall show that the representation of a chord spectrum in a form of deconvolution is the optimal representation of the chord spectrum with regard to the amount of memory needed for the storage of the spectral data. This way we justify such a representation of spectral data and adduce reasons in favor of perceiving chords as chords but not as indivisible sounds.

Recall that according to Kolmogorov, the complexity of data is defined to be the amount of memory required for their storage. Since the spectra considered can be stored as a sequence of impulses, the complexity of a spectral representation can be identified with the number of impulses to be stored.

By complexity of a spectrum \( S \) we understand the number of points in its support \( |S| \). The complexity of \( S \) is denoted by \( |S| \).

By complexity of a deconvolution \( S = T * I \) we understand the total complexity of the factors which is equal to \( |S| + |I| \).

Theorem 3 (Revealing Causality by Optimal Data Representation) Suppose that a spectrum \( S \) is generated by a spectrum \( T \) translated according to an integral distribution \( I \), where \( T \) is a harmonic spectrum or its segment with seven or more partials, and \( I \) corresponds to a two-tone interval, major triad, or minor triad. If the frequency resolution is sufficiently accurate then the spectrum representation corresponding to the spectrum generation \( S \) is the least complex representation of \( S \).

8 Applications to Psychoacoustics and Music Theory

Some statements of music theory can be explained as the prescriptions to provide comprehensible bearing of musical structure. In our model this corresponds to simplifying music recognition.

For example, we can explain the prohibition of parallel fifths in the counterpoint. Parallel fifths imply parallel motion of the partials associated with the voices. This makes the separation of voices more difficult. The resulting psychoacoustic effect is "timbral" rather than harmonical. This can break the homogeneity of musical texture, and therefore parallel fifths are avoided (or used intentionally). The timbres of parallel leading of parts is used in pipe organs where several pipes tuned in a chord are turned on by a single key.

The model explains the nature of interval bearing. The interval bearing can be understood as the capability to recognize the distance between the tones which are similar in spectral structure. In other words, the interval bearing is nothing but correlative perception in the frequency domain. The emphasized condition is rather a generalization than a restriction. Indeed, all musical tones, having the same ratio of harmonics, meet this condition. On the other hand, the idea of distance is applicable to similar sounds with no pitch salience, as bell-like sounds, or even band-pass noises. Therefore, we eliminate the idea of absolute pitch from the definition of interval bearing.
According to our model, the function of inter-
val hearing is decomposing acoustical streams and
tracking parallel acoustical processes. This is ex-
tremely important for the orientation in the acous-
tical environment.

Note the role of logarithmic scales in our consid-
eration. Owing to the use of logarithm, patterns
with a linear structure (as tone spectra with mul-
tiplication of partial frequencies) are non-linearly
compressed, becoming irredissoluble. Therefore, the
role of logarithmic scales in perception can be ex-
plained as providing the indecomposibility of pat-
terns. In particular, as follows from Lemma 3, har-
monic spectra are irredissoluble which meets the per-
ception of a musical tone as an entirety.

On the other hand, Theorem 3 explains the per-
ception of chords as composed sounds. It is noteworthy that the optimality in the representa-
tion corresponds to the causality of data genera-
tion. This way we obtain a way of getting seman-
tical knowledge based on general principles of data
processing.

Note that if we supposed the model sensitivity to
the phase of the signal, we couldn’t prove the ir-
redissolubility of harmonic spectra and interval distri-
butions of chords. Indeed, if we considered spectra
with complex coefficients, by the fundamental the-
orem of algebra the associated polynomials would be
always factored into linear terms, implying re-
dissolubility of all spectra. Similar difficulties arise
even for spectra with negative integral coefficients
(see the example after lemma 4 where negative coef-
cients, corresponding to the past 140°, are consid-
ered). This may explain the insensitivity of audio
perception to the phase of the signal: otherwise, the
signal decomposition would not correspond to
physical causality in the signal generation.

Finally, we would like to mention that the fac-
torization method is efficient for justifying the prin-
ciple of correlation of perception theoretically, but
may fail in applications.

The first reason is the difference of spectral and
polynomial approximations. In a sense, polynomial
approximations are stable with respect to deviations
of coefficients but they are not stable with
respect to changes of the degree of the polynomial.
In spectra, partials can deviate from theoretical fre-
frequencies, implying changes of the degree of the asso-
ciated polynomials. However, those spectra are usu-
ally considered as approximating theoretical spe-
tra, whereas the associated polynomials may have
quite different factorizations.

Another practical disadvantage of the factoriza-
tion approach is that factorizing real spectra requires
much computing when the associated polynomials
are of high degree.

Therefore, our method of spectrum decompo-
sition by correlation analysis can be more stable
and reliable in applications. On the other hand,
by virtue of the isomorphism between spectra and
polynomials, the correlation method may be used to
factorize polynomials.

9 Conclusions
We have suggested an "artificial perception" ap-
proach to data processing. It is proposed:
1. Representing data in terms of generative ele-
ments and their transformations.
2. Finding generative elements by discovering the
messages correlated under deformations of the
data.
3. Performing a directional search for the ap-
propriate transformations of the data by the
method of variable resolution.
4. Overcoming the ambiguity in the data repre-
sentation, applying the criterion of the least
complexity of data representation.
5. Justifying the approach by a series of math-
eratical statements showing the efficiency of
chord spectrum representations used.
6. This approach has been tested in chord recog-
nition and has been applied to explaining some
perception phenomena. In particular, we have
obtained 90% reliability in chord recognition.
Besides, we justify the logarithmic scaling in
pitch perception, the insensitivity of the ear to the
phase of the signal, and the prohibition of
parallel voice leading in strict counterpoint.

References
[Chaie and Jaffe, 1986] Chaie C. and Jaffe
D. Source Separation and Note Identification
in Polyphonic Music. Proc. of the IEEE-
IECEJ/IAS Int. Conf. on Acoustics, Speech,
and Signal Processing, Tokyo, 1986, vol. 2,
1286-1292.
[Mont-Reynaud and Mellinger, 1989] Mont-
Reynaud B. and Mellinger D. Source Separa-
tion by Frequency Co-Modulation. Proc. of
the First Int. Conf. on Music Perception and
Cognition, Kyoto, Japan, 17-19 October, 1989,
99-102.
[Tangianlu, 1993] Tangianlu A. S. Artificial Per-
ception and Music Recognition. Berlin:
Springer-Verlag, 1993.